Reference

Experiment Design and Analysis

ReliaSoft

© 1992-2025 Hottinger Bruel & Kjaer UK Ltd ("HBK UK Ltd"), at Technology Centre, Advanced Manufacturing Park, Brunel Way, Catcliffe, Rotherham S60 5WG. HBK UK Ltd is an affiliate of Hottinger Bruel & Kjaer Inc. ALL RIGHTS RESERVED.

This license agreement ("License Agreement") sets forth the terms and conditions that govern the distribution and use of the HBK UK Ltd software documentation (the "HBK Software Documentation"), including any and all derivative works. Any unauthorized distribution or use of the HBK Software Documentation is strictly prohibited.

Licensee hereby acknowledges and agrees to the following terms and conditions of this License Agreement for its use of the HBK Software Documentation:

HBK UK Ltd grants the Licensee who accepts and abides by the terms of this License Agreement a non-exclusive, non-transferable, royalty-free, and revocable license to the HBK Software Documentation. Unless otherwise indicated, all HBK Software Documentation are copyrighted and are the property of HBK UK Ltd or, where applicable, the relevant member of HBK UK Ltd's corporate group. They are licensed to you and for your use only as an individual who has purchased the software ("Licensee"). Notwithstanding this License Agreement, the Licensee shall not have a license to the trademarks, logos, or any other intellectual property of HBK UK Ltd and/or its affiliates or Licensor(s).

Licensee may print a single copy of the HBK Software Documentation for his/her reference. Licensee may reprint the HBK Software Documentation, as needed, if the original printed copy is damaged and/or destroyed.

Except as provided above, no part of the HBK Software Documentation, either text or image, may be used for any purpose other than Licensee's own personal use and reference as a learning aid. Therefore, the reproduction, modification, creation of derivative works, storage in a retrieval system, or retransmission, in any form or by any means, electronic, mechanical or otherwise, for reasons other than Licensee's personal use, is strictly prohibited.

Certain photos and images in the HBK Software Documentation are used under non-transferable licenses obtained by HBK UK Ltd and/or its affiliates and are owned by its Licensor(s) ("Licensor"). Images depicting photos of actual persons are licensed to HBK UK Ltd and/or its affiliates and the signed model releases for these images are on file with the Licensor(s). HBK UK Ltd makes no copyright claims on these images. All ownership and intellectual property rights to the HBK Software Documentation are reserved by either HBK UK Ltd and/or its affiliates or its Licensor(s).

DISCLAIMER: THE HBK SOFTWARE DOCUMENTATION IS PROVIDED “AS IS” WITHOUT WARRANTIES OF ANY KIND, EITHER EXPRESS OR IMPLIED,

INCLUDING, BUT NOT LIMITED TO: 1) ANY IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, OR NON-INFRINGEMENT OF INTELLECTUAL PROPERTY; 2) ANY WARRANTY THAT THE HBK COURSE MATERIALS WILL CONFORM TO SPECIFICATIONS; 3) ANY WARRANTY THAT THE WORK WILL BE ERROR FREE OR VIRUS FREE. IN NO EVENT SHALL HBK UK LTD, ITS AFFILIATES, DISTRIBUTORS, CONTRACTORS, AGENTS, AND ITS LICENSOR(S) BE LIABLE FOR ANY DAMAGES, INCLUDING, BUT NOT LIMITED TO, DIRECT, INDIRECT, SPECIAL OR CONSEQUENTIAL DAMAGES, ARISING OUT OF, RESULTING FROM, OR IN ANY WAY CONNECTED WITH THE HBK SOFTWARE DOCUMENTATION, WHETHER OR NOT BASED UPON WARRANTY, CONTRACT, TORT, OR OTHERWISE, WHETHER OR NOT LOSS WAS SUSTAINED FROM, OR AROSE OUT OF THE RESULTS OF, OR USE OF, THE HBK SOFTWARE DOCUMENTATION. LICENSEES AGREE TO WAIVE ANY AND ALL CLAIMS AGAINST HBK UK LTD, ITS AFFILIATES, DISTRIBUTORS, CONTRACTORS, AGENTS, AND ITS LICENSOR(S), AND SHALL INDEMNIFY, DEFEND AND HOLD HARMLESS HBK UK LTD, ITS AFFILIATES, DISTRIBUTORS, CONTRACTORS, AGENTS, AND ITS LICENSOR(S) FOR ANY AND ALL LIABILITIES, CLAIMS, DEMANDS, DAMAGES, EXPENSES OR LOSSES THAT MAY ARISE FROM THE LICENSEE'S USE OR DISTRIBUTION OF THE HBK SOFTWARE DOCUMENTATION, INCLUDING ANY LIABILITIES OR DAMAGES FROM DERIVATIVE WORKS OR OTHER PRODUCTS BASED ON, OR RESULTING FROM, THE USE THEREOF.

This License Agreement is subject to change without notice and does not represent any commitment on the part of HBK UK Ltd and/or its affiliates to the Licensee, including any commitment to maintain or update the HBK Software Documentation. The names of companies, products, people, characters, and/or data mentioned in the HBK Software Documentation are not intended to represent any real individual, company, product or event, unless otherwise noted. Any rights not expressly granted herein are reserved for HBK UK Ltd.

TRADEMARKS:

ReliaSoft, Weibull++, BlockSim, Lambda Predict, XFMEA, RCM++ and XFRACAS are all trademarks of Hottinger Bruel & Kjaer Inc.

GlyphWorks, VibeSys, DesignLife and the nCode logos are all trademarks of HBK UK Ltd.

Other product names and services identified in the HBK Software Documentation are trademarks belonging to their respective trademark holders, and are used for illustration purposes. Their use in no way conveys an endorsement or other affiliation with HBK UK Ltd and/or its affiliates.

This License Agreement is subject to change without notice and does not represent any commitment on the part of HBK UK Ltd and/or its affiliates to the Licensee, including any commitment to maintain or update the HBK Software Documentation. The names of companies, products, people, characters, and/or data mentioned in the HBK Software Documentation are not intended to represent any real individual, company, product or event, unless otherwise noted. Any rights not expressly granted herein are reserved for HBK UK Ltd.

TABLE OF CONTENTS

DOE OVERVIEW 1

STATISTICAL BACKGROUND ON DOE 7

SIMPLE LINEAR REGRESSION ANALYSIS 31

MULTIPLE LINEAR REGRESSION ANALYSIS 65

ONE FACTOR DESIGNS 111

GENERAL FULL FACTORIAL DESIGNS 128

RANDOMIZATION AND BLOCKING IN DOE 146

TWO LEVEL FACTORIAL EXPERIMENTS 155

HIGHLY FRACTIONAL FACTORIAL DESIGNS 215

RESPONSE SURFACE METHODS FOR OPTIMIZATION 230

DESIGN EVALUATION AND POWER STUDY 262

OPTIMAL CUSTOM DESIGNS 307

ROBUST PARAMETER DESIGN 317

MIXTURE DESIGN 332

RELIABILITY DOE FOR LIFE TESTS 363

MEASUREMENT SYSTEM ANALYSIS 396

APPENDICES 439

DOE Overview

IN THIS CHAPTER

Why DOE? **2**

Introduction to DOE Principles **2**

DOE Types **3**
For Comparison: One Factor Designs 3
For Factor Screening: Factorial Designs 3
For Optimization: Response Surface Method Designs 5
For Product or Process Robustness: Robust Parameter Designs 5
For Life Tests: Reliability DOE 5
For Experiments with Constraints: Optimal Custom Design 5

Stages of DOE **5**
Planning 5
Screening 6
Optimization 6
Robustness Testing 6
Verification 6

Much of our knowledge about products and processes in the engineering and scientific disciplines is derived from experimentation. An experiment is a series of tests conducted in a systematic manner to increase the understanding of an existing process or to explore a new product or process. Design of experiments (DOE), then, is the tool to develop an experimentation strategy that maximizes learning using a minimum of resources. DOE is widely used in many fields with broad application across all the natural and social sciences. It is extensively used by engineers and scientists involved in the improvement of manufacturing processes to maximize yield and decrease variability. Often engineers also work on products or processes where no scientific theories or principles are directly applicable. Experimental design techniques become extremely important in such studies to develop new products and processes in a cost effective and confident manner.

Why DOE?

With modern technological advances, products and processes are becoming exceedingly complicated. As the cost of experimentation rises rapidly, it is becoming increasingly difficult for the analyst, who is already constrained by resources and time, to investigate the numerous factors that affect these complex processes using trial and error methods. Instead, a technique is needed that identifies the "vital few" factors in the most efficient manner, and then directs the process to its best setting to meet the ever increasing demand for improved quality and increased productivity. DOE techniques provide powerful and efficient methods to achieve these objectives.

Designed experiments are much more efficient than one-factor-at-a-time experiments, which involve changing a single factor at a time to study the effect of the factor on the product or process. While one-factor-at-a-time experiments are easy to understand, they do not allow the investigation of how a factor affects a product or process in the presence of other factors. An *interaction* is the relationship whereby the effect that a factor has on the product or process is altered due to the presence of one or more other factors. Oftentimes interaction effects are more important than the effect of individual factors. This is because the application environment of the product or process includes the presence of many of the factors together instead of isolated occurrences of one of the factors at different times. Consider an example of interaction between two factors in a chemical process, where increasing the temperature alone increases the yield slightly while increasing the pressure alone has no effect. However, in the presence of both higher temperature and higher pressure the yield increases rapidly. In this case, an interaction is said to exist between the two factors affecting the chemical reaction.

The DOE methodology ensures that all factors and their interactions are systematically investigated. Therefore, information obtained from a DOE analysis is much more reliable and complete than results from one-factor-at-a-time experiments that ignore interactions and thus may lead to incorrect conclusions.

Introduction to DOE Principles

The design and analysis of experiments revolves around the understanding of the effects of different variables on another variable. In technical terms, the objective is to establish a *cause-and-effect relationship* between a number of *independent variables* and a *dependent variable* of interest. The dependent variable, in the context of DOE, is called the *response*, and the independent variables are called *factors*. Experiments are run at different factor values, called *levels*. Each run of an experiment involves a combination of the levels of the investigated factors, and each of the combinations is referred to as a *treatment*. When the same number of response obser-

vations are taken for each of the treatments of an experiment, the design of the experiment is said to be *balanced*. Repeated observations at a given treatment are called *replicates*.

The number of treatments of an experiment is determined on the basis of the number of factor levels being investigated. For example, if an experiment involving two factors is to be performed, with the first factor having *m* levels and the second having *n* levels, then *m* x *n* treatment combinations can possibly be run, and the experiment is an *m* x *n* factorial design. If all *m* x *n* combinations are run, then the experiment is a *full factorial*. If only some of the *m* x *n* treatment combinations are run, then the experiment is a *fractional factorial*. In full factorial experiments, all the factors and their interactions can be investigated, whereas in fractional factorial experiments, at least some interactions are not considered because some treatments are not run.

It can be seen that the size of an experiment escalates rapidly as the number of factors (or the number of the levels of the factors) increases. For example, if 2 factors at 3 levels each are to be used, 9 (3x3=9) different treatments are required for a full factorial experiment. If a third factor with 3 levels is added, 27 (3x3x3=27) treatments are required, and 81 (3x3x3x3=81) treatments are required if a fourth factor with three levels is added. If only two levels are used for each factor, then in the four-factor case, 16 (2x2x2x2=16) treatments are required. For this reason, many experiments are restricted to two levels, and these designs are given a special treatment in this reference. Using a fractional design further reduces the number of required treatments.

DOE Types

For Comparison: One Factor Designs

With these designs, only one factor is under investigation, and the objective is to determine whether the response is significantly different at different factor levels. The factor can be qualitative or quantitative. In the case of qualitative factors (e.g., different suppliers, different materials, etc.), no extrapolations (i.e., predictions) can be performed outside the tested levels, and only the effect of the factor on the response can be determined. On the other hand, data from tests where the factor is quantitative (such as temperature, voltage, load, etc.) can be used for both effect investigation and prediction, provided that sufficient data is available. (In Weibull++ DOE folios, predictions for one factor designs can be performed using the multiple linear regression folio or free form folio.)

For Factor Screening: Factorial Designs

In factorial designs, multiple factors are investigated simultaneously during the test. As in one factor designs, qualitative and/or quantitative factors can be considered. The objective of these designs is to identify the factors that have a significant effect on the response, as well as

investigate the effect of interactions (depending on the experiment design used). Predictions can also be performed when quantitative factors are present, but care must be taken since certain designs are very limited by the choice of the predictive model. For example, in two level designs only a linear relationship can be used between the response and the factors, which may not be realistic.

- **General Full Factorial Designs**

In general full factorial designs, the factors can have different number of levels, and they can be quantitative or qualitative.

- **Two Level Full Factorial Designs**

With these designs, all factors must have only two levels. Restricting the levels to two and running a full factorial experiment reduces the number of treatments (compared to a general full factorial experiment), and it allows for the investigation of all the factors and all their interactions. If all factors are quantitative, then the data from such experiments can be used for predictive purposes, provided a linear model is appropriate for modeling the response (since only two levels are used, curvature cannot be modeled).

- **Two Level Fractional Factorial Design**

This is a special category of two level designs, where not all factor level combinations are considered, and the experimenter can choose which combinations are to be excluded. Based on the excluded combinations, certain interactions cannot be investigated.

- **Plackett-Burman Design**

This is a special category of two level fractional factorial designs, proposed by R. L. Plackett and J. P. Burman [1946], where only a few specifically chosen runs are performed to investigate just the main effects (i.e., no interactions).

- **Taguchi's Orthogonal Arrays**

Taguchi's orthogonal arrays are highly fractional designs, used to estimate main effects using only a few experimental runs. These designs are not only applicable to two level factorial experiments, but also can investigate main effects when factors have more than two levels. Designs are also available to investigate main effects for certain mixed level experiments where the factors included do not have the same number of levels.

For Optimization: Response Surface Method Designs

These are special designs that are used to determine the settings of the factors to achieve an optimum value of the response.

For Product or Process Robustness: Robust Parameter Designs

The famous Taguchi robust design is for robust parameter design. It is is used to design a product or process to be insensitive to noise factors.

For Life Tests: Reliability DOE

This is a special category of DOE where traditional designs, such as the two level designs, are combined with reliability methods to investigate effects of different factors on the life of a unit. In reliability DOE, the response is a life metric (e.g., age, miles, cycles, etc.), and the data may contain censored observations (suspensions, interval data).

For Experiments with Constraints: Optimal Custom Design

The optimal custom design tool can be used to modify the above standard designs to plan an experiment that meets any or all of the following constraints: 1) limited availability of test samples, 2) factor level combinations that cannot be tested, 3) factor level combinations that must be tested or 4) specific factors effects that must be investigated.

Stages of DOE

Designed experiments are usually carried out in five stages: planning, screening, optimization, robustness testing and verification.

Planning

It is important to carefully plan for the course of experimentation before embarking upon the process of testing and data collection. A thorough and precise objective identifying the need to conduct the investigation, an assessment of time and resources available to achieve the objective and an integration of prior knowledge to the experimentation procedure are a few of the goals to keep in mind at this stage. A team composed of individuals from different disciplines related to the product or process should be used to identify possible factors to investigate and determine the most appropriate response(s) to measure. A team-approach promotes synergy that gives a richer set of factors to study and thus a more complete experiment. Carefully planned experiments always lead to increased understanding of the product or process.

Screening

Screening experiments are used to identify the important factors that affect the system under investigation out of the large pool of potential factors. These experiments are carried out in conjunction with prior knowledge of the system to eliminate unimportant factors and focus attention on the key factors that require further detailed analyses. Screening experiments are usually efficient designs requiring a few executions where the focus is not on interactions but on identifying the vital few factors.

Optimization

Once attention is narrowed down to the important factors affecting the process, the next step is to determine the best setting of these factors to achieve the desired objective. Depending on the product or process under investigation, this objective may be to either maximize, minimize or achieve a target value of the response.

Robustness Testing

Once the optimal settings of the factors have been determined, it is important to make the product or process insensitive to variations that are likely to be experienced in the application environment. These variations result from changes in factors that affect the process but are beyond the control of the analyst. Such factors as humidity, ambient temperature, variation in material, etc. are referred to as *noise factors*. It is important to identify sources of such variation and take measures to ensure that the product or process is made insensitive (or robust) to these factors.

Verification

This final stage involves validation of the best settings of the factors by conducting a few follow-up experiment runs to confirm that the system functions as desired and all objectives are met.

Statistical Background on DOE

IN THIS CHAPTER

Basic Concepts **8**
Random Variables and the Normal Distribution 8
Population Mean, Sample Mean and Variance 10
Central Limit Theorem 11
Unbiased and Biased Estimators 12
Degrees of Freedom (dof) 14
Standard Normal Distribution 14
Chi-Squared Distribution 14
Student's t Distribution (t Distribution) 16
F Distribution 17

Hypothesis Testing **17**
Two-sided and One-sided Hypotheses 20

Statistical Inference for a Single Sample **20**
Inference on the Mean of a Population When the Variance Is Known 20
P Value 23
Inference on Mean of a Population When Variance Is Unknown 24
Inference on Variance of a Normal Population 26

Statistical Inference for Two Samples **27**
Inference on the Difference in Population Means When Variances Are Known 27
Inference on the Difference in Population Means When Variances Are Unknown 28
Inference on the Variances of Two Normal Populations 28

Variations occur in nature, be it the tensile strength of a particular grade of steel, caffeine content in your energy drink or the distance traveled by your vehicle in a day. Variations are also seen in the observations recorded during multiple executions of a process, even when all factors are strictly maintained at their respective levels and all the executions are run as identically as possible. The natural variations that occur in a process, even when all conditions are maintained at the same level, are often called *noise*. When the effect of a particular factor on a process is studied, it becomes extremely important to distinguish the changes in the process caused by the

factor from noise. A number of statistical methods are available to achieve this. This chapter covers basic statistical concepts that are useful in understanding the statistical analysis of data obtained from designed experiments. The initial sections of this chapter discuss the normal distribution and related concepts. The assumption of the normal distribution is widely used in the analysis of designed experiments. The subsequent sections introduce the standard normal, chi-squared, t and F distributions that are widely used in calculations related to hypothesis testing and confidence bounds. This chapter also covers hypothesis testing. It is important to gain a clear understanding of hypothesis testing because this concept finds direct application in the analysis of designed experiments to determine whether or not a particular factor is significant [Wu, 2000].

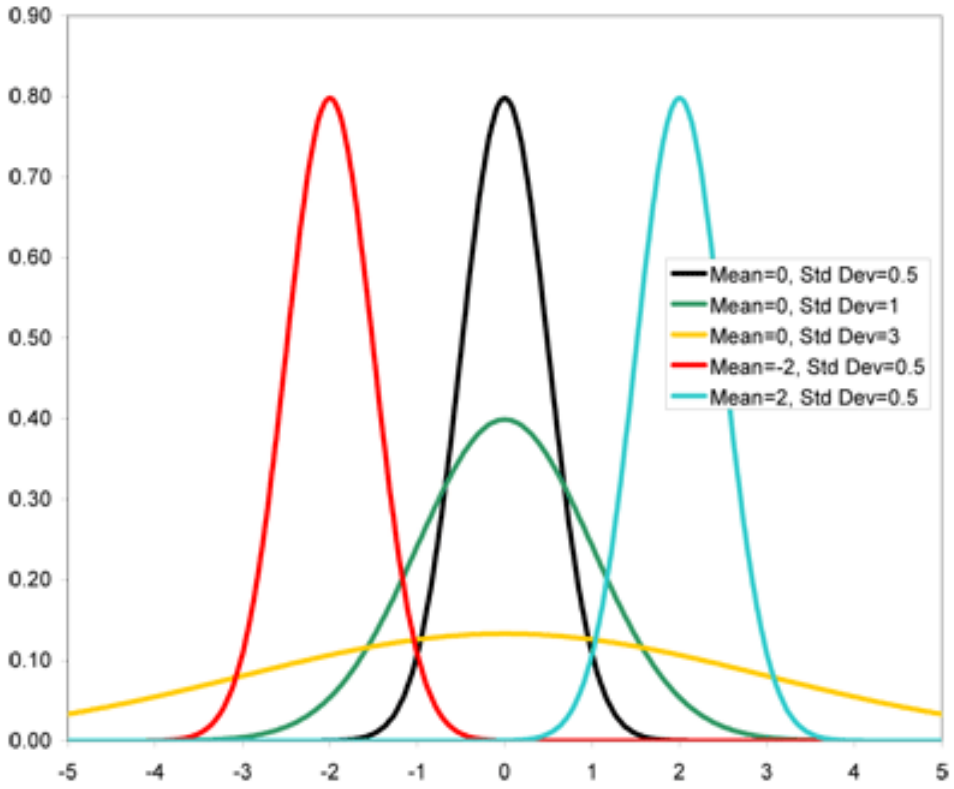
Basic Concepts

Random Variables and the Normal Distribution

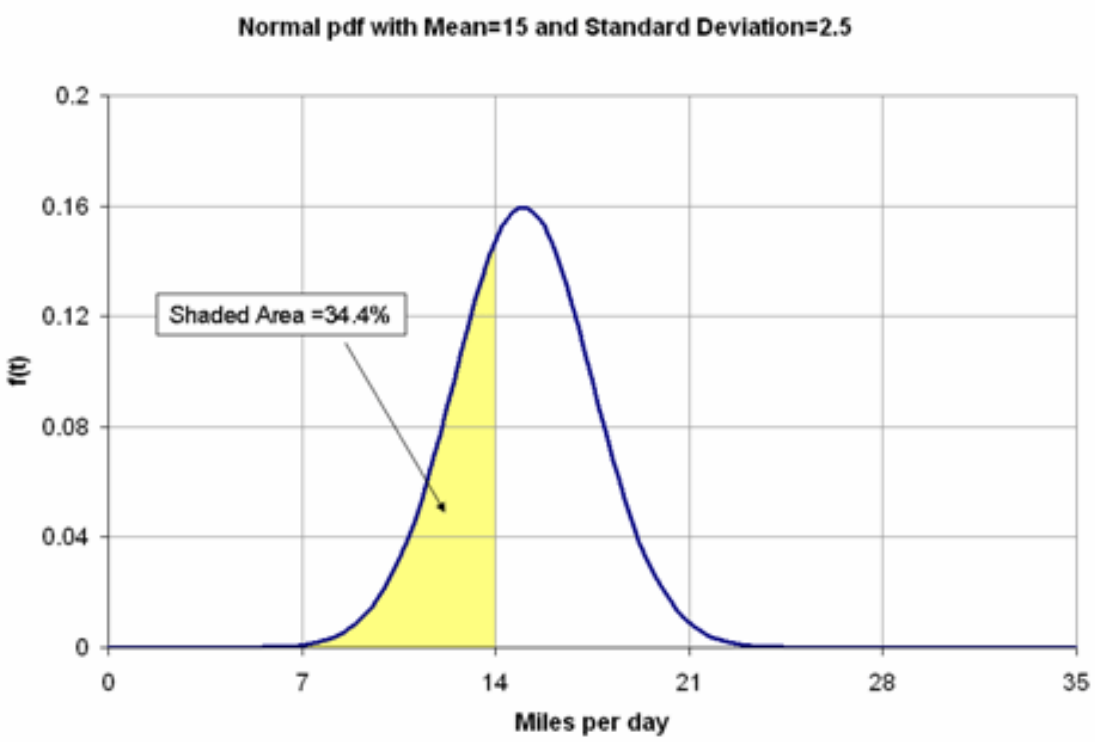
If you record the distance traveled by your car everyday, you'll notice that these values show some variation because your car does not travel the exact same distance every day. If a variable X is used to denote these values then X is considered a *random variable* (because of the diverse and unpredicted values X can have). Random variables are denoted by uppercase letters, while a measured value of the random variable is denoted by the corresponding lowercase letter. For example, if the distance traveled by your car on January 1 was 10.7 miles, then:

$$x = 10.7 \text{ miles}$$

A commonly used distribution to describe the behavior of random variables is the normal distribution. When you calculate the mean and standard deviation for a given data set, a common assumption used is that the data follows a normal distribution. A normal distribution (also referred to as the Gaussian distribution) is a bell-shaped curved (see figure below). The mean and standard deviation are the two parameters of this distribution. The mean determines the location of the distribution on the x-axis and is also called the *location parameter*. The standard deviation determines the spread of the distribution (how narrow or wide) and is thus called the *scale parameter*. The standard deviation, or its square called *variance*, gives an indication of the variability or spread of data. A large value of the standard deviation (or variance) implies that a large amount of variability exists in the data.



Any curve in the image below is also referred to as the probability density function, or *pdf* of the normal distribution, as the area under the curve gives the probability of occurrence of $\boldsymbol{X}$ for a particular interval. For instance, if you obtained the mean and standard deviation for the distance data of your car as 15 miles and 2.5 miles respectively, then the probability that your car travels a distance between 7 miles and 14 miles is given by the area under the curve covered between these two values, which is calculated to be 34.4% (see figure below). This means that on 34.4 days out of every 100 days your car travels, your car can be expected to cover a distance in the range of 7 to 14 miles.



On a normal probability density function, the area under the curve between the values of $Mean - (3 \times StandardDeviation)$ and $Mean + (3 \times StandardDeviation)$ is approximately 99.7% of the total area under the curve. This implies that almost all the time (or 99.7% of the time) the distance traveled will fall in the range of 7.5 miles $(15 - 3 \times 2.5)$ and 22.5 miles $(15 + 3 \times 2.5)$. Similarly, $Mean \pm (2 \times StandardDeviation)$ covers approximately 95% of the area under the curve and $Mean \pm (StandardDeviation)$ covers approximately 68% of the area under the curve.

Population Mean, Sample Mean and Variance

If data for all of the population under investigation is known, then the mean and variance for this population can be calculated as follows:

Population Mean:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

Population Variance:

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

Here, N is the size of the population.

The population standard deviation is the positive square root of the population variance.

Most of the time it is not possible to obtain data for the entire population. For example, it is impossible to measure the height of every male in a country to determine the average height and variance for males of a particular country. In such cases, results for the population have to be estimated using samples. This process is known as *statistical inference*. Mean and variance for a sample are calculated using the following relations:

Sample Mean:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Sample Variance:

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

Here, n is the sample size. The sample standard deviation is the positive square root of the sample variance. The sample mean and variance of a random sample can be used as estimators of the population mean and variance, respectively. The sample mean and variance are referred to as *statistics*. A statistic is any function of observations in a random sample. You may have noticed that the denominator in the calculation of sample variance, unlike the denominator in the calculation of population variance, is $(n-1)$ and not n. The reason for this difference is explained in Biased Estimators.

Central Limit Theorem

The Central Limit Theorem states that for a large sample size, n:

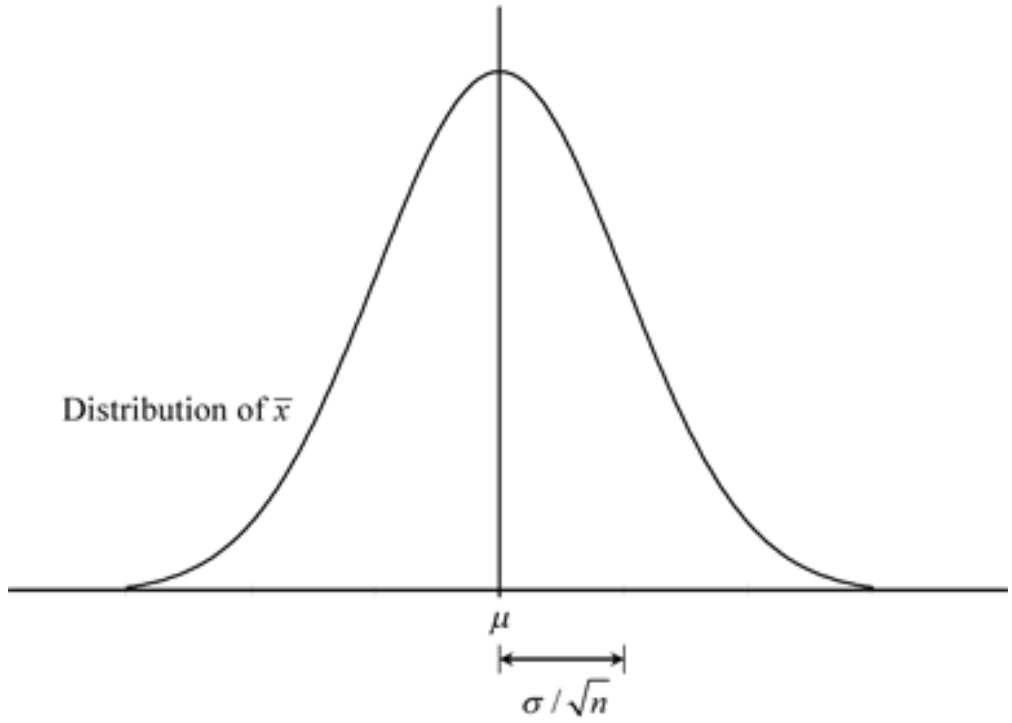
- The sample means from a population are normally distributed with a mean value equal to the population mean, μ, even if the population is not normally distributed.

  What this means is that if random samples are drawn from any population and the sample mean, $\bar{x}$, calculated for each of these samples, then these sample means would follow the normal distribution with a mean (or location parameter) equal to the population mean, μ. Thus, the distribution of the statistic, $\bar{x}$, would be a normal distribution with mean, μ. The distribution of a statistic is called the *sampling distribution*.

- The variance, s^2, of the sample means would be n times smaller than the variance of the population, σ^2.

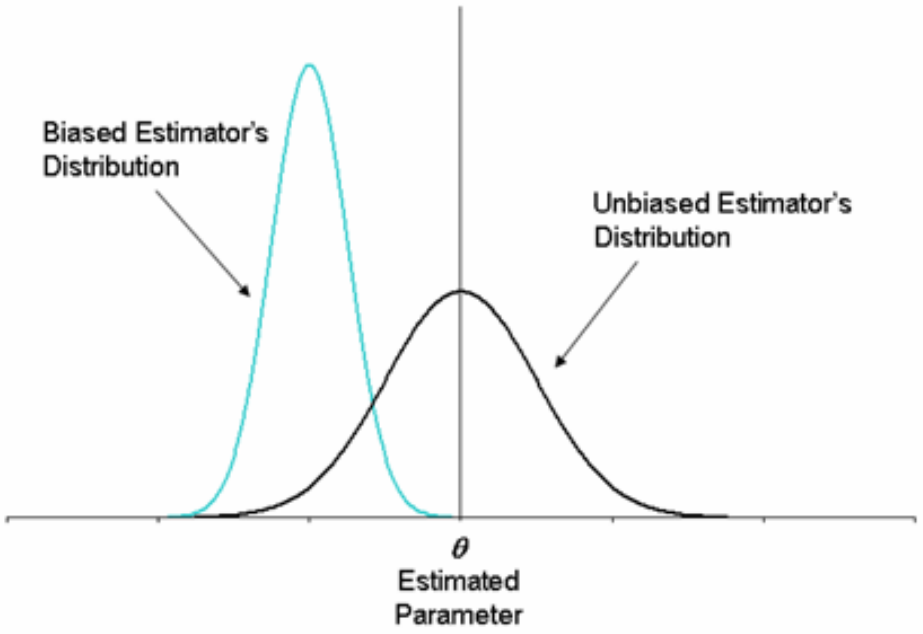
  This implies that the sampling distribution of the sample means would have a variance equal to σ^2/n (or a scale parameter equal to $\sigma/\sqrt{n}$), where σ is the population standard deviation. The standard deviation of the sampling distribution of an estimator is called the standard error of the estimator. Thus the standard error of sample mean $\bar{x}$ is $\sigma/\sqrt{n}$.

In short, the Central Limit Theorem states that the sampling distribution of the sample mean is a normal distribution with parameters μ and $\sigma/\sqrt{n}$ as shown in the figure below.



Unbiased and Biased Estimators

If the mean value of an estimator equals the true value of the quantity it estimates, then the estimator is called an *unbiased estimator* (see figure below). For example, assume that the sample mean is being used to estimate the mean of a population. Using the Central Limit Theorem, the mean value of the sample mean equals the population mean. Therefore, the sample mean is an unbiased estimator of the population mean. If the mean value of an estimator is either less than or greater than the true value of the quantity it estimates, then the estimator is called a *biased estimator*. For example, suppose you decide to choose the smallest observation in a sample to be the estimator of the population mean. Such an estimator would be biased because the average of the values of this estimator would always be less than the true population mean. In other words, the mean of the sampling distribution of this estimator would be less than the true value of the population mean it is trying to estimate. Consequently, the estimator is a biased estimator.



A case of biased estimation is seen to occur when sample variance, s^2, is used to estimate the population variance, σ^2, if the following relation is used to calculate the sample variance:

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$$

The sample variance calculated using this relation is always less than the true population variance. This is because deviations with respect to the sample mean, $\bar{x}$, are used to calculate the sample variance. Sample observations, x_i, tend to be closer to $\bar{x}$ than to μ. Thus, the calculated deviations $(x_i - \bar{x})$ are smaller. As a result, the sample variance obtained is smaller than the population variance. To compensate for this, $(n-1)$ is used as the denominator in place of n in the calculation of sample variance. Thus, the correct formula to obtain the sample variance is:

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

It is important to note that although using $(n-1)$ as the denominator makes the sample variance, s^2, an unbiased estimator of the population variance, σ^2, the sample standard deviation, s, still remains a biased estimator of the population standard deviation, σ. For large sample sizes this bias is negligible.

Degrees of Freedom (dof)

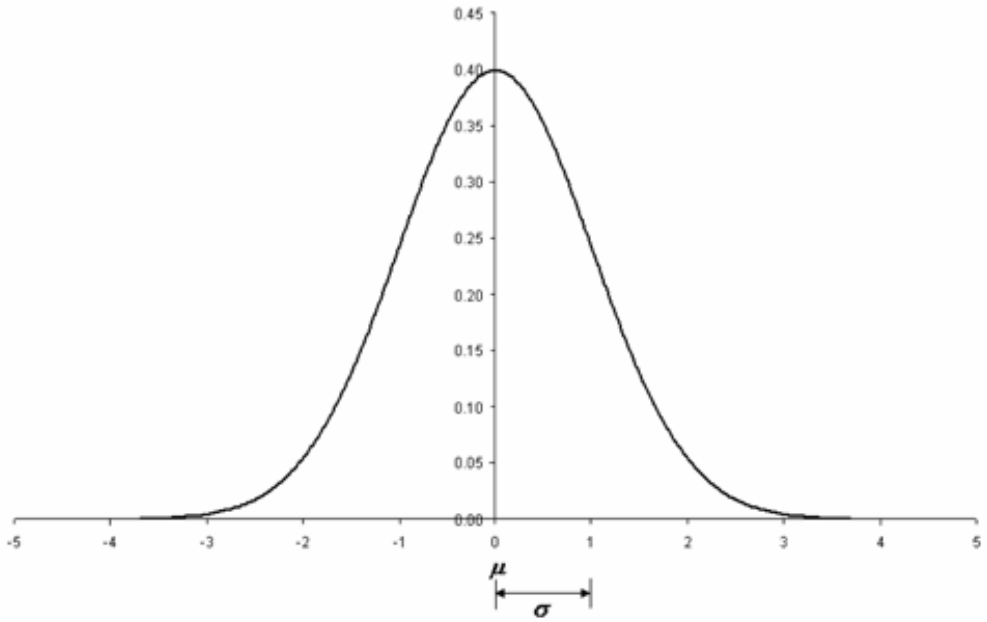
The number of *degrees of freedom* is the number of independent observations made in excess of the unknowns. If there are 3 unknowns and 7 independent observations are taken, then the number of degrees of freedom is 4 (7-3). As another example, two parameters are needed to specify a line. Therefore, there are 2 unknowns. If 10 points are available to fit the line, the number of degrees of freedom is 8 (10-2).

Standard Normal Distribution

A normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$ is called the *standard normal distribution* (see figure below). Standard normal random variables are denoted by Z. If X represents a normal random variable that follows the normal distribution with mean μ and variance σ^2, then the corresponding standard normal random variable is:

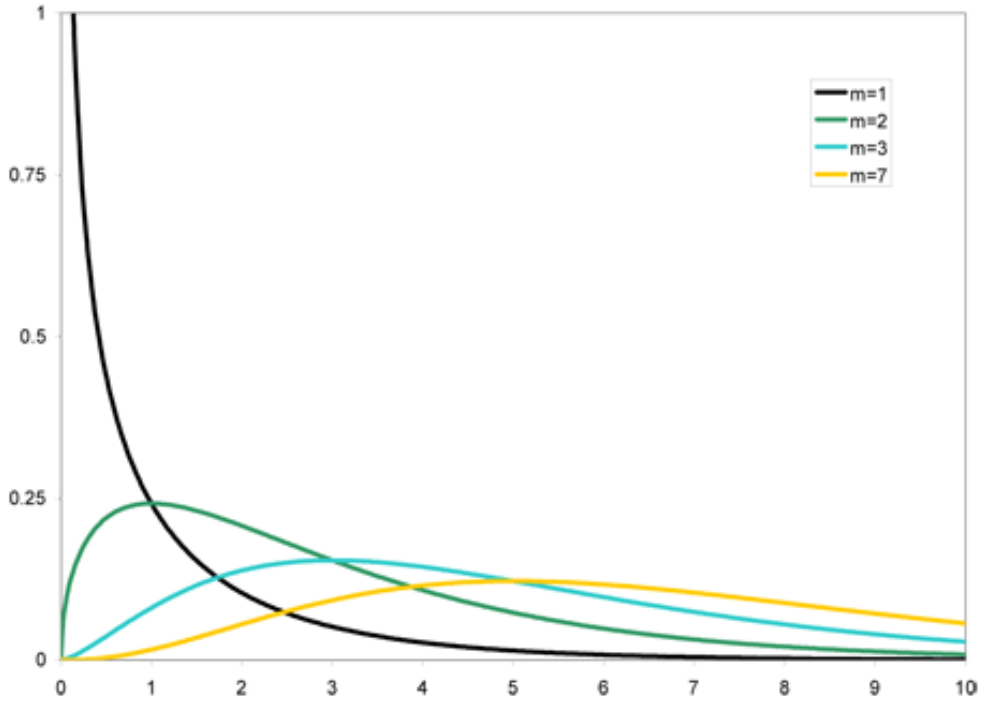
$$Z = (X - \mu)/\sigma$$

Z represents the distance of X from the mean μ in terms of the standard deviation σ.



Chi-Squared Distribution

If Z is a standard normal random variable, then the distribution of Z^2 is a *chi-squared* distribution (see figure below).



A chi-squared random variable is represented by χ^2. Thus:

$$\chi^2 = Z^2$$

The distribution of the variable χ^2 mentioned in the previous equation is also referred to as *centrally distributed chi-squared* with one degree of freedom. The degree of freedom is 1 here because the chi-squared random variable is obtained from a single standard normal random variable Z. The previous equation may also be represented by including the degree of freedom in the equation as:

$$\chi_1^2 = Z^2$$

If Z_1 , Z_2 , Z_3 ... Z_m are m independent standard normal random variables, then:

$$\chi_m^2 = Z_1^2 + Z_2^2 + Z_3^2 \ldots + Z_m^2$$

is also a chi-squared random variable. The distribution of χ_m^2 is said to be *centrally distributed chi-squared* with m degrees of freedom, as the chi-squared random variable is obtained from m independent standard normal random variables. If X is a normal random variable, then the distribution of X^2 is said to be *non-centrally distributed* chi-squared with one degree of freedom. Therefore, X^2 is a chi-squared random variable and can be represented as:

$$\chi_1^2 = X^2$$

If X_1 , X_2 , X_3 ... X_m are m independent normal random variables then:

$$\chi_m^2 = X_1^2 + X_2^2 + X_3^2 \ldots + X_m^2$$

is a non-centrally distributed chi-squared random variable with m degrees of freedom.

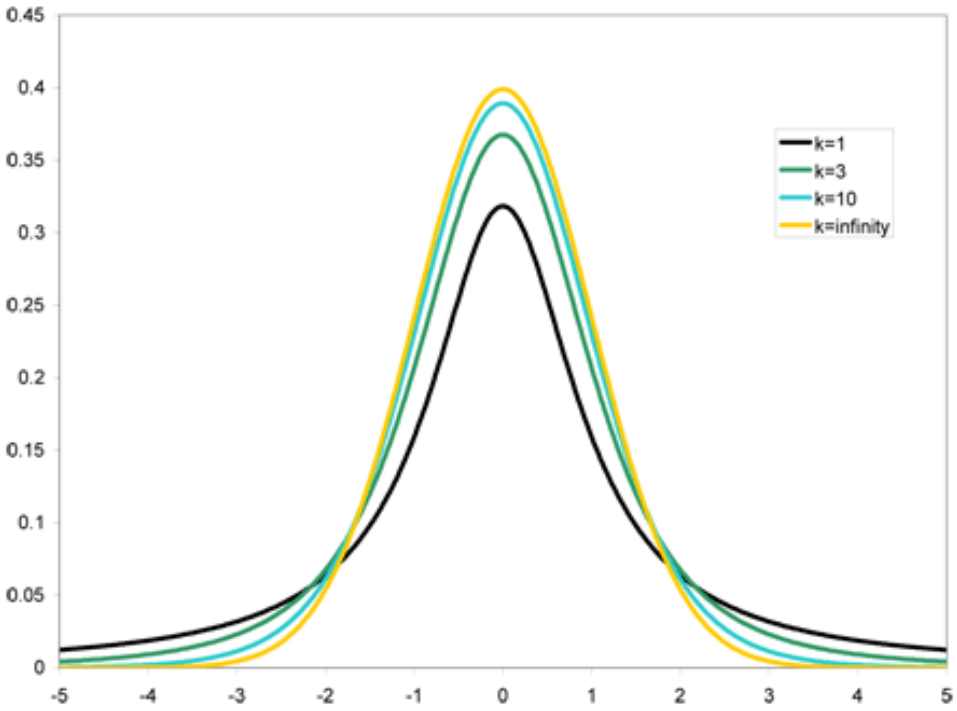
Student's t Distribution (t Distribution)

If Z is a standard normal random variable, χ_k^2 is a chi-squared random variable with k degrees of freedom, and both of these random variables are independent, then the distribution of the random variable T such that:

$$T = \frac{Z}{\sqrt{\chi_k^2/k}}$$

is said to follow the t distribution with k degrees of freedom.

The t distribution is similar in appearance to the standard normal distribution (see figure below). Both of these distributions are symmetric, reaching a maximum at the mean value of zero. However, the t distribution has heavier tails than the standard normal distribution, implying that it has more probability in the tails. As the degrees of freedom, k, of the t distribution approach infinity, the distribution approaches the standard normal distribution.

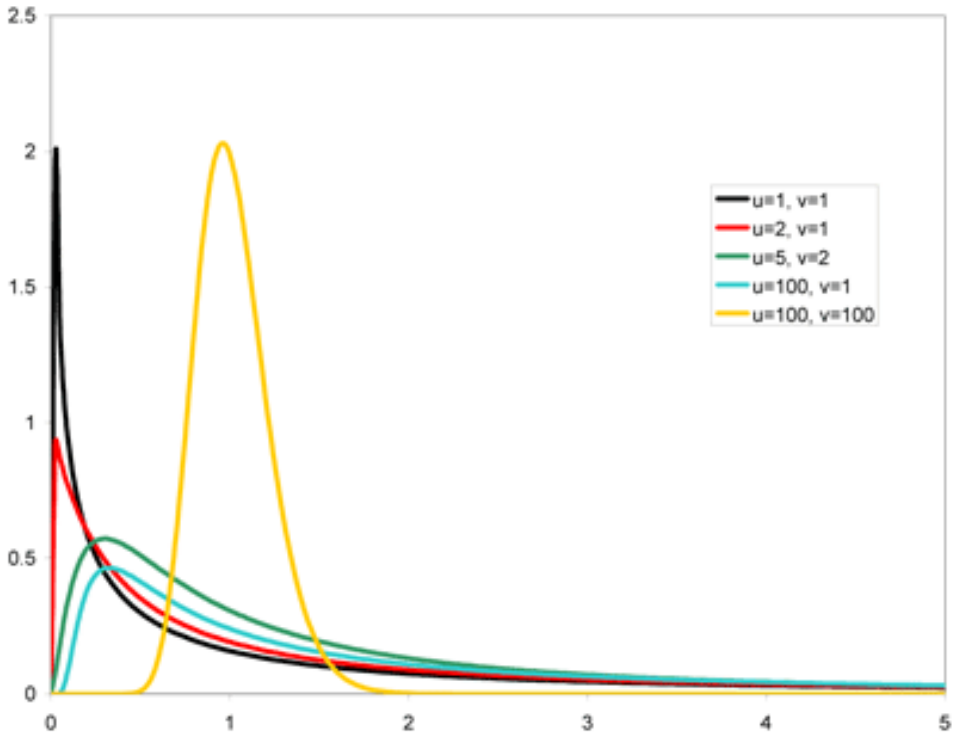


F Distribution

If χ_u^2 and χ_v^2 are two independent chi-squared random variables with u and v degrees of freedom, respectively, then the distribution of the random variable F such that:

$$F = \frac{\chi_u^2/u}{\chi_v^2/v}$$

is said to follow the F distribution with u degrees of freedom in the numerator and v degrees of freedom in the denominator. The F distribution resembles the chi-squared distribution (see the following figure). This is because the F random variable, like the chi-squared random variable, is non-negative and the distribution is skewed to the right (a right skew means that the distribution is unsymmetrical and has a right tail). The F random variable is usually abbreviated by including the degrees of freedom as $F_{u,v}$.



Hypothesis Testing

A statistical hypothesis is a statement about the population under study or about the distribution of a quantity under consideration. The null hypothesis, H_0 , is the hypothesis to be tested. It is a statement about a theory that is believed to be true but has not been proven. For instance, if a

new product design is thought to perform consistently, regardless of the region of operation, then the null hypothesis may be stated as

H_0 : New product design performance is not affected by region

Statements in H_0 always include exact values of parameters under consideration. For example:

H_0 : The population mean is 100

Or simply:

$$H_0 : \mu = 100$$

Rejection of the null hypothesis, H_0 , leads to the possibility that the alternative hypothesis, H_1 , may be true. Given the previous null hypothesis, the alternate hypothesis may be:

H_1 : New product design performance is affected by region

In the case of the example regarding inference on the population mean, the alternative hypothesis may be stated as:

H_1 : The population mean is not 100

Or simply:

$$H_1 : \mu \neq 100$$

Hypothesis testing involves the calculation of a test statistic based on a random sample drawn from the population. The test statistic is then compared to the critical value(s) and used to make a decision about the null hypothesis. The critical values are set by the analyst.

The outcome of a hypothesis test is that we either *reject* H_0 or we *fail to reject* H_0 . Failing to reject H_0 implies that we did not find sufficient evidence to reject H_0 . It does not necessarily mean that there is a high probability that H_0 is true. As such, the terminology *accept* H_0 is not preferred.

For example, assume that an analyst wants to know if the mean of a certain population is 100 or not. The statements for this hypothesis can be stated as follows:

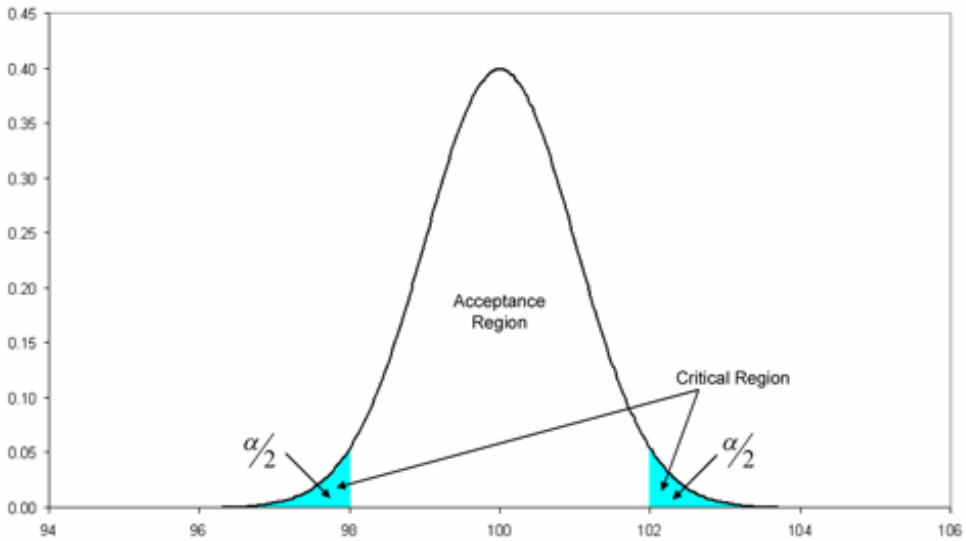
$$H_0 : \quad \mu = 100$$
$$H_1 : \quad \mu \neq 100$$

The analyst decides to use the sample mean as the test statistic for this test. The analyst further decides that if the sample mean lies between 98 and 102 it can be concluded that the population

mean is 100. Thus, the critical values set for this test by the analyst are 98 and 102. It is also decided to draw out a random sample of size 25 from the population.

Now assume that the true population mean is $\mu = 100$ and the true population standard deviation is $\sigma = 5$. This information is not known to the analyst. Using the Central Limit Theorem, the test statistic (sample mean) will follow a normal distribution with a mean equal to the population mean, μ, and a standard deviation of $\sigma/\sqrt{n}$, where n is the sample size. Therefore, the distribution of the test statistic has a mean of 100 and a standard deviation of $5/\sqrt{25} = 1$. This distribution is shown in the figure below.

The unshaded area in the figure bound by the critical values of 98 and 102 is called the *acceptance region*. The acceptance region gives the probability that a random sample drawn from the population would have a sample mean that lies between 98 and 102. Therefore, this is the region that will lead to the "acceptance" of H_0. On the other hand, the shaded area gives the probability that the sample mean obtained from the random sample lies outside of the critical values. In other words, it gives the probability of rejection of the null hypothesis when the true mean is 100. The shaded area is referred to as the critical region or the rejection region. Rejection of the null hypothesis H_0 when it is true is referred to as type I error. Thus, there is a 4.56% chance of making a type I error in this hypothesis test. This percentage is called the significance level of the test and is denoted by α. Here $\alpha = 0.0456$ or 4.56% (area of the shaded region in the figure). The value of α is set by the analyst when he/she chooses the critical values.



A type II error is also defined in hypothesis testing. This error occurs when the analyst fails to reject the null hypothesis when it is actually false. Such an error would occur if the value of the sample mean obtained is in the acceptance region bounded by 98 and 102 even though the true population mean is not 100. The probability of occurrence of type II error is denoted by β.

Two-sided and One-sided Hypotheses

As seen in the previous section, the critical region for the hypothesis test is split into two parts, with equal areas in each tail of the distribution of the test statistic. Such a hypothesis, in which the values for which we can reject H_0 are in both tails of the probability distribution, is called a two-sided hypothesis. The hypothesis for which the critical region lies only in one tail of the probability distribution is called a one-sided hypothesis. For instance, consider the following hypothesis test:

$$H_0: \quad \mu = 100$$
$$H_1: \quad \mu > 100$$

This is an example of a one-sided hypothesis. Here the critical region lies entirely in the right tail of the distribution.

The hypothesis test may also be set up as follows:

$$H_0: \quad \mu = 100$$
$$H_1: \quad \mu < 100$$

This is also a one-sided hypothesis. Here the critical region lies entirely in the left tail of the distribution.

Statistical Inference for a Single Sample

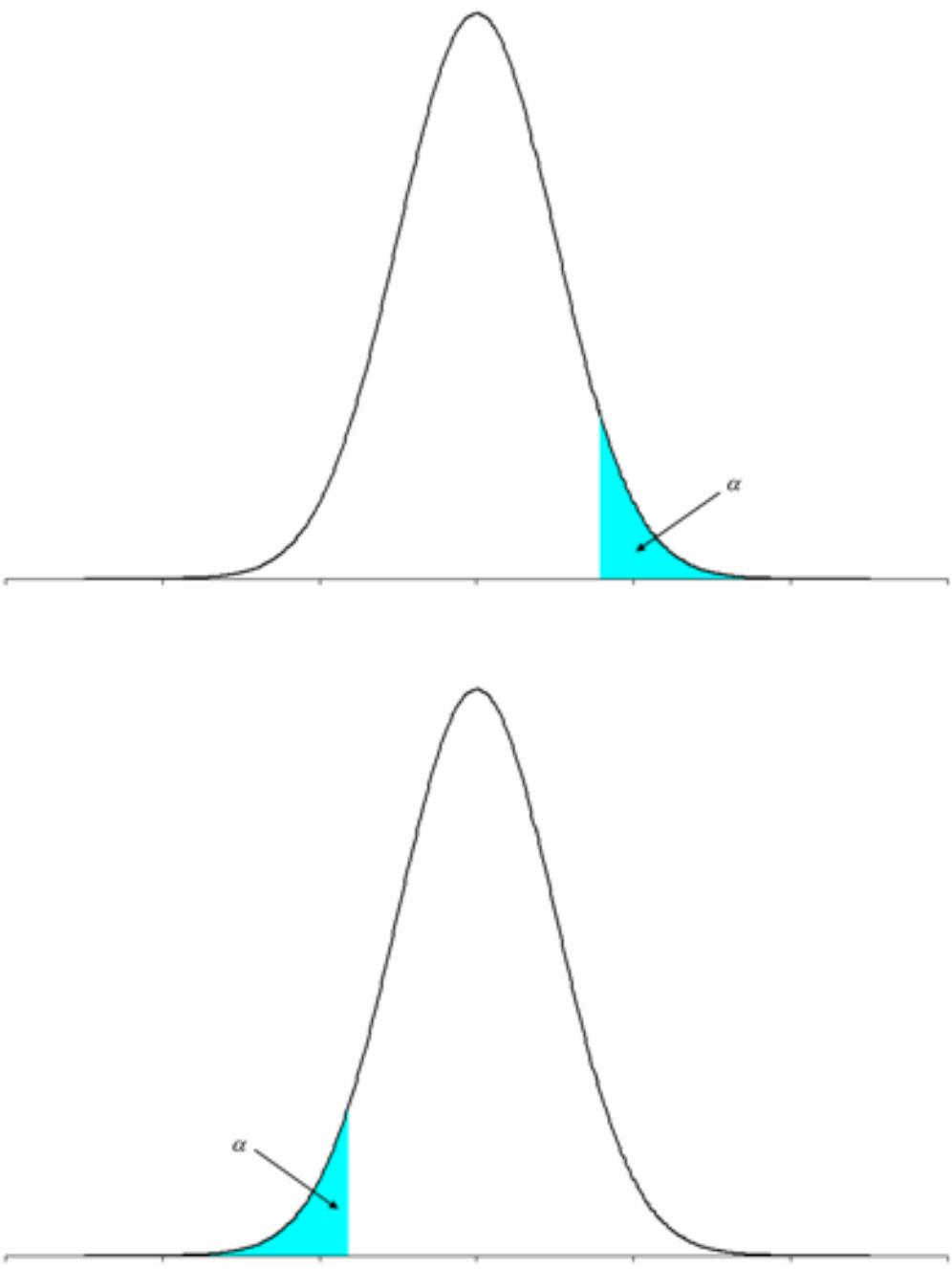
Hypothesis testing forms an important part of statistical inference. As stated previously, statistical inference refers to the process of estimating results for the population based on measurements from a sample. In the next sections, statistical inference for a single sample is discussed briefly.

Inference on the Mean of a Population When the Variance Is Known

The test statistic used in this case is based on the standard normal distribution. If $\bar{X}$ is the calculated sample mean, then the standard normal test statistic is:

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

where μ_0 is the hypothesized population mean, σ is the population standard deviation and n is the sample size.



For example, assume that an analyst wants to know if the mean of a population, μ, is 100. The population variance, σ^2, is known to be 25. The hypothesis test may be conducted as follows:

1) The statements for this hypothesis test may be formulated as:

H_0 : $\mu = 100$
H_1 : $\mu \neq 100$

It is a clear that this is a two-sided hypothesis. Thus the critical region will lie in both of the tails of the probability distribution.

2) Assume that the analyst chooses a significance level of 0.05. Thus $\alpha = 0.05$. The significance level determines the critical values of the test statistic. Here the test statistic is based on the standard normal distribution. For the two-sided hypothesis these values are obtained as:

$$z_{\alpha/2} = z_{0.025} = 1.96$$

and

$$-z_{\alpha/2} = -z_{0.025} = -1.96$$

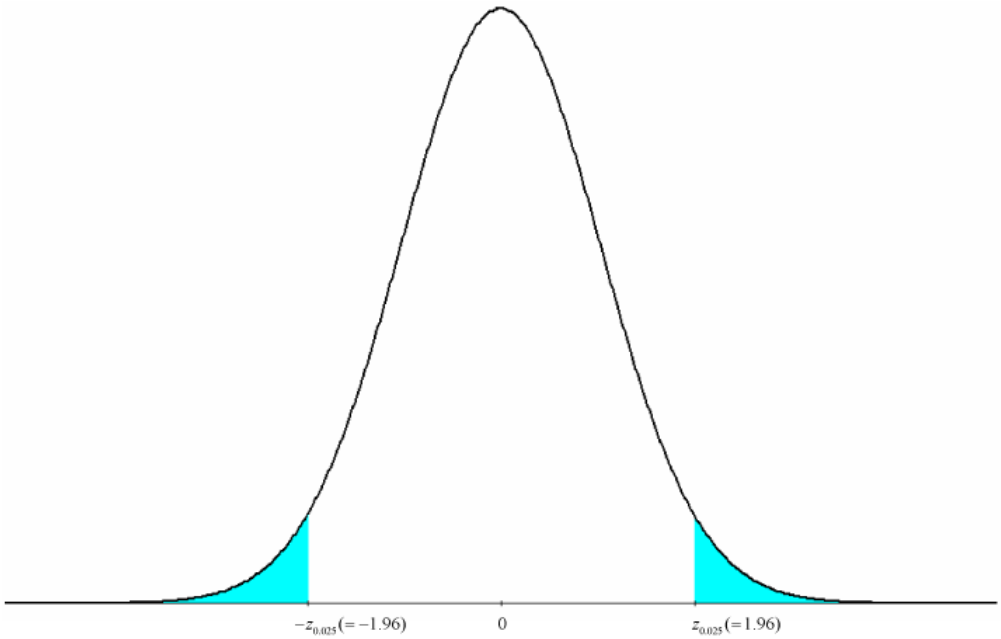
These values and the critical regions are shown in figure below. The analyst would fail to reject H_0 if the test statistic, Z_0 , is such that:

$$-z_{\alpha/2} \leq Z_0 \leq z_{\alpha/2}$$

or

$$-1.96 \leq Z_0 \leq 1.96$$

3) Next the analyst draws a random sample from the population. Assume that the sample size, n , is 25 and the sample mean is obtained as $\bar{x} = 103$.



4) The value of the test statistic corresponding to the sample mean value of 103 is:

$$z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{103 - 100}{5/\sqrt{25}} = 3$$

Since this value does not lie in the acceptance region $-1.96 \leq Z_0 \leq 1.96$, we reject $H_0 : \mu = 100$ at a significance level of 0.05.

P Value

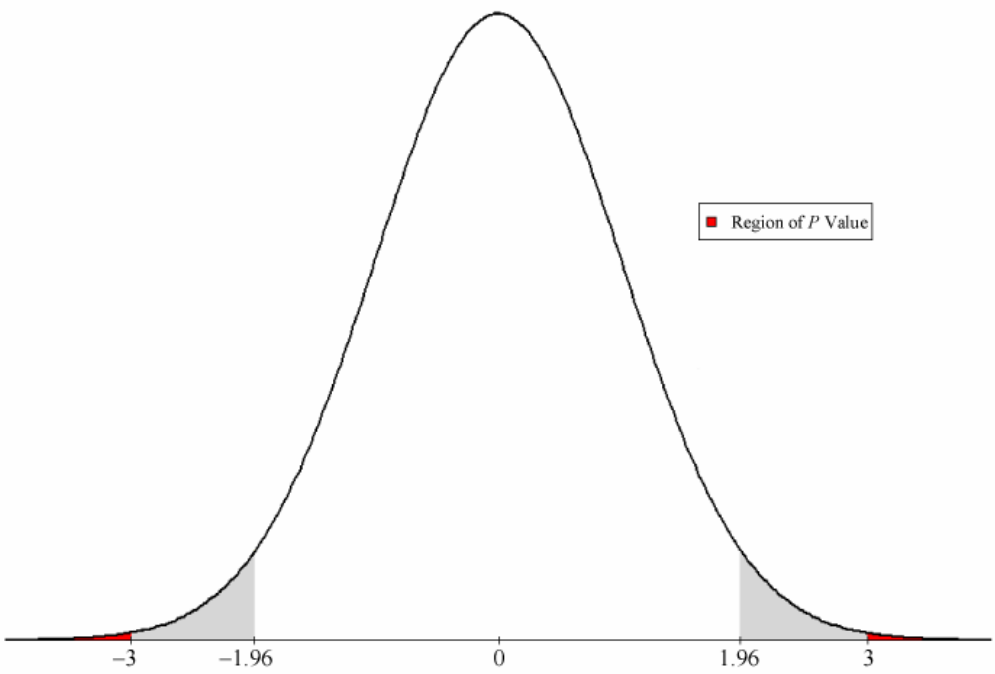
In the previous example the null hypothesis was rejected at a significance level of 0.05. This statement does not provide information as to how far out the test statistic was into the critical region. At times it is necessary to know if the test statistic was just into the critical region or was far out into the region. This information can be provided by using the p value.

The p value is the probability of occurrence of the values of the test statistic that are either equal to the one obtained from the sample or more unfavorable to H_0 than the one obtained from the sample. It is the lowest significance level that would lead to the rejection of the null hypothesis, H_0, at the given value of the test statistic. The value of the test statistic is referred to as significant when H_0 is rejected. The p value is the smallest α at which the statistic is significant and H_0 is rejected.

For instance, in the previous example the test statistic was obtained as $z_0 = 3$. Values that are more unfavorable to H_0 in this case are values greater than 3. Then the required probability is the probability of getting a test statistic value either equal to or greater than 3 (this is abbreviated as $P(Z \geq 3)$). This probability is shown in figure below as the dark shaded area on the right tail of the distribution and is equal to 0.0013 or 0.13% (i.e., $P(Z \geq 3) = 0.0013$). Since this is a two-sided test the p value is:

$$p\,value = 2 \times 0.0013 = 0.0026$$

Therefore, the smallest α (corresponding to the test static value of 3) that would lead to the rejection of H_0 is 0.0026.



Inference on Mean of a Population When Variance Is Unknown

When the variance, σ^2, of a population (that can be assumed to be normally distributed) is unknown the sample variance, S^2, is used in its place in the calculation of the test statistic. The test statistic used in this case is based on the t distribution and is obtained using the following relation:

$$T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

The test statistic follows the t distribution with $n-1$ degrees of freedom.

For example, assume that an analyst wants to know if the mean of a population, μ, is less than 50 at a significance level of 0.05. A random sample drawn from the population gives the sample mean, $\bar{x}$, as 47.7 and the sample standard deviation, s, as 5. The sample size, n, is 25. The hypothesis test may be conducted as follows:

1) The statements for this hypothesis test may be formulated as:

$$H_0: \quad \mu = 50$$
$$H_1: \quad \mu < 50$$

It is clear that this is a one-sided hypothesis. Here the critical region will lie in the left tail of the probability distribution.

2) Significance level, $\alpha = 0.05$. Here, the test statistic is based on the t distribution. Thus, for the one-sided hypothesis the critical value is obtained as:

$$-t_{\alpha,dof} = -t_{0.05,n-1} = -t_{0.05,24} = -1.7109$$

This value and the critical regions are shown in the figure below. The analyst would fail to reject H_0 if the test statistic T_0 is such that:

$$T_0 > -t_{0.05,24}$$

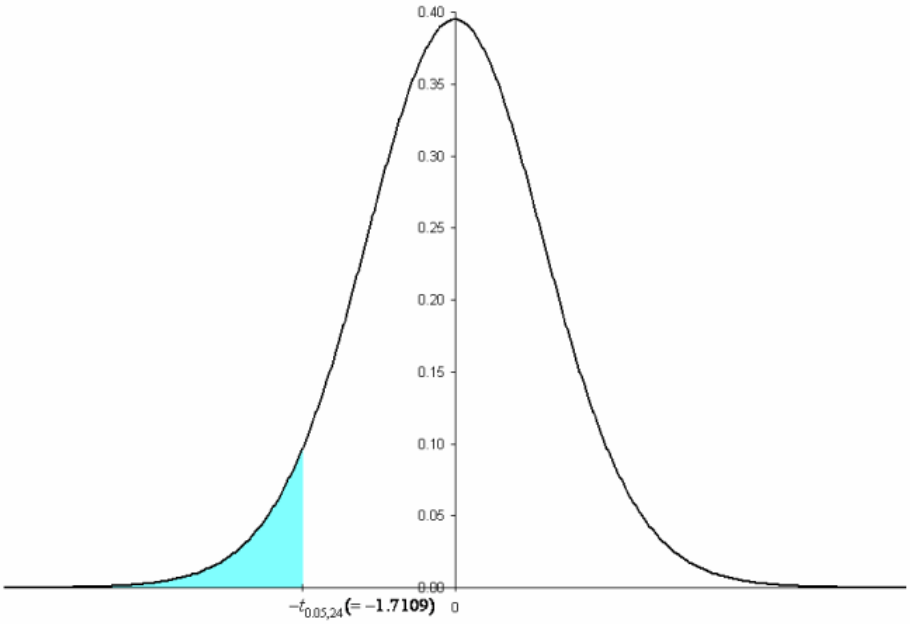
3) The value of the test statistic, T_0 , corresponding to the given sample data is:

$$\begin{aligned} t_0 &= \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \\ &= \frac{47.7 - 50}{5/\sqrt{25}} \\ &= -2.3 \end{aligned}$$

Since T_0 is less than the critical value of -1.7109, $H_0 : \mu = 50$ is rejected and it is concluded that at a significance level of 0.05 the population mean is less than 50.

4) P value

In this case the p value is the probability that the test statistic is either less than or equal to -2.3 (since values less than -2.3 are unfavorable to H_0). This probability is equal to 0.0152.



Inference on Variance of a Normal Population

The test statistic used in this case is based on the chi-squared distribution. If S^2 is the calculated sample variance and σ_0^2 the hypothesized population variance then the Chi-Squared test statistic is:

$$\chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

The test statistic follows the chi-squared distribution with $n-1$ degrees of freedom.

For example, assume that an analyst wants to know if the variance of a population exceeds 1 at a significance level of 0.05. A random sample drawn from the population gives the sample variance as 2. The sample size, n, is 20. The hypothesis test may be conducted as follows:

1) The statements for this hypothesis test may be formulated as:

$$H_0: \quad \sigma^2 = 1$$
$$H_1: \quad \sigma^2 > 1$$

This is a one-sided hypothesis. Here the critical region will lie in the right tail of the probability distribution.

2) Significance level, $\alpha = 0.05$. Here, the test statistic is based on the chi-squared distribution. Thus for the one-sided hypothesis the critical value is obtained as:

$$\chi^2_{\alpha,n-1} = \chi^2_{0.05,19} = 30.1435$$

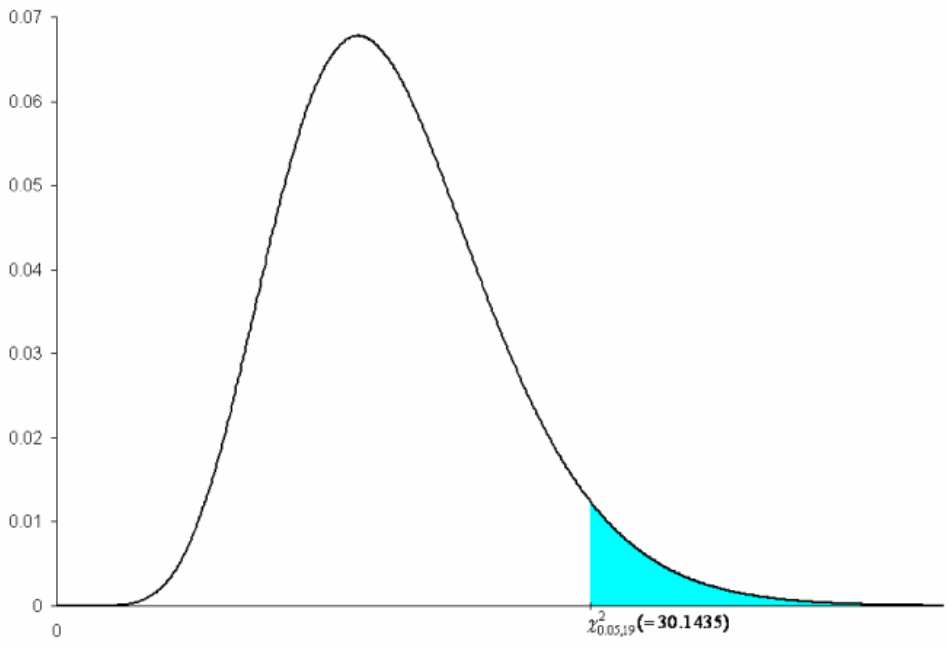
This value and the critical regions are shown in the figure below. The analyst would fail to reject H_0 if the test statistic χ_0^2 is such that:

$$\chi_0^2 < \chi^2_{0.05,19}$$

3) The value of the test statistic χ_0^2 corresponding to the given sample data is:

$$\chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2}$$
$$= \frac{(20-1)2}{1} = 38$$

Since χ_0^2 is greater than the critical value of 30.1435, $H_0: \sigma^2 = 1$ is rejected and it is concluded that at a significance level of 0.05 the population variance exceeds 1.



4) P value

In this case the p value is the probability that the test statistic is greater than or equal to 38 (since values greater than 38 are unfavorable to H_0). This probability is determined to be 0.0059.

Statistical Inference for Two Samples

Inference on the Difference in Population Means When Variances Are Known

The test statistic used here is based on the standard normal distribution. Let μ_1 and μ_2 represent the means of two populations, and σ_1^2 and σ_2^2 their variances, respectively. Let Δ_0 be the hypothesized difference in the population means and $\bar{X}_1$ and $\bar{X}_2$ be the sample means obtained from two samples of sizes n_1 and n_2 drawn randomly from the two populations, respectively. The test statistic can be obtained as:

$$Z_0 = \frac{\bar{X}_1 - \bar{X}_2 - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

The statements for the hypothesis test are:

$$H_0: \quad \mu_1 - \mu_2 = \Delta_0$$
$$H_1: \quad \mu_1 - \mu_2 \neq \Delta_0$$

If $\Delta_0 = 0$, then the hypothesis will test for the equality of the two population means.

Inference on the Difference in Population Means When Variances Are Unknown

If the population variances can be assumed to be equal then the following test statistic based on the t distribution can be used. Let $\bar{X}_1$, $\bar{X}_2$, S_1^2 and S_2^2 be the sample means and variances obtained from randomly drawn samples of sizes n_1 and n_2 from the two populations, respectively. The weighted average, S_p^2 , of the two sample variances is:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

S_p^2 has (n_1 + n_2 -- 2) degrees of freedom. The test statistic can be calculated as:

$$T_0 = \frac{\bar{X}_1 - \bar{X}_2 - \Delta_0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

T_0 follows the t distribution with (n_1 + n_2 -- 2) degrees of freedom. This test is also referred to as the two-sample pooled t test. If the population variances cannot be assumed to be equal then the following test statistic is used:

$$T_0^* = \frac{\bar{X}_1 - \bar{X}_2 - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

T_0^* follows the t distribution with v degrees of freedom. v is defined as follows:

$$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{(S_1^2/n_1)^2}{n_1+1} + \frac{(S_2^2/n_2)^2}{n_2+1}} - 2$$

Inference on the Variances of Two Normal Populations

The test statistic used here is based on the F distribution. If S_1^2 and S_2^2 are the sample variances drawn randomly from the two populations and n_1 and n_2 are the two sample sizes, respectively, then the test statistic that can be used to test the equality of the population variances is:

$$F_0 = \frac{S_1^2}{S_2^2}$$

The test statistic follows the F distribution with (n_1 -- 1) degrees of freedom in the numerator and (n_2 -- 1) degrees of freedom in the denominator.

For example, assume that an analyst wants to know if the variances of two normal populations are equal at a significance level of 0.05. Random samples drawn from the two populations give the sample standard deviations as 1.84 and 2, respectively. Both the sample sizes are 20. The hypothesis test may be conducted as follows:

1) The statements for this hypothesis test may be formulated as:

$$H_0 : \quad \sigma_1^2 = \sigma_2^2$$
$$H_1 : \quad \sigma_1^2 \neq \sigma_2^2$$

It is clear that this is a two-sided hypothesis and the critical region will be located on both sides of the probability distribution.

2) Significance level $\alpha = 0.05$. Here the test statistic is based on the F distribution. For the two-sided hypothesis the critical values are obtained as:

$$f_{\alpha/2,n_1-1,n_2-1} = f_{0.025,19,19} = 2.53$$

and

$$f_{1-\alpha/2,n_1-1,n_2-1} = f_{0.975,19,19} = 0.40$$

These values and the critical regions are shown in the figure below. The analyst would fail to reject H_0 if the test statistic F_0 is such that:

$$f_{1-\alpha/2,n_1-1,n_2-1} \leq F_0 \leq f_{\alpha/2,n_1-1,n_2-1}$$

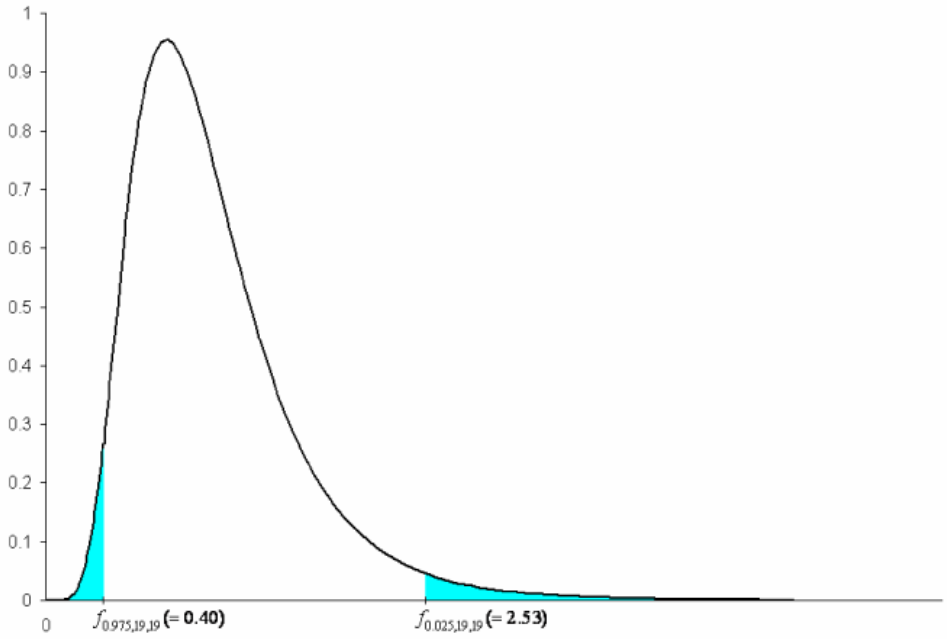
or

$$0.40 \leq F_0 \leq 2.53$$

3) The value of the test statistic F_0 corresponding to the given data is:

$$\begin{aligned} F_0 &= \frac{S_1^2}{S_2^2} \\ &= \frac{1.84^2}{2^2} \\ &= 0.8464 \end{aligned}$$

Since F_0 lies in the acceptance region, the analyst fails to reject $H_0 : \sigma_1^2 = \sigma_2^2$ at a significance level of 0.05.


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
$f_{0.975,19,19}$ (= 0.40)
$f_{0.025,19,19}$ (= 2.53)

Simple Linear Regression Analysis

IN THIS CHAPTER

Simple Linear Regression Analysis **32**
Fitted Regression Line 35

Hypothesis Tests in Simple Linear Regression **38**
t Tests 38
Analysis of Variance Approach to Test the Significance of Regression 42

Confidence Intervals in Simple Linear Regression **48**
Confidence Interval on Regression Coefficients 48
Confidence Interval on Fitted Values 48
Confidence Interval on New Observations 49

Measures of Model Adequacy **51**
Coefficient of Determination (R^2) 51
Residual Analysis 53
Lack-of-Fit Test 57

Transformations **63**

Regression analysis is a statistical technique that attempts to explore and model the relationship between two or more variables. For example, an analyst may want to know if there is a relationship between road accidents and the age of the driver. Regression analysis forms an important part of the statistical analysis of the data obtained from designed experiments and is discussed briefly in this chapter. Every experiment analyzed in a Weibull++ DOE foilo includes regression results for each of the responses. These results, along with the results from the analysis of variance (explained in the One Factor Designs and General Full Factorial Designs chapters), provide information that is useful to identify significant factors in an experiment and explore the nature of the relationship between these factors and the response. Regression analysis forms the basis for all Weibull++ DOE folio calculations related to the sum of squares used in the analysis of variance. The reason for this is explained in Appendix B. Additionally, DOE folios also include a regression tool to see if two or more variables are related, and to explore the nature of the relationship between them.

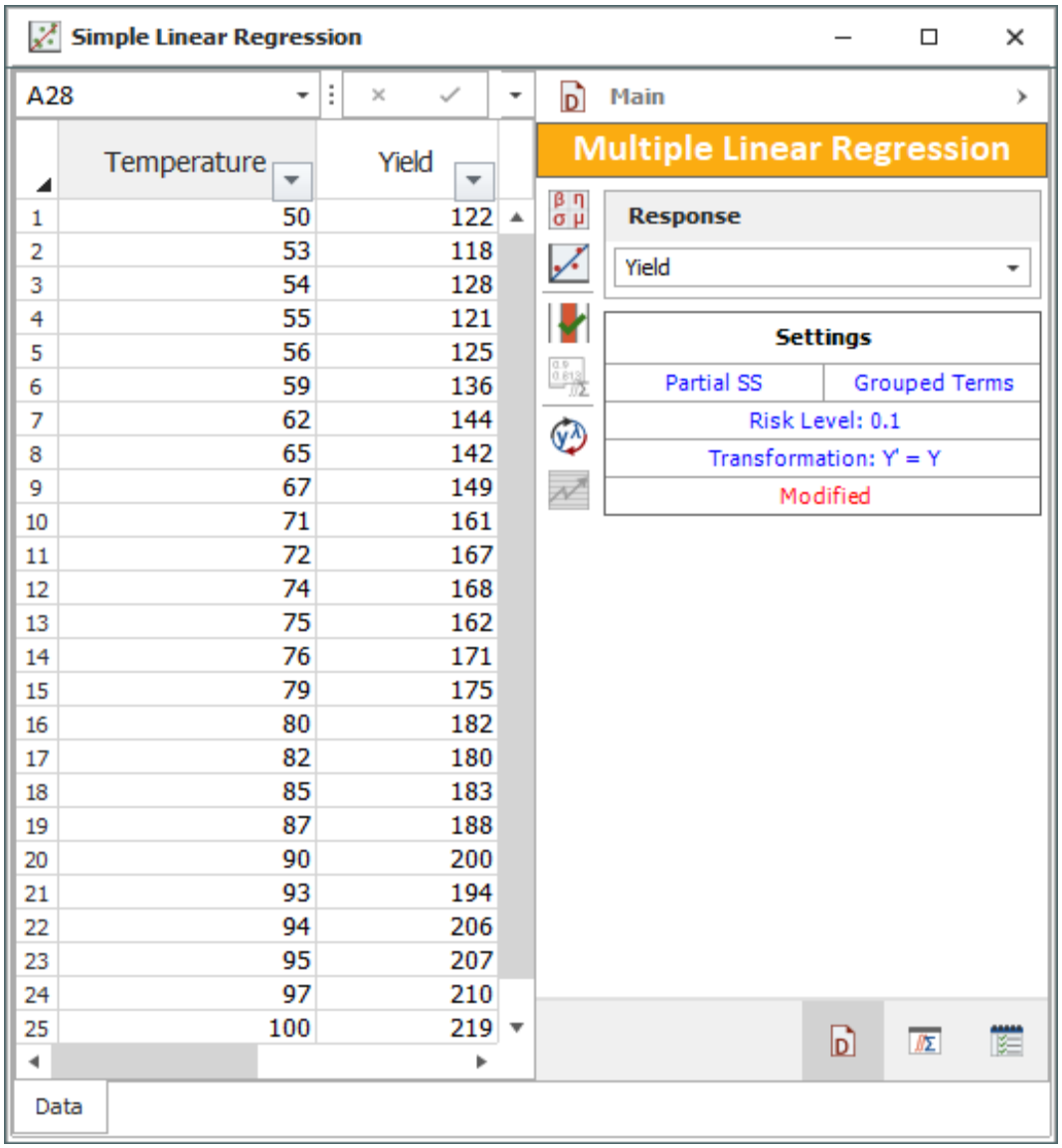
This chapter discusses simple linear regression analysis while a subsequent chapter focuses on multiple linear regression analysis.

Simple Linear Regression Analysis

A linear regression model attempts to explain the relationship between two or more variables using a straight line. Consider the data obtained from a chemical process where the yield of the process is thought to be related to the reaction temperature (see the table below).

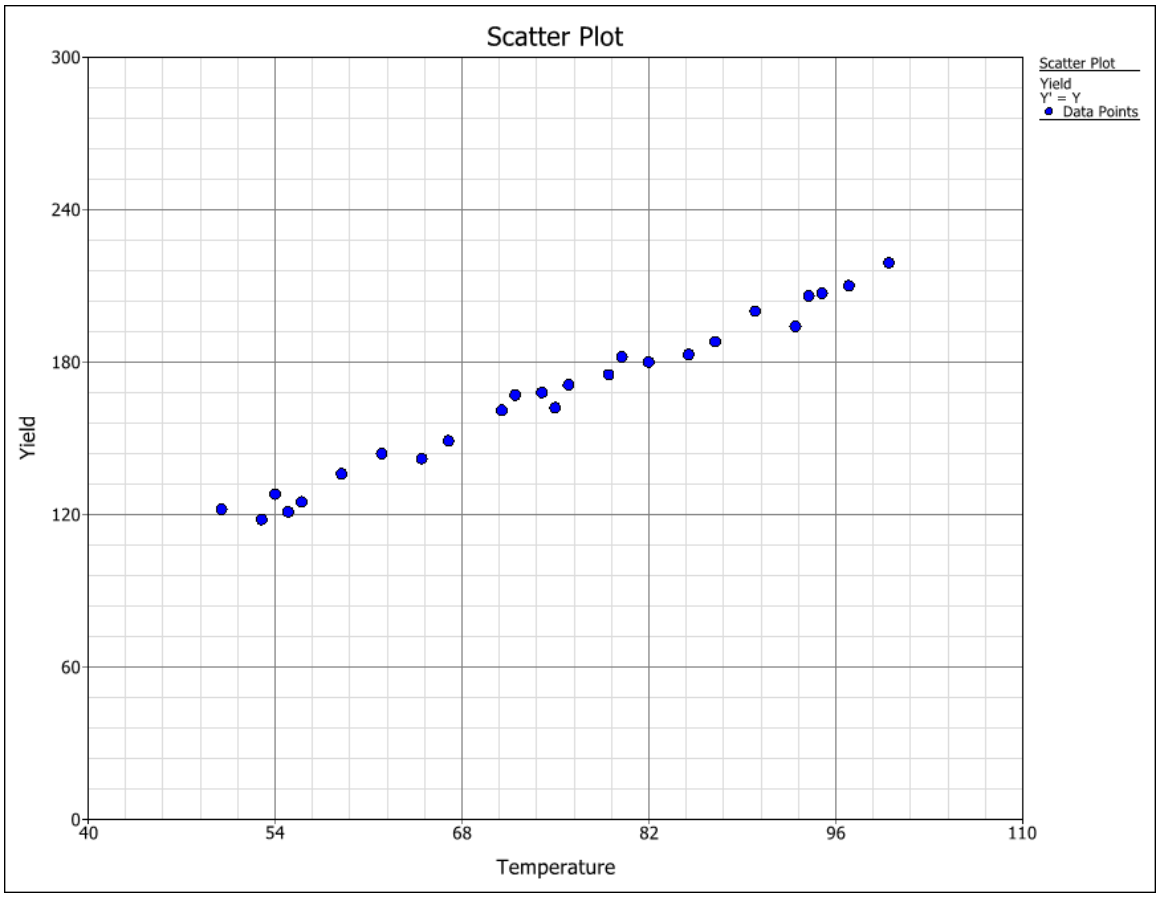
Observation Number	Temperature (x_i)	Yield (y_i)
1	50	122
2	53	118
3	54	128
4	55	121
5	56	125
6	59	136
7	62	144
8	65	142
9	67	149
10	71	161
11	72	167
12	74	168
13	75	162
14	76	171
15	79	175
16	80	182
17	82	180
18	85	183
19	87	188
20	90	200
21	93	194
22	94	206
23	95	207
24	97	210
25	100	219

This data can be entered in the DOE folio as shown in the following figure:



	Temperature	Yield
1	50	122
2	53	118
3	54	128
4	55	121
5	56	125
6	59	136
7	62	144
8	65	142
9	67	149
10	71	161
11	72	167
12	74	168
13	75	162
14	76	171
15	79	175
16	80	182
17	82	180
18	85	183
19	87	188
20	90	200
21	93	194
22	94	206
23	95	207
24	97	210
25	100	219

And a scatter plot can be obtained as shown in the following figure. In the scatter plot yield, y_i is plotted for different temperature values, x_i .



It is clear that no line can be found to pass through all points of the plot. Thus no functional relation exists between the two variables x and Y. However, the scatter plot does give an indication that a straight line may exist such that all the points on the plot are scattered randomly around this line. A statistical relation is said to exist in this case. The statistical relation between x and Y may be expressed as follows:

$$Y = \beta_0 + \beta_1 x + \epsilon$$

The above equation is the linear regression model that can be used to explain the relation between x and Y that is seen on the scatter plot above. In this model, the mean value of Y (abbreviated as $E(Y)$) is assumed to follow the linear relation:

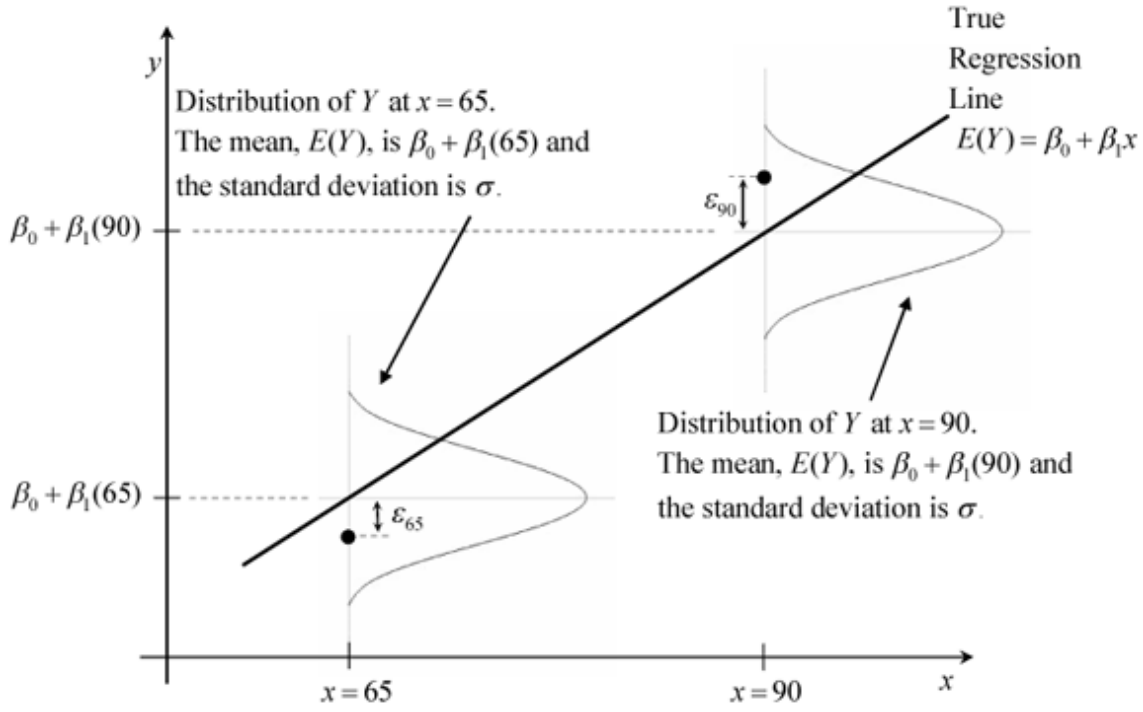
$$E(Y) = \beta_0 + \beta_1 x$$

The actual values of Y (which are observed as yield from the chemical process from time to time and are random in nature) are assumed to be the sum of the mean value, $E(Y)$, and a random error term, ϵ:

$$\begin{aligned} Y &= E(Y) + \epsilon \\ &= \beta_0 + \beta_1 x + \epsilon \end{aligned}$$

The regression model here is called a *simple* linear regression model because there is just one independent variable, x, in the model. In regression models, the independent variables are also referred to as regressors or predictor variables. The dependent variable, Y, is also referred to as the response. The slope, β_1, and the intercept, β_0, of the line $E(Y) = \beta_0 + \beta_1 x$ are called *regression coefficients*. The slope, β_1, can be interpreted as the change in the mean value of Y for a unit change in x.

The random error term, ϵ, is assumed to follow the normal distribution with a mean of 0 and variance of σ^2. Since Y is the sum of this random term and the mean value, $E(Y)$, which is a constant, the variance of Y at any given value of x is also σ^2. Therefore, at any given value of x, say x_i, the dependent variable Y follows a normal distribution with a mean of $\beta_0 + \beta_1 x_i$ and a standard deviation of σ. This is illustrated in the following figure.



Fitted Regression Line

The true regression line is usually not known. However, the regression line can be estimated by estimating the coefficients β_1 and β_0 for an observed data set. The estimates, $\hat{\beta}_1$ and $\hat{\beta}_0$, are calculated using least squares. (For details on least square estimates, refer to Hahn & Shapiro (1967).) The estimated regression line, obtained using the values of $\hat{\beta}_1$ and $\hat{\beta}_0$, is called the *fitted line*. The least square estimates, $\hat{\beta}_1$ and $\hat{\beta}_0$, are obtained using the following equations:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i - \frac{(\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} x_i)}{n}}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

where $\bar{y}$ is the mean of all the observed values and $\bar{x}$ is the mean of all values of the predictor variable at which the observations were taken. $\bar{y}$ is calculated using $\bar{y} = (1/n) \sum)_{i=1}^{n} y_i$ and $\bar{x}$ is calculated using $\bar{x} = (1/n) \sum)_{i=1}^{n} x_i$.

Once $\hat{\beta}_1$ and $\hat{\beta}_0$ are known, the fitted regression line can be written as:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

where $\hat{y}$ is the fitted or estimated value based on the fitted regression model. It is an estimate of the mean value, $E(Y)$. The fitted value,$\hat{y}_i$, for a given value of the predictor variable, x_i , may be different from the corresponding observed value, y_i . The difference between the two values is called the *residual*, e_i :

$$e_i = y_i - \hat{y}_i$$

Calculation of the Fitted Line Using Least Square Estimates

The least square estimates of the regression coefficients can be obtained for the data in the preceding table as follows:

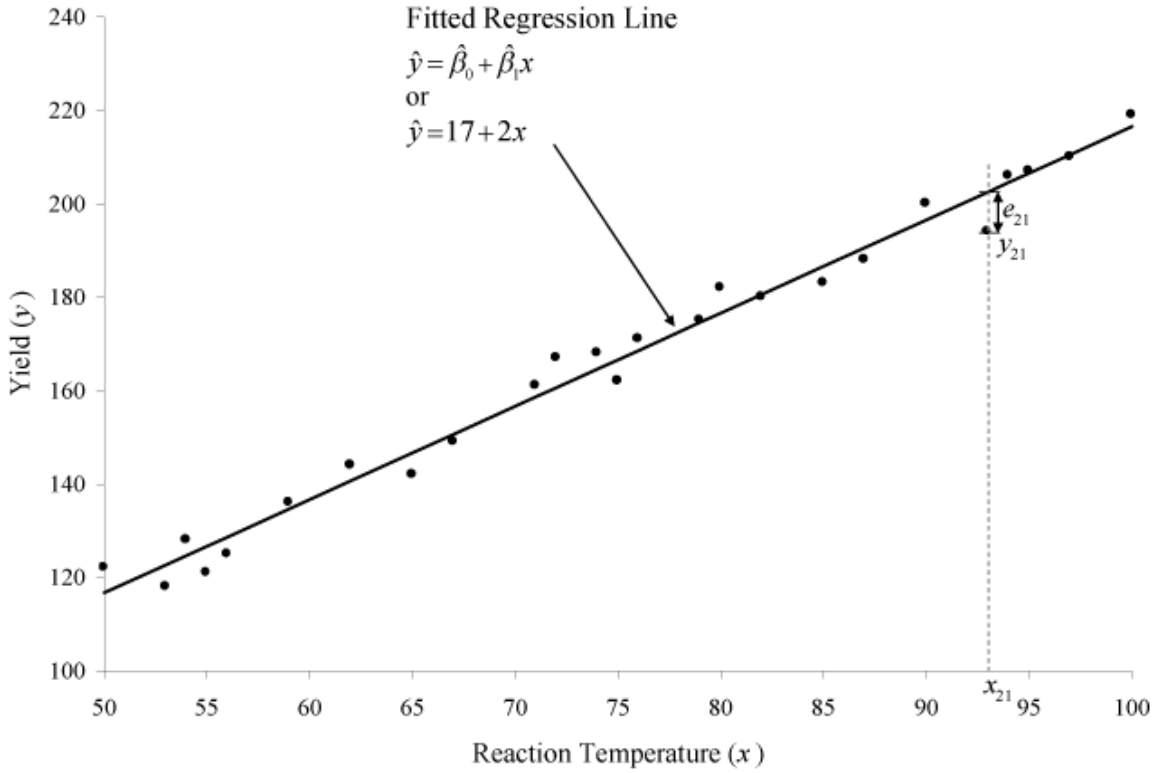
$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} y_i x_i - \frac{(\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} x_i)}{n}}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \\ &= \frac{322516 - \frac{4158 x 1871}{25}}{5697.36} \\ &= 1.9952 \approx 2.00\end{aligned}$$

$$\begin{aligned}\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ &= 166.32 - 2x74.84 \\ &= 17.0016 \approx 17.00\end{aligned}$$

Knowing $\hat{\beta}_0$ and $\hat{\beta}_1$, the fitted regression line is:

$$\begin{aligned}\hat{y} &= \hat{\beta}_0 + \hat{\beta}_1 x \\ &= 17.0016 + 1.9952 \times x \\ &\approx 17 + 2x\end{aligned}$$

This line is shown in the figure below.



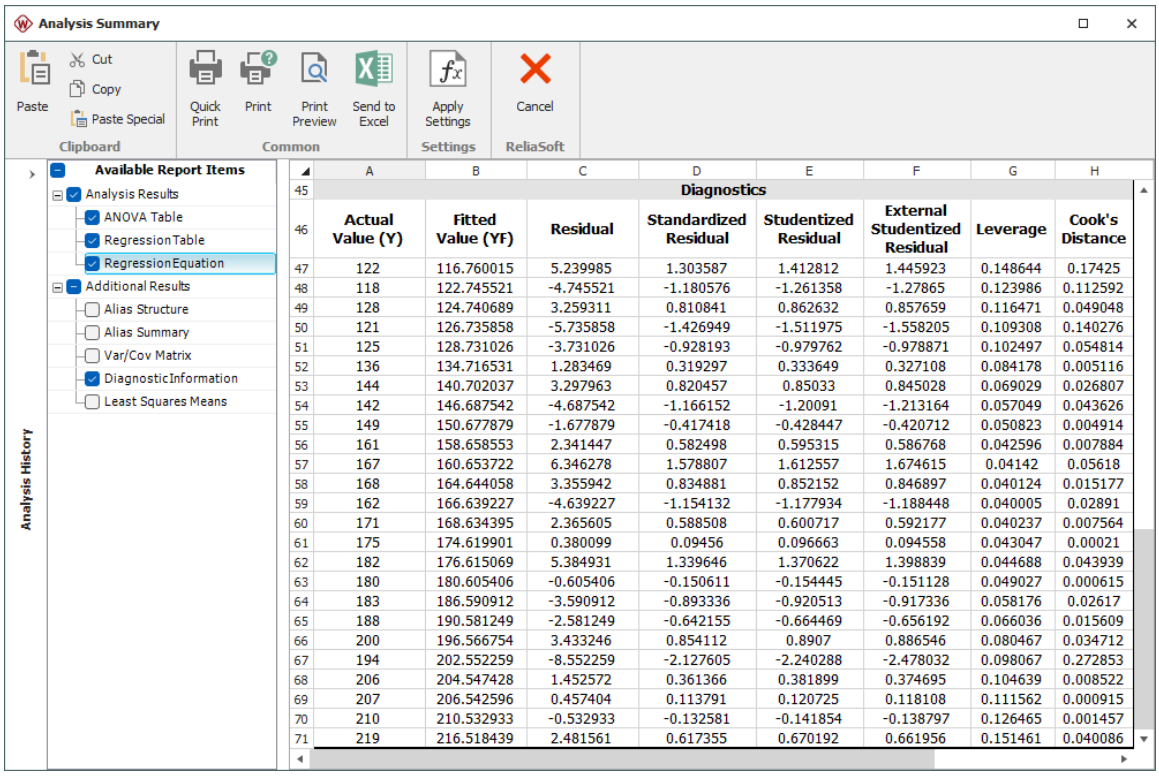
Once the fitted regression line is known, the fitted value of $\boldsymbol{Y}$ corresponding to any observed data point can be calculated. For example, the fitted value corresponding to the 21st observation in the preceding table is:

$$\begin{aligned}\hat{y}_{21} &= \hat{\beta}_0 + \hat{\beta}_1 x_{21} \\ &= (17.0016) + (1.9952) \times 93 \\ &= 202.6\end{aligned}$$

The observed response at this point is $y_{21} = 194$. Therefore, the residual at this point is:

$$\begin{aligned}e_{21} &= y_{21} - \hat{y}_{21} \\ &= 194 - 202.6 \\ &= -8.6\end{aligned}$$

In DOE folios, fitted values and residuals can be calculated. The values are shown in the figure below.



Diagnostics

	Actual Value (Y)	Fitted Value (YF)	Residual	Standardized Residual	Studentized Residual	External Studentized Residual	Leverage	Cook's Distance
47	122	116.760015	5.239985	1.303587	1.412812	1.445923	0.148644	0.17425
48	118	122.745521	-4.745521	-1.180576	-1.261358	-1.27865	0.123986	0.112592
49	128	124.740689	3.259311	0.810841	0.862632	0.857659	0.116471	0.049048
50	121	126.735858	-5.735858	-1.426949	-1.511975	-1.558205	0.109308	0.140276
51	125	128.731026	-3.731026	-0.928193	-0.979762	-0.978871	0.102497	0.054814
52	136	134.716531	1.283469	0.319297	0.333649	0.327108	0.084178	0.005116
53	144	140.702037	3.297963	0.820457	0.85033	0.845028	0.069029	0.026807
54	142	146.687542	-4.687542	-1.166152	-1.20091	-1.213164	0.057049	0.043626
55	149	150.677879	-1.677879	-0.417418	-0.428447	-0.420712	0.050823	0.004914
56	161	158.658553	2.341447	0.582498	0.595315	0.586768	0.042596	0.007884
57	167	160.653722	6.346278	1.578807	1.612557	1.674615	0.04142	0.05618
58	168	164.644058	3.355942	0.834881	0.852152	0.846897	0.040124	0.015177
59	162	166.639227	-4.639227	-1.154132	-1.177934	-1.188448	0.040005	0.02891
60	171	168.634395	2.365605	0.588508	0.600717	0.592177	0.040237	0.007564
61	175	174.619901	0.380099	0.09456	0.096663	0.094558	0.043047	0.00021
62	182	176.615069	5.384931	1.339646	1.370622	1.398839	0.044688	0.043939
63	180	180.605406	-0.605406	-0.150611	-0.154445	-0.151128	0.049027	0.000615
64	183	186.590912	-3.590912	-0.893336	-0.920513	-0.917336	0.058176	0.02617
65	188	190.581249	-2.581249	-0.642155	-0.664469	-0.656192	0.066036	0.015609
66	200	196.566754	3.433246	0.854112	0.8907	0.886546	0.080467	0.034712
67	194	202.552259	-8.552259	-2.127605	-2.240288	-2.478032	0.098067	0.272853
68	206	204.547428	1.452572	0.361366	0.381899	0.374695	0.104639	0.008522
69	207	206.542596	0.457404	0.113791	0.120725	0.118108	0.111562	0.000915
70	210	210.532933	-0.532933	-0.132581	-0.141854	-0.138797	0.126465	0.001457
71	219	216.518439	2.481561	0.617355	0.670192	0.661956	0.151461	0.040086

Hypothesis Tests in Simple Linear Regression

The following sections discuss hypothesis tests on the regression coefficients in simple linear regression. These tests can be carried out if it can be assumed that the random error term, ϵ, is normally and independently distributed with a mean of zero and variance of σ^2.

t Tests

The t tests are used to conduct hypothesis tests on the regression coefficients obtained in simple linear regression. A statistic based on the t distribution is used to test the two-sided hypothesis that the true slope, β_1, equals some constant value, $\beta_{1,0}$. The statements for the hypothesis test are expressed as:

$$H_0: \quad \beta_1 = \beta_{1,0}$$
$$H_1: \quad \beta_1 \neq \beta_{1,0}$$

The test statistic used for this test is:

$$T_0 = \frac{\hat{\beta}_1 - \beta_{1,0}}{se(\hat{\beta}_1)}$$

where $\hat{\beta}_1$ is the least square estimate of β_1 , and $se(\hat{\beta}_1)$ is its standard error. The value of $se(\hat{\beta}_1)$ can be calculated as follows:

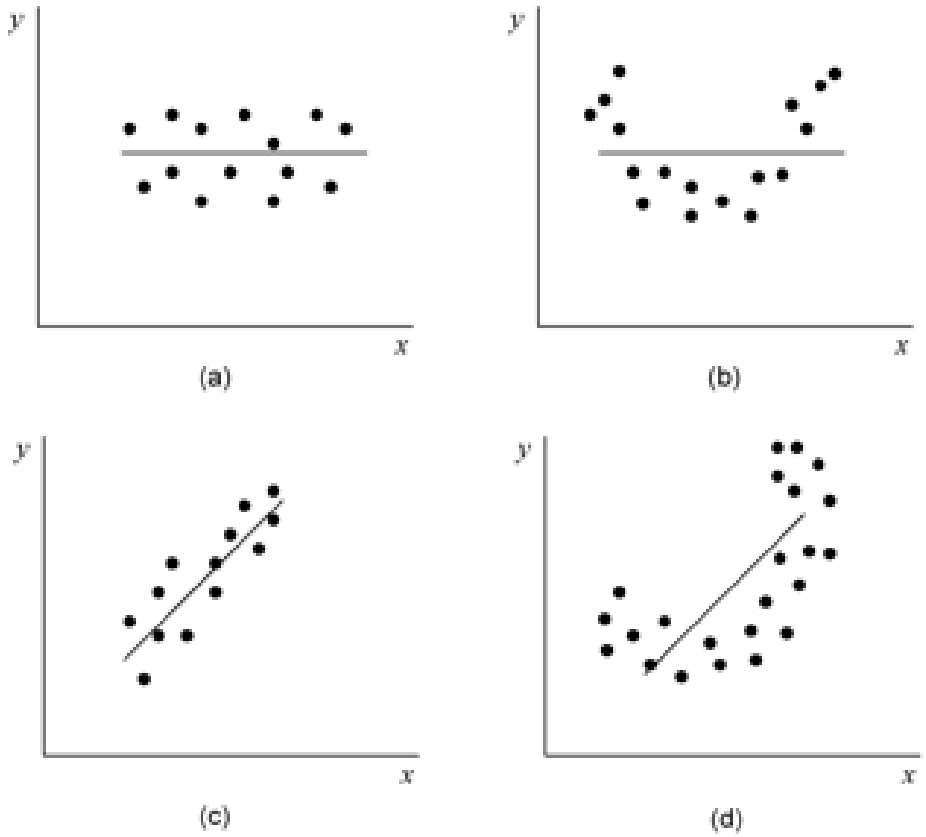
$$se(\hat{\beta}_1) = \sqrt{\frac{\frac{\sum_{i=1}^{n} e_i^2}{n-2}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

The test statistic, T_0 , follows a t distribution with $(n-2)$ degrees of freedom, where n is the total number of observations. The null hypothesis, H_0 , is accepted if the calculated value of the test statistic is such that:

$$-t_{\alpha/2,n-2} < T_0 < t_{\alpha/2,n-2}$$

where $t_{\alpha/2,n-2}$ and $-t_{\alpha/2,n-2}$ are the critical values for the two-sided hypothesis. $t_{\alpha/2,n-2}$ is the percentile of the t distribution corresponding to a cumulative probability of $(1-\alpha/2)$ and α is the significance level.

If the value of $\beta_{1,0}$ used is zero, then the hypothesis tests for the significance of regression. In other words, the test indicates if the fitted regression model is of value in explaining variations in the observations or if you are trying to impose a regression model when no true relationship exists between x and Y . Failure to reject $H_0 : \beta_1 = 0$ implies that no linear relationship exists between x and Y . This result may be obtained when the scatter plots of against are as shown in (a) of the following figure and (b) of the following figure. (a) represents the case where no model exits for the observed data. In this case you would be trying to fit a regression model to noise or random variation. (b) represents the case where the true relationship between x and Y is not linear. (c) and (d) represent the case when $H_0 : \beta_1 = 0$ is rejected, implying that a model does exist between x and Y . (c) represents the case where the linear model is sufficient. In the following figure, (d) represents the case where a higher order model may be needed.



A similar procedure can be used to test the hypothesis on the intercept. The test statistic used in this case is:

$$T_0 = \frac{\hat{\beta}_0 - \beta_{0,0}}{se(\hat{\beta}_0)}$$

where $\hat{\beta}_0$ is the least square estimate of β_0 , and $se(\hat{\beta}_0)$ is its standard error which is calculated using:

$$se(\hat{\beta}_0) = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-2}\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]}$$

Example

The test for the significance of regression for the data in the preceding table is illustrated in this example. The test is carried out using the t test on the coefficient β_1 . The hypothesis to be tested is $H_0 : \beta_1 = 0$. To calculate the statistic to test H_0 , the estimate, $\hat{\beta}_1$, and the standard

$\hat{\beta}_1$ was obtained in this section. The standard error can be calculated as follows:

$$se(\hat{\beta}_1) = \sqrt{\frac{\frac{\sum_{i=1}^{n} e_i^2}{n-2}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$
$$= \sqrt{\frac{(371.627/23)}{5679.36}}$$
$$= 0.0533$$

Then, the test statistic can be calculated using the following equation:

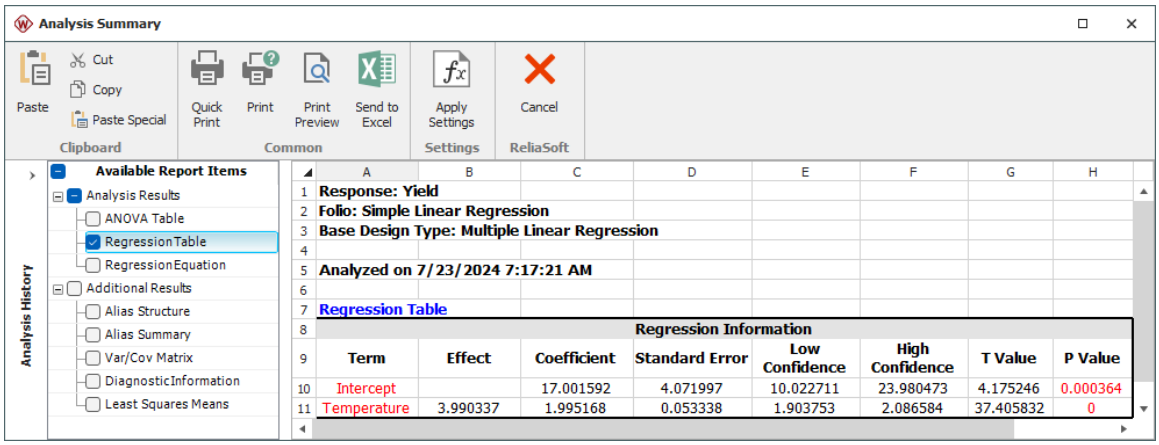
$$t_0 = \frac{\hat{\beta}_1 - \beta_{1,0}}{se(\hat{\beta}_0)} = \frac{1.9952 - 0}{0.0533} = 37.4058$$

The p value corresponding to this statistic based on the t distribution with 23 (n-2 = 25-2 = 23) degrees of freedom can be obtained as follows:

$$pvalue = 2 \times (1 - P(T \leq t_0)$$
$$= 2 \times (1 - 0.999999)$$
$$= 0$$

Assuming that the desired significance level is 0.1, since p value < 0.1, $H_0 : \beta_1 = 0$ is rejected indicating that a relation exists between temperature and yield for the data in the preceding table. Using this result along with the scatter plot, it can be concluded that the relationship between temperature and yield is linear.

In Weibull++ DOE folios, information related to the t test is displayed in the Regression Information table as shown in the following figure. In this table the t test for β_1 is displayed in the row for the term Temperature because β_1 is the coefficient that represents the variable temperature in the regression model. The columns labeled Standard Error, T Value and P Value represent the standard error, the test statistic for the test and the p value for the t test, respectively. These values have been calculated for β_1 in this example. The Coefficient column represents the estimate of regression coefficients. The Effect column represents values obtained by multiplying the coefficients by a factor of 2. This value is useful in the case of two factor experiments and is explained in Two Level Factorial Experiments. Columns Low Confidence and High Confidence represent the limits of the confidence intervals for the regression coefficients and are explained in Confidence Interval on Regression Coefficients.



Analysis of Variance Approach to Test the Significance of Regression

The analysis of variance (ANOVA) is another method to test for the significance of regression. As the name implies, this approach uses the variance of the observed data to determine if a regression model can be applied to the observed data. The observed variance is partitioned into components that are then used in the test for significance of regression.

Sum of Squares

The total variance (i.e., the variance of all of the observed data) is estimated using the observed data. As mentioned in Statistical Background, the variance of a population can be estimated using the sample variance, which is calculated using the following relationship:

$$s^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}$$

The quantity in the numerator of the previous equation is called the *sum of squares*. It is the sum of the square of deviations of all the observations, y_i , from their mean, $\bar{y}$. In the context of ANOVA this quantity is called the *total sum of squares* (abbreviated SS_T) because it relates to the total variance of the observations. Thus:

$$SS_T = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

The denominator in the relationship of the sample variance is the number of degrees of freedom associated with the sample variance. Therefore, the number of degrees of freedom associated with SS_T , $dof(SS_T)$, is $n-1$. The sample variance is also referred to as a *mean square* because it is obtained by dividing the sum of squares by the respective degrees of freedom. Therefore, the total mean square (abbreviated MS_T) is:

$$MS_T = \frac{SS_T}{dof(SS_T)} = \frac{SS_T}{n-1}$$

When you attempt to fit a regression model to the observations, you are trying to explain some of the variation of the observations using this model. If the regression model is such that the resulting fitted regression line passes through all of the observations, then you would have a "perfect" model (see (a) of the figure below). In this case the model would explain all of the variability of the observations. Therefore, the model sum of squares (also referred to as the regression sum of squares and abbreviated SS_R) equals the total sum of squares; i.e., the model explains all of the observed variance:

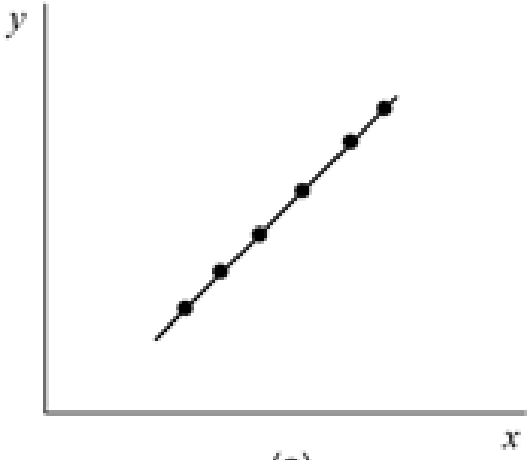
$$SS_R = SS_T$$

For the perfect model, the regression sum of squares, SS_R, equals the total sum of squares, SS_T, because all estimated values, $\hat{y}_i$, will equal the corresponding observations, y_i. SS_R can be calculated using a relationship similar to the one for obtaining SS_T by replacing y_i by $\hat{y}_i$ in the relationship of SS_T. Therefore:

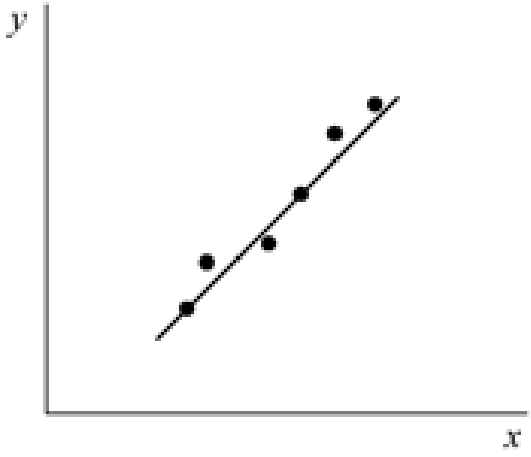
$$SS_R = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

The number of degrees of freedom associated with SS_R is 1.

Based on the preceding discussion of ANOVA, a perfect regression model exists when the fitted regression line passes through all observed points. However, this is not usually the case, as seen in (b) of the following figure.



(a)



(b)

In both of these plots, a number of points do not follow the fitted regression line. This indicates that a part of the total variability of the observed data still remains unexplained. This portion of the total variability or the total sum of squares, that is not explained by the model, is called the *residual sum of squares* or the *error sum of squares* (abbreviated SS_E). The deviation for this sum of squares is obtained at each observation in the form of the residuals, e_i . The error sum of squares can be obtained as the sum of squares of these deviations:

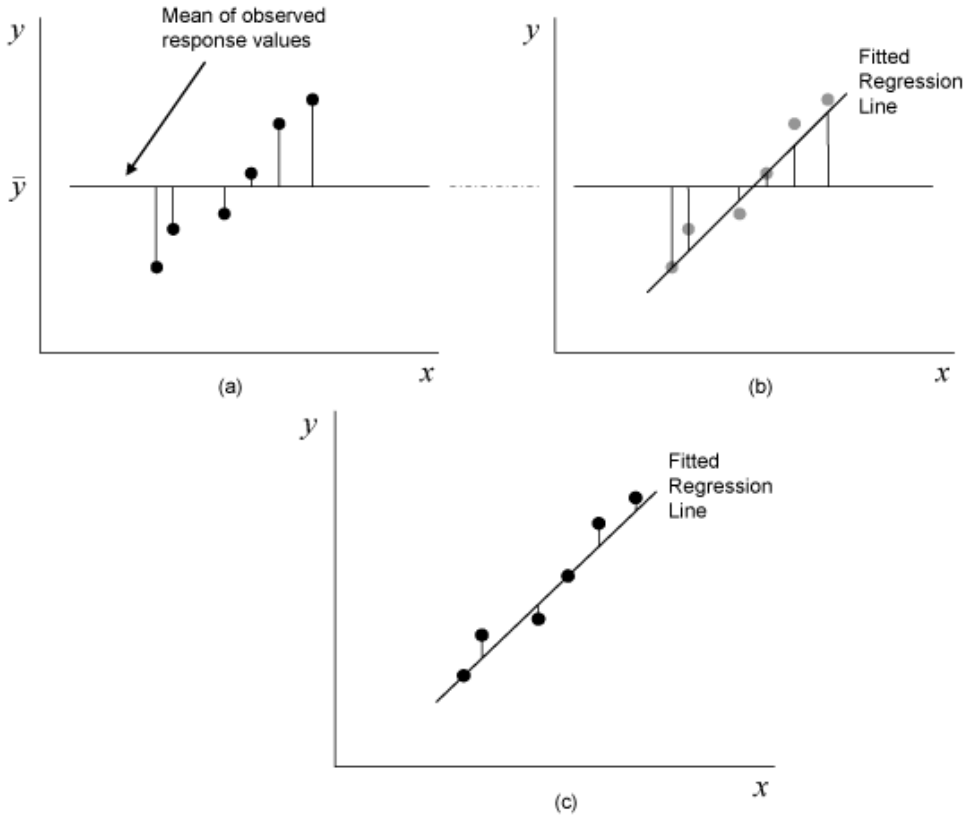
$$SS_E = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

The number of degrees of freedom associated with SS_E , $dof(SS_E)$, is $(n-2)$. The total variability of the observed data (i.e., total sum of squares, SS_T) can be written using the portion of the variability explained by the model, SS_R , and the portion unexplained by the model, SS_E , as:

$$SS_T = SS_R + SS_E$$

The above equation is also referred to as the analysis of variance identity and can be expanded as follows:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$



Mean Squares

As mentioned previously, mean squares are obtained by dividing the sum of squares by the respective degrees of freedom. For example, the error mean square, MS_E, can be obtained as:

$$MS_E = \frac{SS_E}{dof(SS_E)} = \frac{SS_E}{n-2}$$

The error mean square is an estimate of the variance, σ^2, of the random error term, ϵ, and can be written as:

$$\hat{\sigma}^2 = \frac{SS_E}{n-2}$$

Similarly, the regression mean square, MS_R, can be obtained by dividing the regression sum of squares by the respective degrees of freedom as follows:

$$MS_R = \frac{SS_R}{dof(SS_R)} = \frac{SS_R}{1}$$

F Test

To test the hypothesis $H_0 : \beta_1 = 0$, the statistic used is based on the F distribution. It can be shown that if the null hypothesis H_0 is true, then the statistic:

$$F_0 = \frac{MS_R}{MS_E} = \frac{SS_R/1}{SS_E/(n-2)}$$

follows the F distribution with 1 degree of freedom in the numerator and $(n-2)$ degrees of freedom in the denominator. H_0 is rejected if the calculated statistic, F_0, is such that:

$$F_0 > f_{\alpha,1,n-2}$$

where $f_{\alpha,1,n-2}$ is the percentile of the F distribution corresponding to a cumulative probability of $(1-\alpha)$ and α is the significance level.

Example

The analysis of variance approach to test the significance of regression can be applied to the yield data in the preceding table. To calculate the statistic, F_0, for the test, the sum of squares have to be obtained. The sum of squares can be calculated as shown next. The total sum of squares can be calculated as:

$$\begin{aligned} SS_T &= \sum_{i=1}^{n} (y_i - \bar{y})^2 \\ &= \sum_{i=1}^{25} (y_i - 166.32)^2 \\ &= 22979.44 \end{aligned}$$

The regression sum of squares can be calculated as:

$$\begin{aligned} SS_R &= \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \\ &= \sum_{i=1}^{25} (\hat{y}_i - 166.32)^2 \\ &= 22607.81 \end{aligned}$$

The error sum of squares can be calculated as:

$$\begin{aligned} SS_E &= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \\ &= \sum_{i=1}^{25} (y_i - \hat{y}_i)^2 \\ &= 371.63 \end{aligned}$$

Knowing the sum of squares, the statistic to test $H_0 : \beta_1 = 0$ can be calculated as follows:

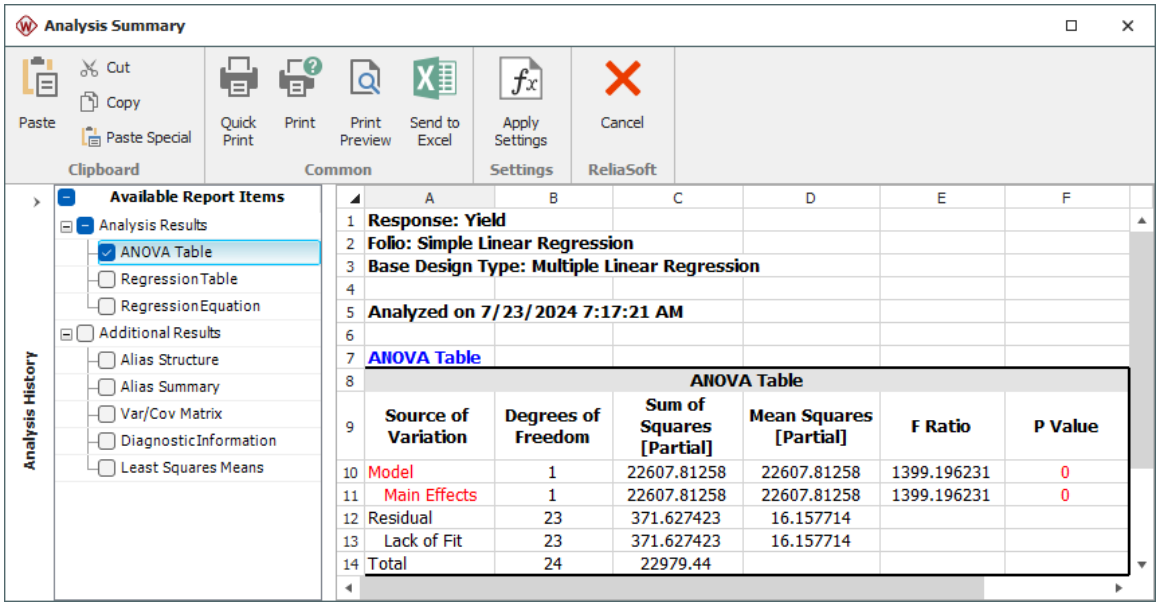
$$\begin{aligned} f_0 &= \frac{MS_R}{MS_E} \\ &= \frac{SS_R/1}{SS_E/(n-2)} \\ &= \frac{22607.81/1}{371.63/(25-2)} \\ &= 1399.20 \end{aligned}$$

The critical value at a significance level of 0.1 is $f_{0.05,1,23} = 2.937$. Since $f_0 > f_{\alpha,1,n-2}$, $H_0 : \beta_1 = 0$ is rejected and it is concluded that β_1 is not zero. Alternatively, the p value can also be used. The p value corresponding to the test statistic, f_0, based on the F distribution with one degree of freedom in the numerator and 23 degrees of freedom in the denominator is:

$$\begin{aligned} p\ value &= 1 - P(F \leq f_0) \\ &= 1 - 0.999999 \\ &= 4.17E - 22 \end{aligned}$$

Assuming that the desired significance is 0.1, since the p value < 0.1, then $H_0 : \beta_1 = 0$ is rejected, implying that a relation does exist between temperature and yield for the data in the preceding table. Using this result along with the scatter plot of the above figure, it can be concluded that the relationship that exists between temperature and yield is linear. This result is displayed in the ANOVA table as shown in the following figure. Note that this is the same result that was obtained from the t test in the section t Tests. The ANOVA and Regression Information tables in Weibull++ DOE folios represent two different ways to test for the significance of the regression model. In the case of multiple linear regression models these tables are expanded to allow tests on individual variables used in the model. This is done using extra sum of

squares. Multiple linear regression models and the application of extra sum of squares in the analysis of these models are discussed in Multiple Linear Regression Analysis.



ANOVA Table

Source of Variation	Degrees of Freedom	Sum of Squares [Partial]	Mean Squares [Partial]	F Ratio	P Value
Model	1	22607.81258	22607.81258	1399.196231	0
Main Effects	1	22607.81258	22607.81258	1399.196231	0
Residual	23	371.627423	16.157714		
Lack of Fit	23	371.627423	16.157714		
Total	24	22979.44			

Confidence Intervals in Simple Linear Regression

A confidence interval represents a closed interval where a certain percentage of the population is likely to lie. For example, a 90% confidence interval with a lower limit of A and an upper limit of B implies that 90% of the population lies between the values of A and B. Out of the remaining 10% of the population, 5% is less than A and 5% is greater than B. (For details refer to the Life Data Analysis Reference Book.) This section discusses confidence intervals used in simple linear regression analysis.

Confidence Interval on Regression Coefficients

A 100 $(1-\alpha)$ percent confidence interval on β_1 is obtained as follows:

$$\hat{\beta}_1 \pm t_{\alpha/2,n-2} \cdot se(\hat{\beta}_1)$$

Similarly, a 100 $(1-\alpha)$ percent confidence interval on β_0 is obtained as:

$$\hat{\beta}_0 \pm t_{\alpha/2,n-2} \cdot se(\hat{\beta}_0)$$

Confidence Interval on Fitted Values

A 100 $(1-\alpha)$ percent confidence interval on any fitted value, $\hat{y}_i$, is obtained as follows:

$$\hat{y}_i \pm t_{\alpha/2,n-2}\sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{(x_i-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]}$$

It can be seen that the width of the confidence interval depends on the value of x_i and will be a minimum at $x_i = \bar{x}$ and will widen as $|x_i - \bar{x}|$ increases.

Confidence Interval on New Observations

For the data in the preceding table, assume that a new value of the yield is observed after the regression model is fit to the data. This new observation is independent of the observations used to obtain the regression model. If x_p is the level of the temperature at which the new observation was taken, then the estimate for this new value based on the fitted regression model is:

$$\begin{aligned}\hat{y}_p &= \hat{\beta}_0 + \hat{\beta}_1 x_p \\ &= 17.0016 + 1.9952 \times x_p\end{aligned}$$

If a confidence interval needs to be obtained on $\hat{y}_p$, then this interval should include both the error from the fitted model and the error associated with future observations. This is because $\hat{y}_p$ represents the estimate for a value of Y that was not used to obtain the regression model. The confidence interval on $\hat{y}_p$ is referred to as the *prediction interval*. A 100 $(1-\alpha)$ percent prediction interval on a new observation is obtained as follows:

$$\hat{y}_p \pm t_{\alpha/2,n-2}\sqrt{\hat{\sigma}^2\left[1+\frac{1}{n}+\frac{(x_p-\bar{x})^2}{\sum_{i=1}^{n}(x_p-\bar{x})^2}\right]}$$

Example

To illustrate the calculation of confidence intervals, the 95% confidence intervals on the response at $x = 93$ for the data in the preceding table is obtained in this example. A 95% prediction interval is also obtained assuming that a new observation for the yield was made at $x = 91$.

The fitted value, $\hat{y}_i$, corresponding to $x = 93$ is:

$$\begin{aligned}\hat{y}_{21} &= \hat{\beta}_0 + \hat{\beta}_1 x_{21} \\ &= 17.0016 + 1.9952 \times 93 \\ &= 202.6\end{aligned}$$

The 95% confidence interval $(\alpha = 0.05)$ on the fitted value, $\hat{y}_{21} = 202.6$, is:

$$
\begin{aligned}
&= \hat{y}_i \pm t_{\alpha/2,n-2} \sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right]} \\
&= 202.6 \pm t_{0.025,23} \sqrt{MS_E \left[\frac{1}{25} + \frac{(93 - 74.84)^2}{5679.36} \right]} \\
&= 202.6 \pm 2.069 \sqrt{16.16 \left[\frac{1}{25} + \frac{(93 - 74.84)^2}{5679.36} \right]} \\
&= 202.6 \pm 2.602
\end{aligned}
$$

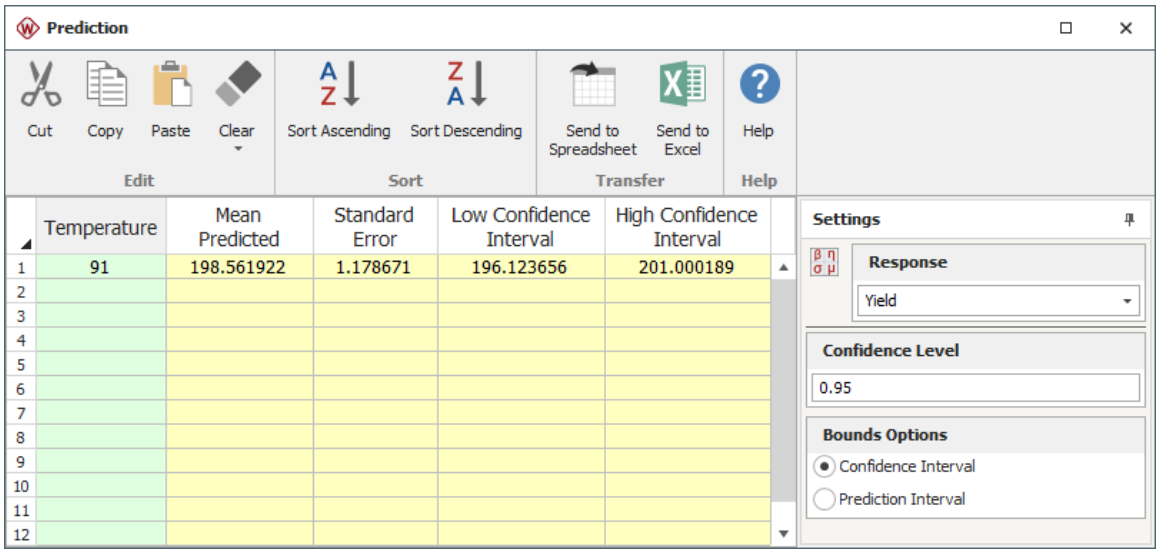
The 95% limits on $\hat{y}_{21}$ are 199.95 and 205.2, respectively. The estimated value based on the fitted regression model for the new observation at $x = 91$ is:

$$
\begin{aligned}
\hat{y}_p &= \hat{\beta}_0 + \hat{\beta}_1 x_p \\
&= 17.0016 + 1.9952 \times 91 \\
&= 198.6
\end{aligned}
$$

The 95% prediction interval on $\hat{y}_p = 198.6$ is:

$$
\begin{aligned}
&= \hat{y}_p \pm t_{\alpha/2,n-2} \sqrt{\hat{\sigma}^2 \left[1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum_{i=1}^{n} (x_p - \bar{x})^2} \right]} \\
&= 198.6 \pm t_{0.025,23} \sqrt{MS_E \left[1 + \frac{1}{25} + \frac{(93 - 74.84)^2}{5679.36} \right]} \\
&= 198.6 \pm 2.069 \sqrt{16.16 \left[1 + \frac{1}{25} + \frac{(93 - 74.84)^2}{5679.36} \right]} \\
&= 198.6 \pm 2.069 \times 4.1889 \\
&= 198.6 \pm 8.67
\end{aligned}
$$

The 95% limits on $\hat{y}_p$ are 189.9 and 207.2, respectively. In Weibull++ DOE folios, confidence and prediction intervals can be calculated from the control panel. The prediction interval values calculated in this example are shown in the figure below as Low Prediction Interval and High Prediction Interval, respectively. The columns labeled Mean Predicted and Standard Error represent the values of $\hat{y}_p$ and the standard error used in the calculations.



	Temperature	Mean Predicted	Standard Error	Low Confidence Interval	High Confidence Interval
1	91	198.561922	1.178671	196.123656	201.000189
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					

Measures of Model Adequacy

It is important to analyze the regression model before inferences based on the model are undertaken. The following sections present some techniques that can be used to check the appropriateness of the model for the given data. These techniques help to determine if any of the model assumptions have been violated.

Coefficient of Determination (*R*^2)

The coefficient of determination is a measure of the amount of variability in the data accounted for by the regression model. As mentioned previously, the total variability of the data is measured by the total sum of squares, SS_T . The amount of this variability explained by the regression model is the regression sum of squares, SS_R . The coefficient of determination is the ratio of the regression sum of squares to the total sum of squares.

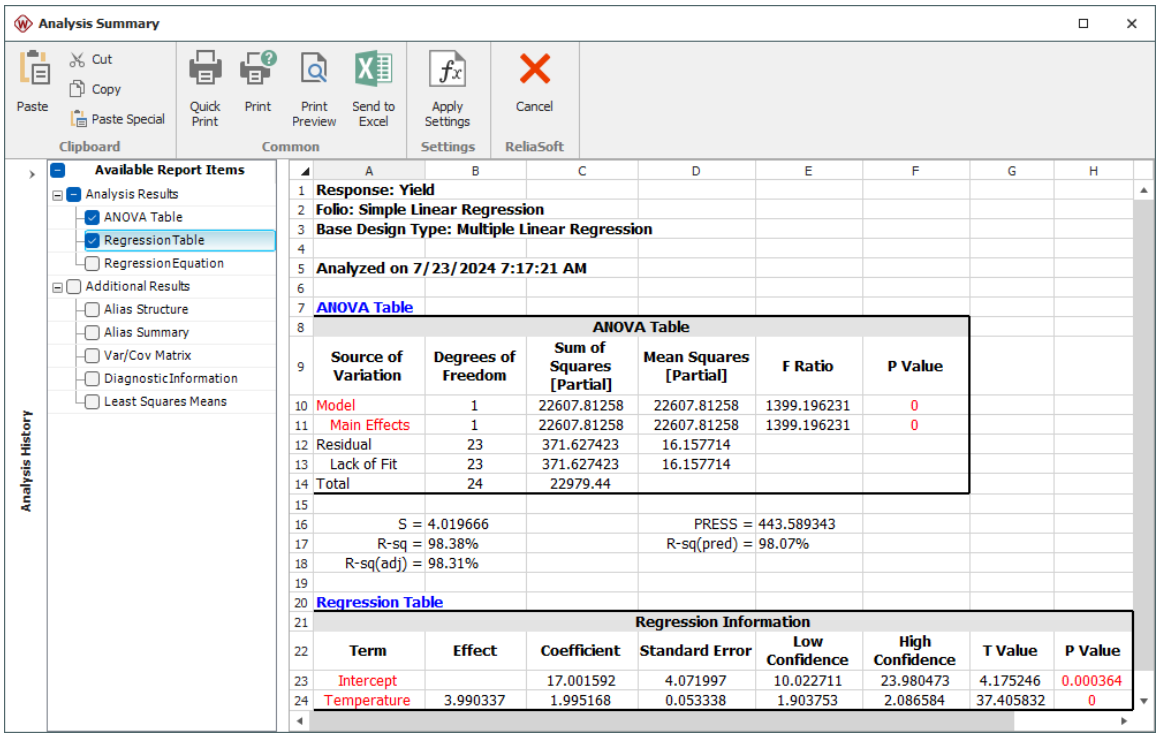
$$R^2 = \frac{SS_R}{SS_T}$$

R^2 can take on values between 0 and 1 since $R^2 = \frac{SS_R}{SS_T}$. For the yield data example, R^2 can be calculated as:

$$\begin{aligned} R^2 &= \frac{SS_R}{SS_T} \\ &= \frac{22607.81}{22979.44} \\ &= 0.98 \end{aligned}$$

Therefore, 98% of the variability in the yield data is explained by the regression model, indicating a very good fit of the model. It may appear that larger values of R^2 indicate a better fitting regression model. However, R^2 should be used cautiously as this is not always the case. The value of R^2 increases as more terms are added to the model, even if the new term does not contribute significantly to the model. Therefore, an increase in the value of R^2 cannot be taken as a sign to conclude that the new model is superior to the older model. Adding a new term may make the regression model worse if the error mean square, MS_E, for the new model is larger than the MS_E of the older model, even though the new model will show an increased value of R^2. In the results obtained from the DOE folio, R^2 is displayed as R-sq under the ANOVA table (as shown in the figure below), which displays the complete analysis sheet for the data in the preceding table.

The other values displayed with are S, R-sq(adj), PRESS and R-sq(pred). These values measure different aspects of the adequacy of the regression model. For example, the value of S is the square root of the error mean square, MS_E, and represents the "standard error of the model." A lower value of S indicates a better fitting model. The values of S, R-sq and R-sq(adj) indicate how well the model fits the observed data. The values of PRESS and R-sq(pred) are indicators of how well the regression model predicts new observations. R-sq(adj), PRESS and R-sq(pred) are explained in Multiple Linear Regression Analysis.



	A	B	C	D	E	F	G	H
1	**Response: Yield**							
2	**Folio: Simple Linear Regression**							
3	**Base Design Type: Multiple Linear Regression**							
4								
5	**Analyzed on 7/23/2024 7:17:21 AM**							
6								
7	**ANOVA Table**							

ANOVA Table

Source of Variation	Degrees of Freedom	Sum of Squares [Partial]	Mean Squares [Partial]	F Ratio	P Value
Model	1	22607.81258	22607.81258	1399.196231	0
Main Effects	1	22607.81258	22607.81258	1399.196231	0
Residual	23	371.627423	16.157714		
Lack of Fit	23	371.627423	16.157714		
Total	24	22979.44			

S =	4.019666	PRESS =	443.589343
R-sq =	98.38%	R-sq(pred) =	98.07%
R-sq(adj) =	98.31%		

Regression Table

Regression Information

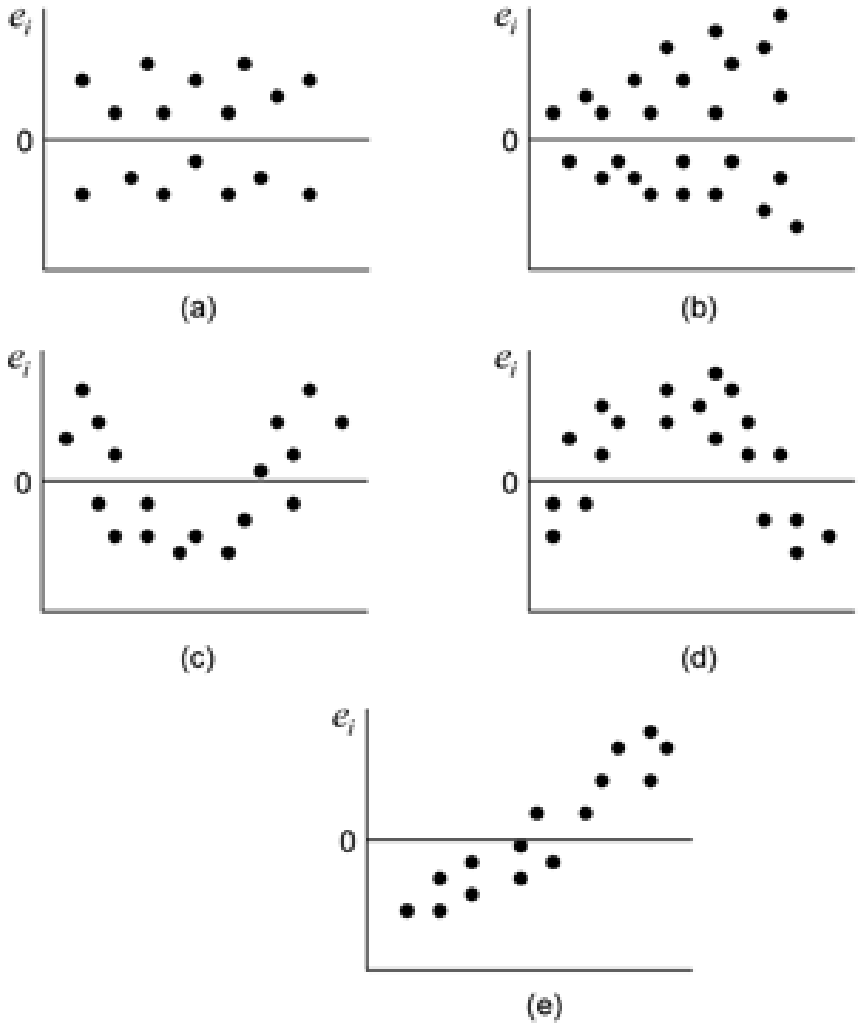
Term	Effect	Coefficient	Standard Error	Low Confidence	High Confidence	T Value	P Value
Intercept		17.001592	4.071997	10.022711	23.980473	4.175246	0.000364
Temperature	3.990337	1.995168	0.053338	1.903753	2.086584	37.405832	0

Residual Analysis

In the simple linear regression model the true error terms, ϵ_i , are never known. The residuals, e_i , may be thought of as the observed error terms that are similar to the true error terms. Since the true error terms, ϵ_i , are assumed to be normally distributed with a mean of zero and a variance of σ^2 , in a good model the observed error terms (i.e., the residuals, e_i) should also follow these assumptions. Thus the residuals in the simple linear regression should be normally distributed with a mean of zero and a constant variance of σ^2 . Residuals are usually plotted against the fitted values, $\hat{y}_i$, against the predictor variable values, x_i , and against time or run-order sequence, in addition to the normal probability plot. Plots of residuals are used to check for the following:

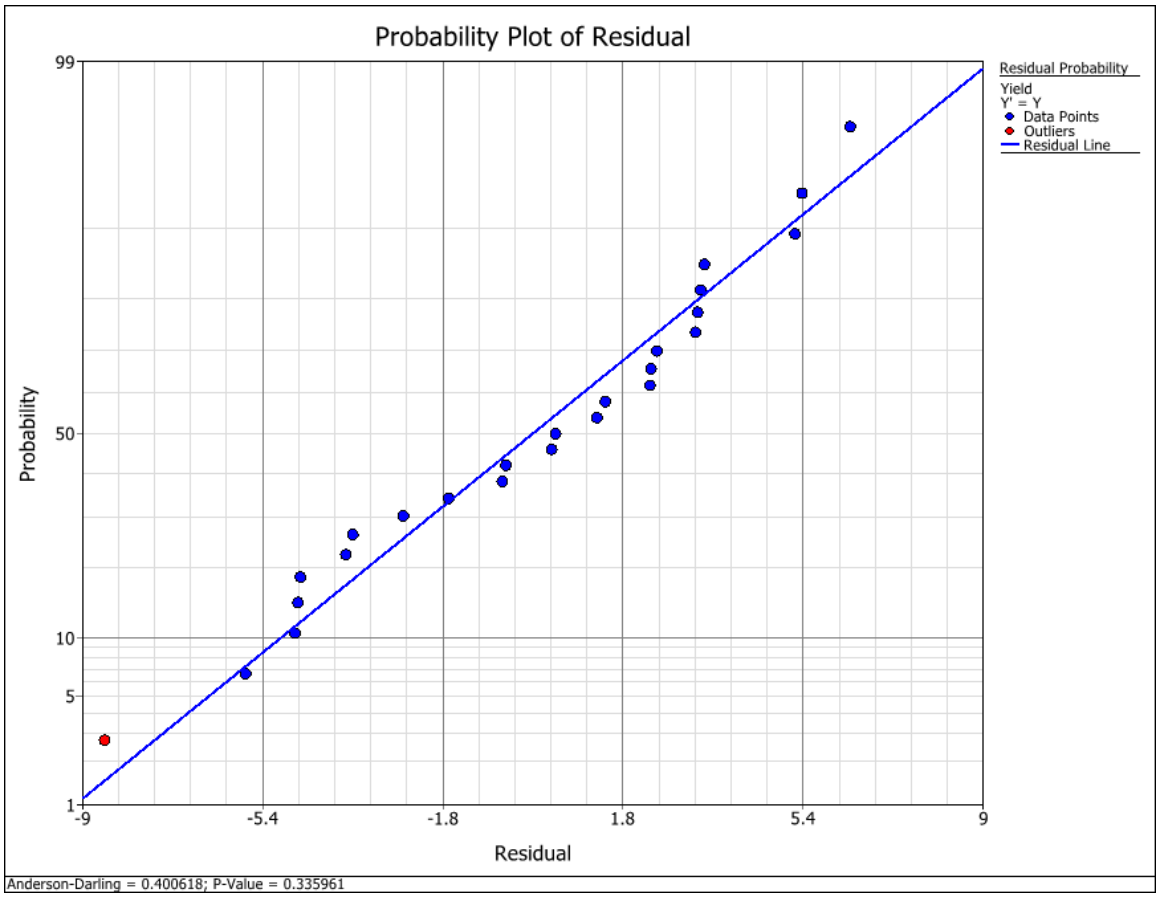
1. Residuals follow the normal distribution.
2. Residuals have a constant variance.
3. Regression function is linear.
4. A pattern does not exist when residuals are plotted in a time or run-order sequence.
5. There are no outliers.

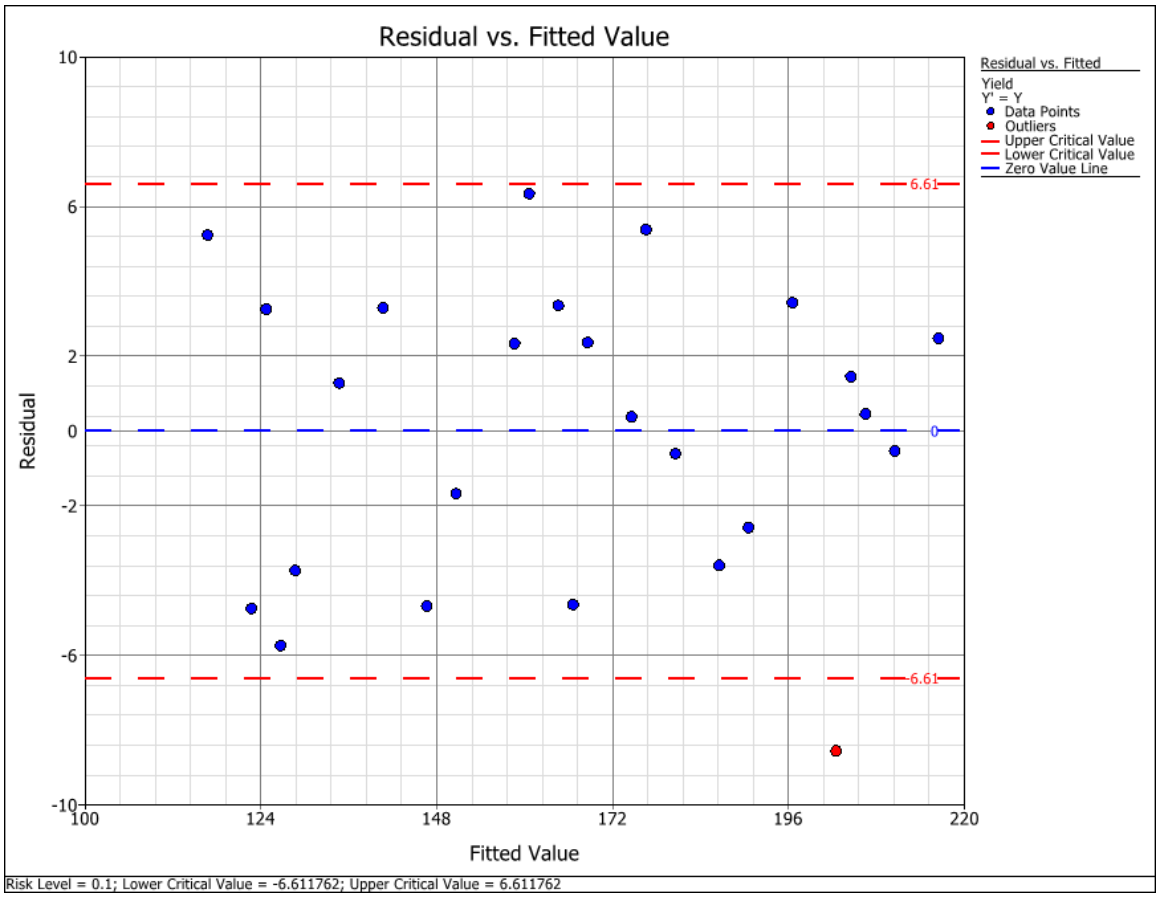
Examples of residual plots are shown in the following figure. (a) is a satisfactory plot with the residuals falling in a horizontal band with no systematic pattern. Such a plot indicates an appropriate regression model. (b) shows residuals falling in a funnel shape. Such a plot indicates increase in variance of residuals and the assumption of constant variance is violated here. Transformation on Y may be helpful in this case (see Transformations). If the residuals follow the pattern of (c) or (d), then this is an indication that the linear regression model is not adequate. Addition of higher order terms to the regression model or transformation on x or Y may be required in such cases. A plot of residuals may also show a pattern as seen in (e), indicating that the residuals increase (or decrease) as the run order sequence or time progresses. This may be due to factors such as operator-learning or instrument-creep and should be investigated further.

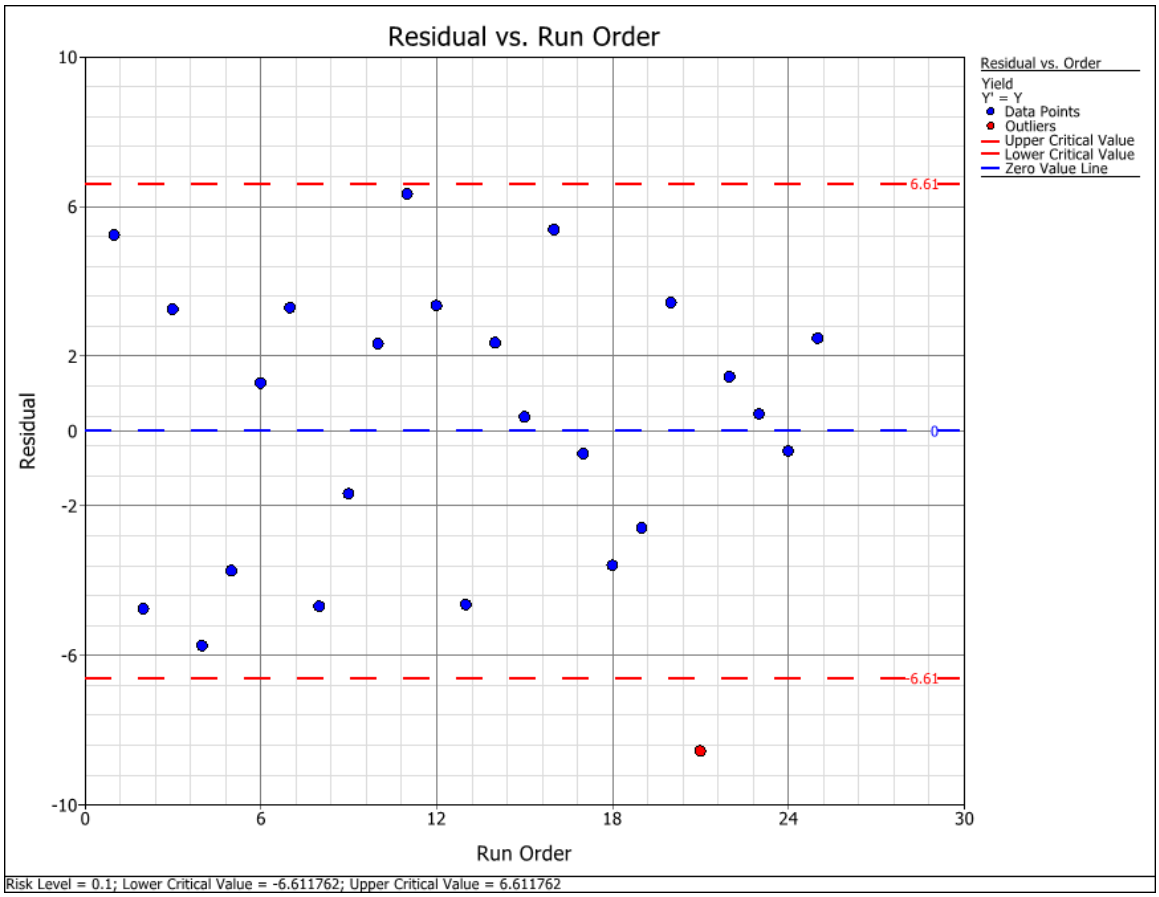


Example

Residual plots for the data of the preceding table are shown in the following figures. One of the following figures is the normal probability plot. It can be observed that the residuals follow the normal distribution and the assumption of normality is valid here. In one of the following figures the residuals are plotted against the fitted values, $\hat{y}_i$, and in one of the following figures the residuals are plotted against the run order. Both of these plots show that the 21st observation seems to be an outlier. Further investigations are needed to study the cause of this outlier.


Probability Plot of Residual
Probability
99
50
10
5
1
-9
-5.4
-1.8
1.8
5.4
9
Residual
Residual Probability
Yield
Y' = Y
Data Points
Outliers
Residual Line
Anderson-Darling = 0.400618; P-Value = 0.335961


Residual vs. Fitted Value
Residual vs. Fitted
Yield
Y' = Y
Data Points
Outliers
Upper Critical Value
Lower Critical Value
Zero Value Line
Residual
Fitted Value
10
6
2
0
-2
-6
-10
100
124
148
172
196
220
6.61
6.61
0
Risk Level = 0.1; Lower Critical Value = -6.611762; Upper Critical Value = 6.611762



Lack-of-Fit Test

As mentioned in Analysis of Variance Approach, ANOVA, a perfect regression model results in a fitted line that passes exactly through all observed data points. This perfect model will give us a zero error sum of squares ($SS_E = 0$). Thus, no error exists for the perfect model. However, if you record the response values for the same values of x_i for a second time, in conditions maintained as strictly identical as possible to the first time, observations from the second time will not all fall along the perfect model. The deviations in observations recorded for the second time constitute the "purely" random variation or noise. The sum of squares due to pure error (abbreviated SS_{PE}) quantifies these variations. SS_{PE} is calculated by taking repeated observations at some or all values of x_i and adding up the square of deviations at each level of x using the respective repeated observations at that x value.

Assume that there are n levels of x and m_i repeated observations are taken at each i the level. The data is collected as shown next:

$y_{11}, y_{12}, \ldots, y_{1m_1}$ repeated observations at x_1
$y_{21}, y_{22}, \ldots, y_{2m_2}$ repeated observations at x_2
...
$y_{i1}, y_{i2}, \ldots, y_{im_i}$ repeated observations at x_i
...
$y_{n1}, y_{n2}, \ldots, y_{nm_n}$ repeated observations at x_n

The sum of squares of the deviations from the mean of the observations at i the level of x, x_i, can be calculated as:

$$\sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i)^2$$

where $\bar{y}_i$ is the mean of the m_i repeated observations corresponding to x_i ($\bar{y}_i = (1/m_i) \sum_{j=1}^{m_i} y_{ij}$). The number of degrees of freedom for these deviations is $(m_i - 1)$ as there are m_i observations at i the level of x but one degree of freedom is lost in calculating the mean, $\bar{y}_i$.

The total sum of square deviations (or SS_{PE}) for all levels of x can be obtained by summing the deviations for all x_i as shown next:

$$SS_{PE} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i)^2$$

The total number of degrees of freedom associated with SS_{PE} is:

$$= \sum_{i=1}^{n} (m_i - 1)$$
$$= \sum_{i=1}^{n} m_i - n$$

If all $m_i = m$, (i.e., m repeated observations are taken at all levels of x), then $\sum_{i=1}^{n} m_i = nm$ and the degrees of freedom associated with SS_{PE} are:

$$= nm - n$$

The corresponding mean square in this case will be:

$$MS_{PE} = \frac{SS_{PE}}{nm - n}$$

When repeated observations are used for a perfect regression model, the sum of squares due to pure error, SS_{PE}, is also considered as the error sum of squares, SS_E. For the case when repeated observations are used with imperfect regression models, there are two components of

the error sum of squares, SS_E . One portion is the pure error due to the repeated observations. The other portion is the error that represents variation not captured because of the imperfect model. The second portion is termed as the sum of squares due to lack-of-fit (abbreviated SS_{LOF}) to point to the deficiency in fit due to departure from the perfect-fit model. Thus, for an imperfect regression model:

$$SS_E = SS_{PE} + SS_{LOF}$$

Knowing SS_E and SS_{PE} , the previous equation can be used to obtain SS_{LOF} :

$$SS_{LOF} = SS_E - SS_{PE}$$

The degrees of freedom associated with SS_{LOF} can be obtained in a similar manner using sub-traction. For the case when m repeated observations are taken at all levels of x , the number of degrees of freedom associated with SS_{PE} is:

$$dof(SS_{PE}) = nm - n$$

Since there are nm total observations, the number of degrees of freedom associated with SS_E is:

$$dof(SS_E) = nm - 2$$

Therefore, the number of degrees of freedom associated with SS_{LOF} is:

$$\begin{aligned} &= dof(SS_E) - dof(SS_{PE}) \\ &= (nm - 2) - (nm - n) \\ &= n - 2 \end{aligned}$$

The corresponding mean square, MS_{LOF} , can now be obtained as:

$$MS_{LOF} = \frac{SS_{LOF}}{n-2}$$

The magnitude of SS_{LOF} or MS_{LOF} will provide an indication of how far the regression model is from the perfect model. An F test exists to examine the lack-of-fit at a particular significance level. The quantity MS_{LOF}/MS_{PE} follows an F distribution with $(n-2)$ degrees of freedom in the numerator and $(nm-n)$ degrees of freedom in the denominator when all m_i equal m . The test statistic for the lack-of-fit test is:

$$F_0 = \frac{MS_{LOF}}{MS_{PE}}$$

If the critical value $f_{\alpha,n-2,mn-n}$ is such that:

$$F_0 > f_{\alpha,n-2,nm-n}$$

it will lead to the rejection of the hypothesis that the model adequately fits the data.

Example

Assume that a second set of observations are taken for the yield data of the preceding table. The resulting observations are recorded in the following table. To conduct a lack-of-fit test on this data, the statistic $F_0 = MS_{LOF}/MS_{PE}$, can be calculated as shown next.

Observation Number	Temperature (x_i)	Yield	
		Replicate 1 (y_{i1})	Replicate 2 (y_{i2})
1	50	122	115
2	53	118	127
3	54	128	120
4	55	121	123
5	56	125	132
6	59	136	138
7	62	144	138
8	65	142	143
9	67	149	154
10	71	161	157
11	72	167	166
12	74	168	165
13	75	162	170
14	76	171	173
15	79	175	171
16	80	182	183
17	82	180	184
18	85	183	189
19	87	188	194
20	90	200	202
21	93	194	207
22	94	206	208
23	95	207	210
24	97	210	211
25	100	219	218

Calculation of Least Square Estimates

The parameters of the fitted regression model can be obtained as:

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^{50} y_i x_i - \frac{\left(\sum_{i=1}^{50} y_i\right)\left(\sum_{i=1}^{50} x_i\right)}{50}}{\sum_{i=1}^{50} (x_i - \bar{x})^2} \\ &= \frac{648532 - \frac{8356 \times 3742}{50}}{11358.72} \\ &= 2.04\end{aligned}$$

$$\begin{aligned}\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ &= 167.12 - 2.04 \times 74.84 \\ &= 14.47\end{aligned}$$

Knowing $\hat{\beta}_1$ and $\hat{\beta}_0$, the fitted values, $\hat{y}_i$, can be calculated.

Calculation of the Sum of Squares

Using the fitted values, the sum of squares can be obtained as follows:

$$\begin{aligned}SS_T &= \sum_{i=1}^{50} (y_i - \bar{y})^2 \\ &= 47907.28\end{aligned}$$

$$\begin{aligned}SS_R &= \sum_{i=1}^{50} (\hat{y}_i - \bar{y})^2 \\ &= 47258.91\end{aligned}$$

$$\begin{aligned}SS_E &= \sum_{i=1}^{50} (y_i - \hat{y}_i)^2 \\ &= 648.37\end{aligned}$$

Calculation of MS_{LOF}

The error sum of squares, SS_E, can now be split into the sum of squares due to pure error, SS_{PE}, and the sum of squares due to lack-of-fit, SS_{LOF}. SS_{PE} can be calculated as follows considering that in this example $n = 25$ and $m = 2$:

$$\begin{aligned}SS_{PE} &= \sum_{i=1}^{n} \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i)^2 \\ &= \sum_{i=1}^{25} \sum_{j=1}^{2} (y_{ij} - \bar{y}_i)^2 \\ &= 350\end{aligned}$$

The number of degrees of freedom associated with SS_{PE} is:

$$\begin{aligned} dof(SS_{PE}) &= nm - n \\ &= 25 \times 2 - 25 \\ &= 25 \end{aligned}$$

The corresponding mean square, MS_{PE} , can now be obtained as:

$$\begin{aligned} MS_{PE} &= \frac{SS_{PE}}{dof(SS_{PE})} \\ &= \frac{350}{25} \\ &= 14 \end{aligned}$$

SS_{LOF} can be obtained by subtraction from SS_E as:

$$\begin{aligned} SS_{LOF} &= SS_E - SS_{PE} \\ &= 648.37 - 350 \\ &= 298.37 \end{aligned}$$

Similarly, the number of degrees of freedom associated with SS_{LOF} is:

$$\begin{aligned} dof(SS_{LOF}) &= dof(SS_E) - dof(SS_{PE}) \\ &= (nm - 2) - (nm - n) \\ &= 23 \end{aligned}$$

The lack-of-fit mean square is:

$$\begin{aligned} MS_{LOF} &= \frac{MS_{LOF}}{dof(MS_{LOF})} \\ &= \frac{298.37}{23} \\ &= 12.97 \end{aligned}$$

Calculation of the Test Statistic

The test statistic for the lack-of-fit test can now be calculated as:

$$\begin{aligned} f_0 &= \frac{MS_{LOF}}{MS_{PE}} \\ &= \frac{12.97}{14} \\ &= 0.93 \end{aligned}$$

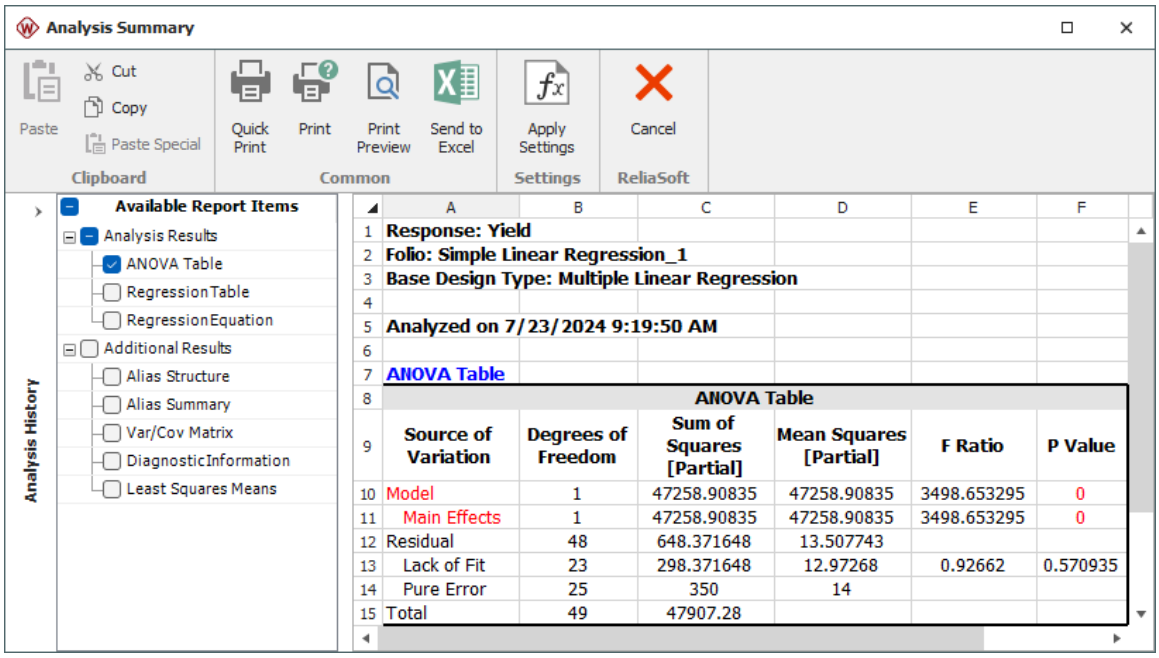
The critical value for this test is:

$$f_{0.05,23,25} = 1.97$$

Since $f_0 < f_{0.05,23,25}$, we fail to reject the hypothesis that the model adequately fits the data. The p value for this case is:

$$p\ value = 1 - P(F \leq f_0)$$
$$= 1 - 0.43$$
$$= 0.57$$

Therefore, at a significance level of 0.05 we conclude that the simple linear regression model, $y = 14.47 + 2.04x$, is adequate for the observed data. The following table presents a summary of the ANOVA calculations for the lack-of-fit test.

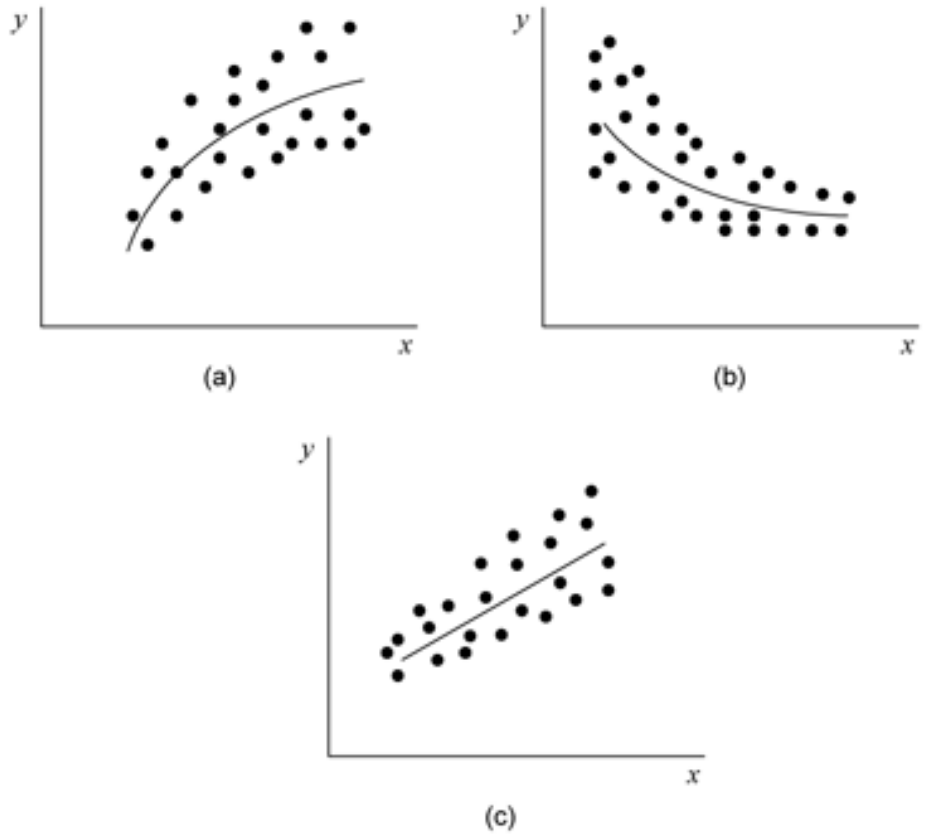


	A	B	C	D	E	F
1	**Response: Yield**					
2	**Folio: Simple Linear Regression_1**					
3	**Base Design Type: Multiple Linear Regression**					
4						
5	**Analyzed on 7/23/2024 9:19:50 AM**					
6						
7	**ANOVA Table**					
8	**ANOVA Table**					
9	**Source of Variation**	**Degrees of Freedom**	**Sum of Squares [Partial]**	**Mean Squares [Partial]**	**F Ratio**	**P Value**
10	Model	1	47258.90835	47258.90835	3498.653295	0
11	Main Effects	1	47258.90835	47258.90835	3498.653295	0
12	Residual	48	648.371648	13.507743		
13	Lack of Fit	23	298.371648	12.97268	0.92662	0.570935
14	Pure Error	25	350	14		
15	Total	49	47907.28			

Transformations

The linear regression model may not be directly applicable to certain data. Non-linearity may be detected from scatter plots or may be known through the underlying theory of the product or process or from past experience. Transformations on either the predictor variable, x, or the response variable, Y, may often be sufficient to make the linear regression model appropriate for the transformed data. If it is known that the data follows the logarithmic distribution, then a logarithmic transformation on Y (i.e., $Y^* = \log(Y)$) might be useful. For data following the Poisson distribution, a square root transformation ($Y^* = \sqrt{Y}$) is generally applicable.

Transformations on Y may also be applied based on the type of scatter plot obtained from the data. The following figure shows a few such examples.



For the scatter plot labeled (a), a square root transformation ($Y^* = \sqrt{Y}$) is applicable. While for the plot labeled (b), a logarithmic transformation (i.e., $Y^* = \log(Y)$) may be applied. For the plot labeled (c), the reciprocal transformation ($Y^* = 1/Y$) is applicable. At times it may be helpful to introduce a constant into the transformation of Y. For example, if Y is negative and the logarithmic transformation on Y seems applicable, a suitable constant, k, may be chosen to make all observed Y positive. Thus the transformation in this case would be $Y^* = \log(k + Y)$.

The Box-Cox method may also be used to automatically identify a suitable power transformation for the data based on the relation:

$$Y^* = Y^{\lambda}$$

Here the parameter λ is determined using the given data such that SS_E is minimized (details on this method are presented in One Factor Designs).

Multiple Linear Regression Analysis

IN THIS CHAPTER

Multiple Linear Regression Model **66**

Estimating Regression Models Using Least Squares **70**
Example 71
Properties of the Least Square Estimators for Beta 76

Hypothesis Tests in Multiple Linear Regression **77**
Test for Significance of Regression 78
Test on Individual Regression Coefficients (t Test) 82
Test on Subsets of Regression Coefficients (Partial F Test) 84
Types of Extra Sum of Squares 85

Confidence Intervals in Multiple Linear Regression **92**
Confidence Interval on Regression Coefficients 92
Confidence Interval on New Observations 93

Measures of Model Adequacy **94**
Coefficient of Multiple Determination, R^2 94
Residual Analysis 96
Outlying x Observations 100
Influential Observations Detection 100
Lack-of-Fit Test 102

Other Topics in Multiple Linear Regression **103**
Polynomial Regression Models 103
Qualitative Factors 103
Multicollinearity 107

This chapter expands on the analysis of simple linear regression models and discusses the analysis of multiple linear regression models. A major portion of the results displayed in Weibull++ DOE folios are explained in this chapter because these results are associated with multiple linear regression. One of the applications of multiple linear regression models is Response Surface

Methodology (RSM). RSM is a method used to locate the optimum value of the response and is one of the final stages of experimentation. It is discussed in Response Surface Methods. Towards the end of this chapter, the concept of using indicator variables in regression models is explained. Indicator variables are used to represent qualitative factors in regression models. The concept of using indicator variables is important to gain an understanding of ANOVA models, which are the models used to analyze data obtained from experiments. These models can be thought of as first order multiple linear regression models where all the factors are treated as qualitative factors. ANOVA models are discussed in the One Factor Designs and General Full Factorial Designs chapters.

Multiple Linear Regression Model

A linear regression model that contains more than one predictor variable is called a *multiple linear regression model*. The following model is a multiple linear regression model with two predictor variables, x_1 and x_2 .

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

The model is linear because it is linear in the parameters β_0 , β_1 and β_2 . The model describes a plane in the three-dimensional space of Y , x_1 and x_2 . The parameter β_0 is the intercept of this plane. Parameters β_1 and β_2 are referred to as *partial regression coefficients*. Parameter β_1 represents the change in the mean response corresponding to a unit change in x_1 when x_2 is held constant. Parameter β_2 represents the change in the mean response corresponding to a unit change in x_2 when x_1 is held constant. Consider the following example of a multiple linear regression model with two predictor variables, x_1 and x_2 :

$$Y = 30 + 5x_1 + 7x_2 + \epsilon$$

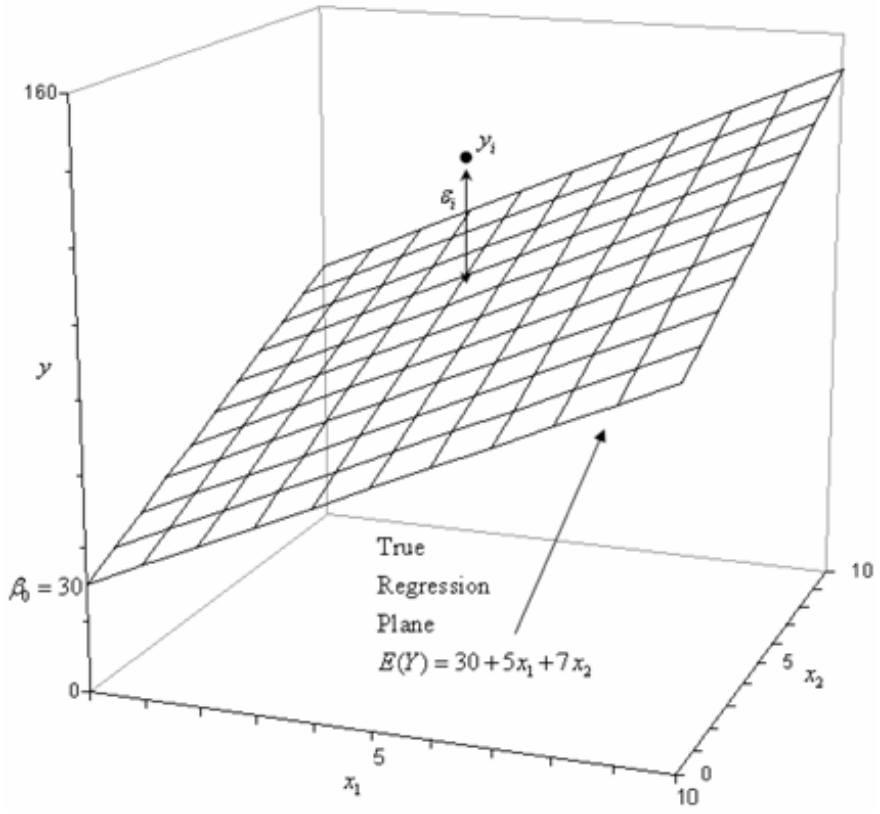
This regression model is a first order multiple linear regression model. This is because the maximum power of the variables in the model is 1. (The regression plane corresponding to this model is shown in the figure below.) Also shown is an observed data point and the corresponding random error, ϵ . The true regression model is usually never known (and therefore the values of the random error terms corresponding to observed data points remain unknown). However, the regression model can be estimated by calculating the parameters of the model for an observed data set. This is explained in Estimating Regression Models Using Least Squares.

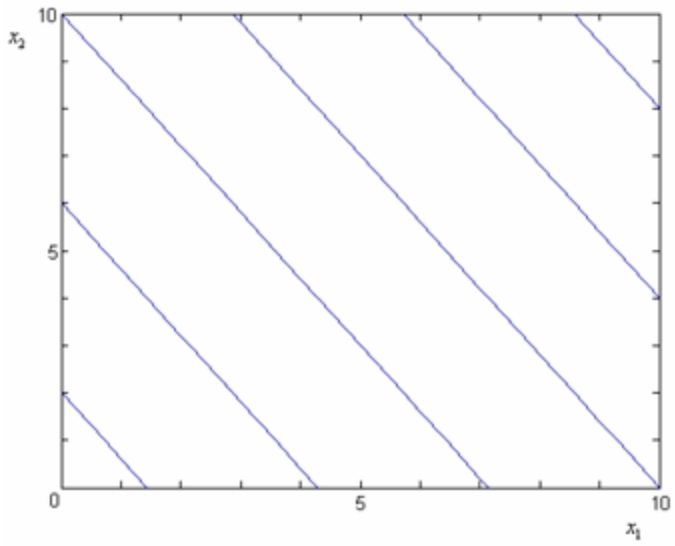
One of the following figures shows the contour plot for the regression model the above equation. The contour plot shows lines of constant mean response values as a function of x_1 and x_2 . The contour lines for the given regression model are straight lines as seen on the plot. Straight contour lines result for first order regression models with no interaction terms.

A linear regression model may also take the following form:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \epsilon$$

A cross-product term, $x_1 x_2$, is included in the model. This term represents an interaction effect between the two variables x_1 and x_2 . Interaction means that the effect produced by a change in the predictor variable on the response depends on the level of the other predictor variable(s). As an example of a linear regression model with interaction, consider the model given by the equation $Y = 30 + 5x_1 + 7x_2 + 3x_1 x_2 + \epsilon$. The regression plane and contour plot for this model are shown in the following two figures, respectively.





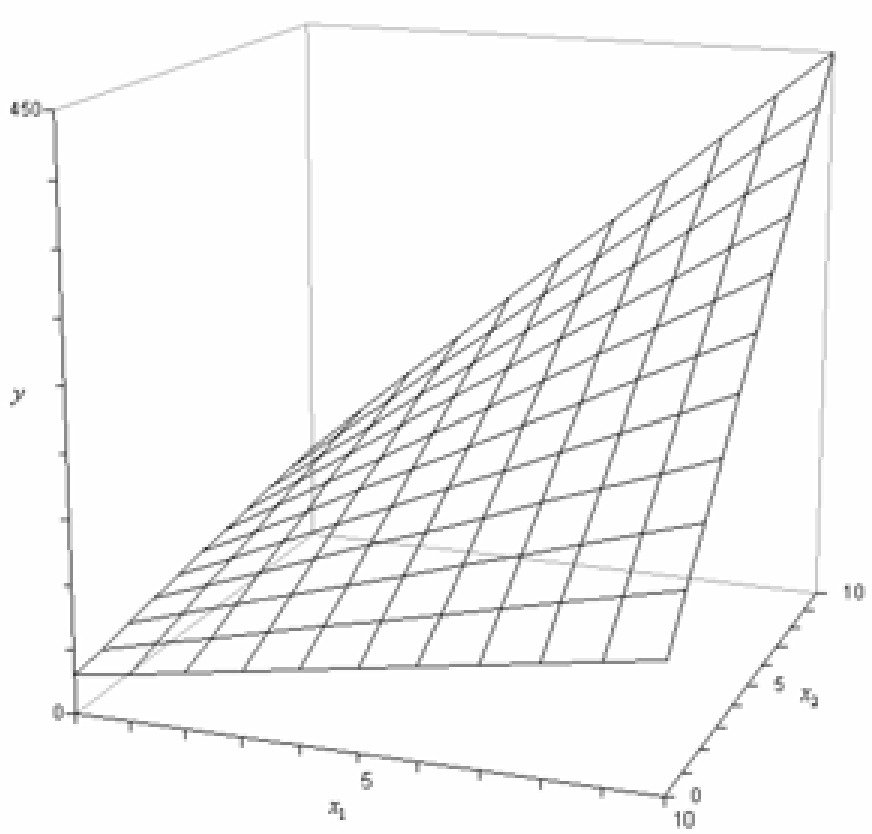
Now consider the regression model shown next:

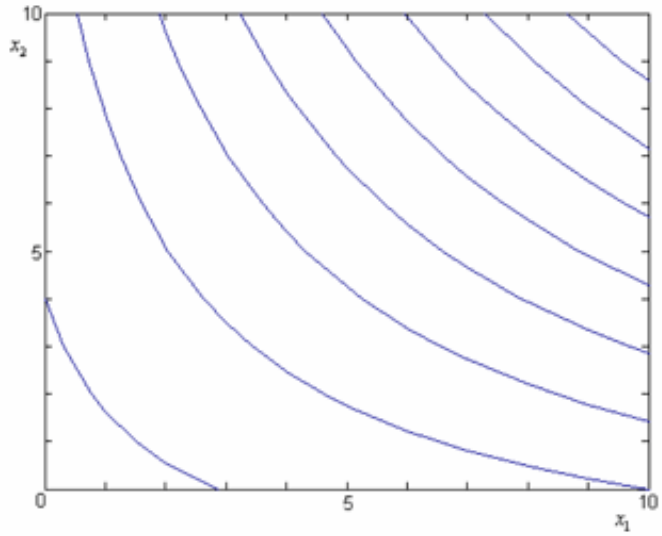
$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_1^3 + \epsilon$$

This model is also a linear regression model and is referred to as a *polynomial regression model*. Polynomial regression models contain squared and higher order terms of the predictor variables making the response surface curvilinear. As an example of a polynomial regression model with an interaction term consider the following equation:

$$Y = 500 + 5x_1 + 7x_2 - 3x_1^2 - 5x_2^2 + 3x_1 x_2 + \epsilon$$

This model is a *second order* model because the maximum power of the terms in the model is two. The regression surface for this model is shown in the following figure. Such regression models are used in RSM to find the optimum value of the response, Y (for details see Response Surface Methods for Optimization). Notice that, although the shape of the regression surface is curvilinear, the regression model is still linear because the model is linear in the parameters. The contour plot for this model is shown in the second of the following two figures.





All multiple linear regression models can be expressed in the following general form:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \epsilon$$

where k denotes the number of terms in the model. For example, the model can be written in the general form using $x_3 = x_1^2$, $x_4 = x_2^3$ and $x_5 = x_1 x_2$ as follows:

$$Y = 500 + 5x_1 + 7x_2 - 3x_3 - 5x_4 + 3x_5 + \epsilon$$

Estimating Regression Models Using Least Squares

Consider a multiple linear regression model with k predictor variables:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \epsilon$$

Let each of the k predictor variables, x_1, x_2 ... x_k, have n levels. Then x_{ij} represents the i th level of the j th predictor variable x_j. For example, x_{51} represents the fifth level of the first predictor variable x_1, while x_{19} represents the first level of the ninth predictor variable, x_9. Observations, y_1, y_2 ... y_n, recorded for each of these n levels can be expressed in the following way:

$$\begin{aligned}
y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \ldots + \beta_k x_{1k} + \epsilon_1 \\
y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \ldots + \beta_k x_{2k} + \epsilon_2 \\
&\vdots \\
y_i &= \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + \epsilon_i \\
&\vdots \\
y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \ldots + \beta_k x_{nk} + \epsilon_n
\end{aligned}$$

The system of n equations shown previously can be represented in matrix notation as follows:

$$y = X\beta + \epsilon$$

where

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdot & \cdot & \cdot & x_{1n} \\ 1 & x_{21} & x_{22} & \cdot & \cdot & \cdot & x_{2n} \\ \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & \cdot & & & & \cdot \\ 1 & x_{n1} & x_{n2} & \cdot & \cdot & \cdot & x_{nn} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ . \\ . \\ \beta_n \end{bmatrix} \text{ and } \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ . \\ . \\ . \\ \epsilon_n \end{bmatrix}$$

The matrix X is referred to as the *design matrix*. It contains information about the levels of the predictor variables at which the observations are obtained. The vector β contains all the regression coefficients. To obtain the regression model, β should be known. β is estimated using least square estimates. The following equation is used:

$$\hat{\beta} = (X'X)^{-1}X'y$$

where $'$ represents the transpose of the matrix while -1 represents the matrix inverse. Knowing the estimates, $\hat{\beta}$, the multiple linear regression model can now be estimated as:

$$\hat{y} = X\hat{\beta}$$

The estimated regression model is also referred to as the *fitted model*. The observations, y_i, may be different from the fitted values $\hat{y}_i$ obtained from this model. The difference between these two values is the residual, e_i. The vector of residuals, e, is obtained as:

$$e = y - \hat{y}$$

The fitted model can also be written as follows, using $\hat{\beta} = (X'X)^{-1}X'y$:

$$\begin{aligned} \hat{y} &= X\hat{\beta} \\ &= X(X'X)^{-1}X'y \\ &= Hy \end{aligned}$$

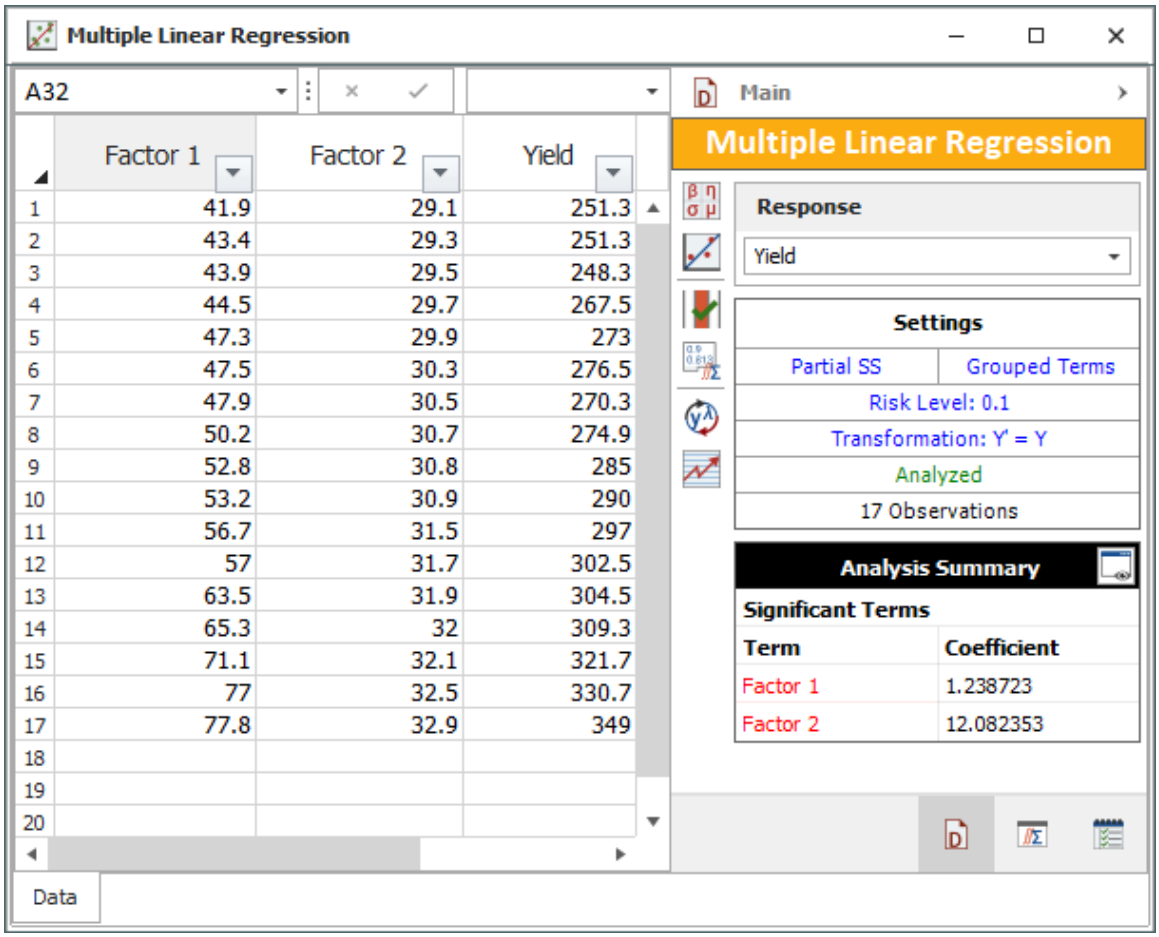
where $H = X(X'X)^{-1}X'$. The matrix, H, is referred to as the hat matrix. It transforms the vector of the observed response values, y, to the vector of fitted values, $\hat{y}$.

Example

An analyst studying a chemical process expects the yield to be affected by the levels of two factors, x_1 and x_2. Observations recorded for various levels of the two factors are shown in the following table. The analyst wants to fit a first order regression model to the data. Interaction between x_1 and x_2 is not expected based on knowledge of similar processes. Units of the factor levels and the yield are ignored for the analysis.

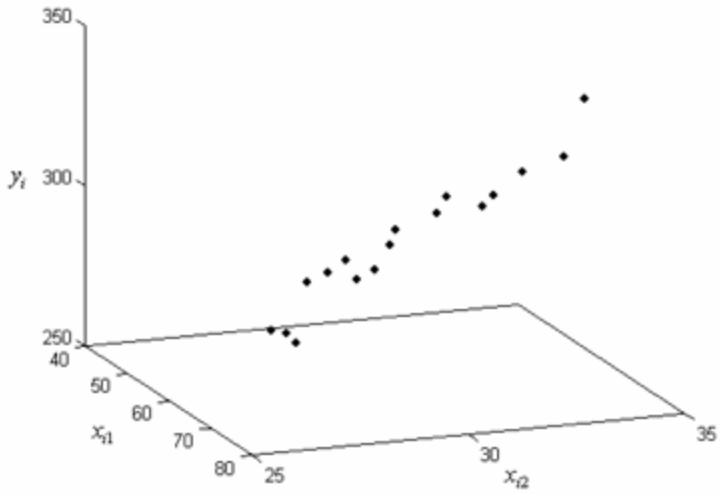
Observation Number	Factor 1 (x_{i1})	Factor 2 (x_{i2})	Yield (y_i)
1	41.9	29.1	251.3
2	43.4	29.3	251.3
3	43.9	29.5	248.3
4	44.5	29.7	267.5
5	47.3	29.9	273.0
6	47.5	30.3	276.5
7	47.9	30.5	270.3
8	50.2	30.7	274.9
9	52.8	30.8	285.0
10	53.2	30.9	290.0
11	56.7	31.5	297.0
12	57.0	31.7	302.5
13	63.5	31.9	304.5
14	65.3	32.0	309.3
15	71.1	32.1	321.7
16	77.0	32.5	330.7
17	77.8	32.9	349.0

The data of the above table can be entered into the Weibull++ DOE folio using the multiple linear regression folio tool as shown in the following figure.



	Factor 1	Factor 2	Yield
1	41.9	29.1	251.3
2	43.4	29.3	251.3
3	43.9	29.5	248.3
4	44.5	29.7	267.5
5	47.3	29.9	273
6	47.5	30.3	276.5
7	47.9	30.5	270.3
8	50.2	30.7	274.9
9	52.8	30.8	285
10	53.2	30.9	290
11	56.7	31.5	297
12	57	31.7	302.5
13	63.5	31.9	304.5
14	65.3	32	309.3
15	71.1	32.1	321.7
16	77	32.5	330.7
17	77.8	32.9	349
18			
19			
20			

A scatter plot for the data is shown next.



The first order regression model applicable to this data set having two predictor variables is:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

where the dependent variable, Y, represents the yield and the predictor variables, x_1 and x_2, represent the two factors respectively. The X and y matrices for the data can be obtained as:

$$X = \begin{bmatrix} 1 & 41.9 & 29.1 \\ 1 & 43.4 & 29.3 \\ . & . & . \\ . & . & . \\ . & . & . \\ 1 & 77.8 & 32.9 \end{bmatrix} \quad y = \begin{bmatrix} 251.3 \\ 251.3 \\ . \\ . \\ . \\ 349.0 \end{bmatrix}$$

The least square estimates, $\hat{\beta}$, can now be obtained:

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'y \\ &= \begin{bmatrix} 17 & 941 & 525.3 \\ 941 & 54270 & 29286 \\ 525.3 & 29286 & 16254 \end{bmatrix}^{-1} \begin{bmatrix} 4902.8 \\ 276610 \\ 152020 \end{bmatrix} \\ &= \begin{bmatrix} -153.51 \\ 1.24 \\ 12.08 \end{bmatrix} \end{aligned}$$

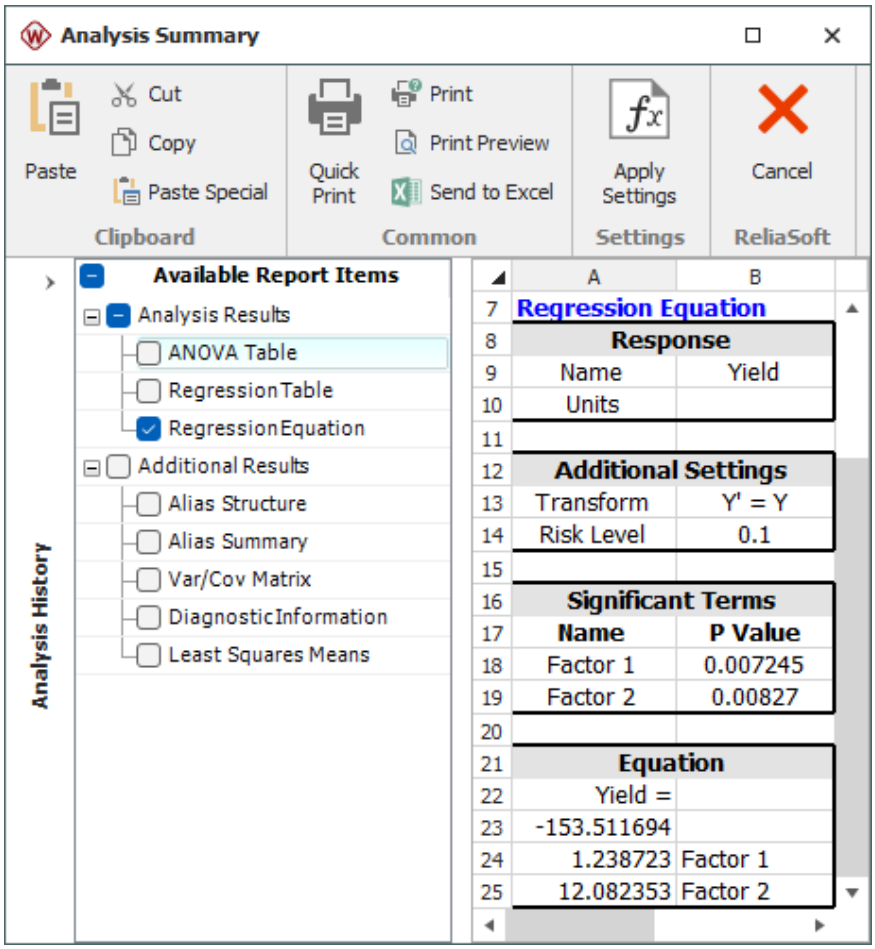
Thus:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} -153.51 \\ 1.24 \\ 12.08 \end{bmatrix}$$

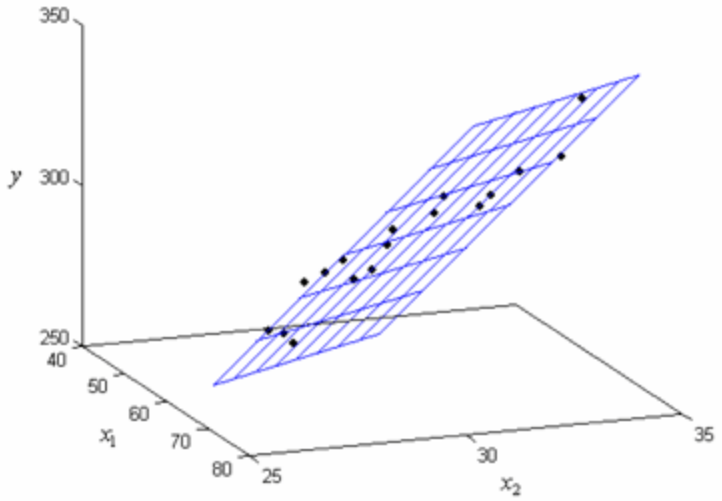
and the estimated regression coefficients are $\hat{\beta}_0 = -153.51$, $\hat{\beta}_1 = 1.24$ and $\hat{\beta}_2 = 12.08$. The fitted regression model is:

$$\begin{aligned} \hat{y} &= \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \\ &= -153.5 + 1.24x_1 + 12.08x_2 \end{aligned}$$

The fitted regression model can be viewed in the Weibull++ DOE folio, as shown next.



A plot of the fitted regression plane is shown in the following figure.



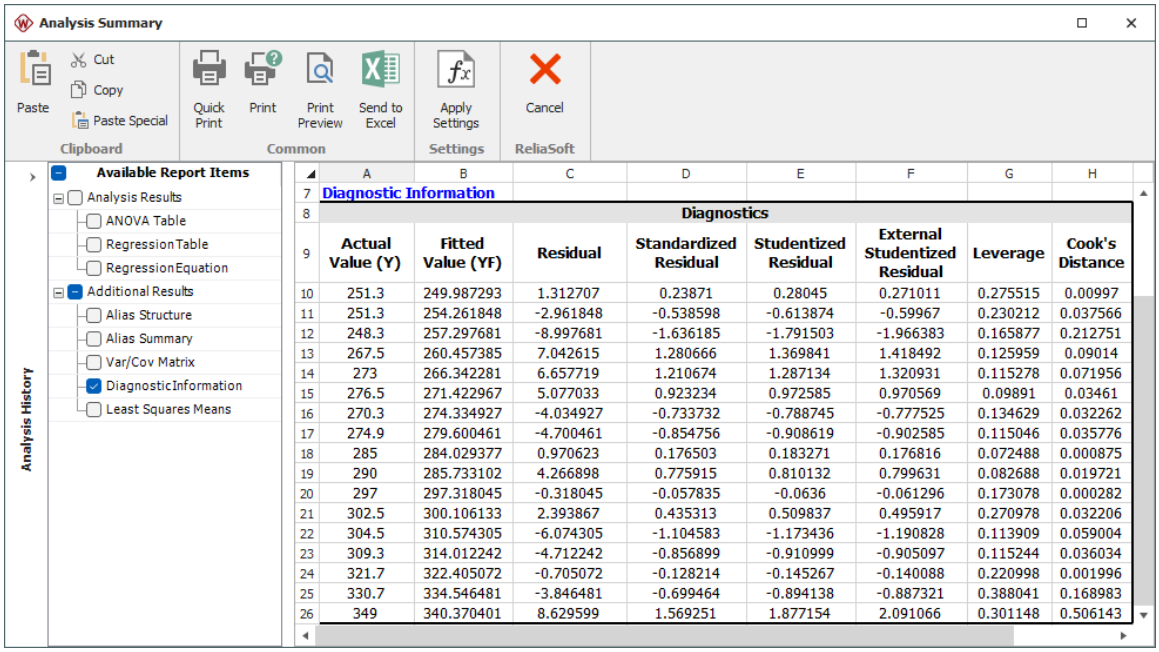
The fitted regression model can be used to obtain fitted values, $\hat{y}_i$, corresponding to an observed response value, y_i. For example, the fitted value corresponding to the fifth observation is:

$$\begin{aligned}\hat{y}_i &= -153.5 + 1.24x_{i1} + 12.08x_{i2}\\ \hat{y}_5 &= -153.5 + 1.24x_{51} + 12.08x_{52}\\ &= -153.5 + 1.24(47.3) + 12.08(29.9)\\ &= 266.3\end{aligned}$$

The observed fifth response value is $y_5 = 273.0$. The residual corresponding to this value is:

$$\begin{aligned}e_i &= y_i - \hat{y}_i\\ e_5 &= y_5 - \hat{y}_5\\ &= 273.0 - 266.3\\ &= 6.7\end{aligned}$$

In Weibull++ DOE folios, fitted values and residuals are shown in the Diagnostic Information table of the detailed summary of results. The values are shown in the following figure.



Diagnostic Information

Diagnostics

Actual Value (Y)	Fitted Value (YF)	Residual	Standardized Residual	Studentized Residual	External Studentized Residual	Leverage	Cook's Distance
251.3	249.987293	1.312707	0.23871	0.28045	0.271011	0.275515	0.00997
251.3	254.261848	-2.961848	-0.538598	-0.613874	-0.59967	0.230212	0.037566
248.3	257.297681	-8.997681	-1.636185	-1.791503	-1.966383	0.165877	0.212751
267.5	260.457385	7.042615	1.280666	1.369841	1.418492	0.125959	0.09014
273	266.342281	6.657719	1.210674	1.287134	1.320931	0.115278	0.071956
276.5	271.422967	5.077033	0.923234	0.972585	0.970569	0.09891	0.03461
270.3	274.334927	-4.034927	-0.733732	-0.788745	-0.777525	0.134629	0.032262
274.9	279.600461	-4.700461	-0.854756	-0.908619	-0.902585	0.115046	0.035776
285	284.029377	0.970623	0.176503	0.183271	0.176816	0.072488	0.000875
290	285.733102	4.266898	0.775915	0.810132	0.799631	0.082688	0.019721
297	297.318045	-0.318045	-0.057835	-0.0636	-0.061296	0.173078	0.000282
302.5	300.106133	2.393867	0.435313	0.509837	0.495917	0.270978	0.032206
304.5	310.574305	-6.074305	-1.104583	-1.173436	-1.190828	0.113909	0.059004
309.3	314.012242	-4.712242	-0.856899	-0.910999	-0.905097	0.115244	0.036034
321.7	322.405072	-0.705072	-0.128214	-0.145267	-0.140088	0.220998	0.001996
330.7	334.546481	-3.846481	-0.699464	-0.894138	-0.887321	0.388041	0.168983
349	340.370401	8.629599	1.569251	1.877154	2.091066	0.301148	0.506143

The fitted regression model can also be used to predict response values. For example, to obtain the response value for a new observation corresponding to 47 units of x_1 and 31 units of x_2, the value is calculated using:

$$\begin{aligned}\hat{y}(47,31) &= -153.5 + 1.24(47) + 12.08(31)\\ &= 279.26\end{aligned}$$

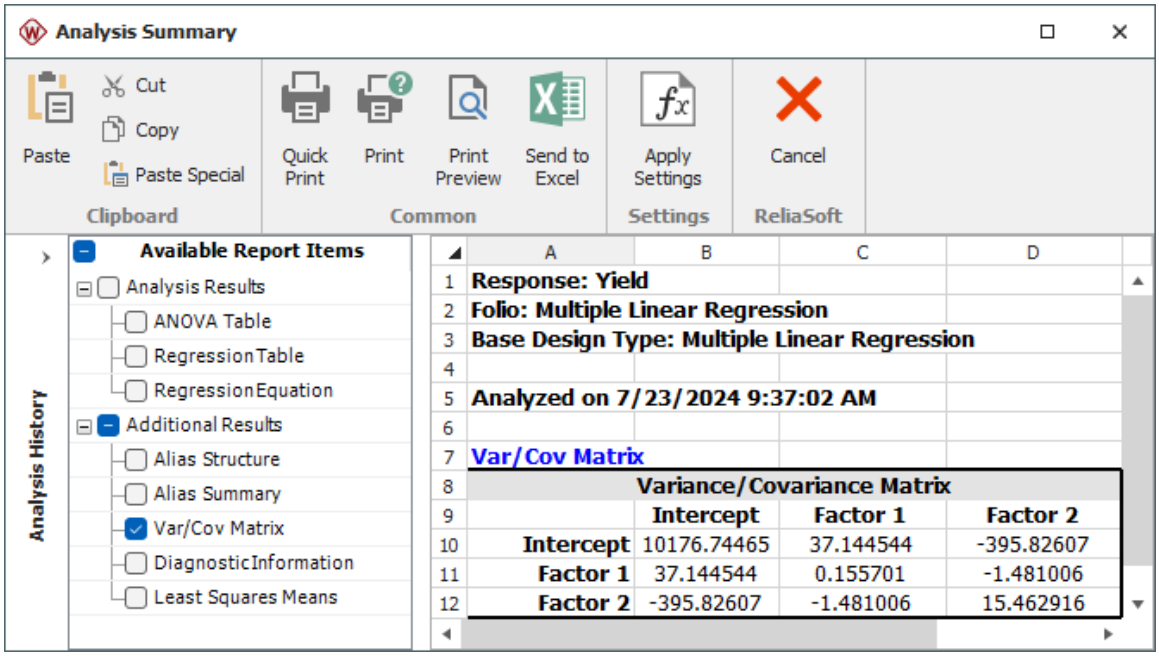
Properties of the Least Square Estimators for Beta

The least square estimates, $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ ... $\hat{\beta}_k$, are unbiased estimators of β_0, β_1, β_2 ... β_k, provided that the random error terms, ϵ_i, are normally and independently distributed. The vari-

$(X'X)^{-1}$ matrix. The variance-covariance matrix of the estimated regression coefficients is obtained as follows:

$$C = \hat{\sigma}^2 (X'X)^{-1}$$

C is a symmetric matrix whose diagonal elements, C_{jj}, represent the variance of the estimated j th regression coefficient, $\hat{\beta}_j$. The off-diagonal elements, C_{ij}, represent the covariance between the i th and j th estimated regression coefficients, $\hat{\beta}_i$ and $\hat{\beta}_j$. The value of $\hat{\sigma}^2$ is obtained using the error mean square, MS_E. The variance-covariance matrix for the data in the table (see Estimating Regression Models Using Least Squares) can be viewed in the DOE folio, as shown next.



Calculations to obtain the matrix are given in this example. The positive square root of C_{jj} represents the estimated standard deviation of the j th regression coefficient, $\hat{\beta}_j$, and is called the estimated standard error of $\hat{\beta}_j$ (abbreviated $se(\hat{\beta}_j)$).

$$se(\hat{\beta}_j) = \sqrt{C_{jj}}$$

Hypothesis Tests in Multiple Linear Regression

This section discusses hypothesis tests on the regression coefficients in multiple linear regression. As in the case of simple linear regression, these tests can only be carried out if it can be assumed that the random error terms, ϵ_i, are normally and independently distributed with a

. Three types of hypothesis tests can be carried out for multiple linear regression models:

1. Test for significance of regression: This test checks the significance of the whole regression model.
2. t test: This test checks the significance of individual regression coefficients.
3. F test: This test can be used to simultaneously check the significance of a number of regression coefficients. It can also be used to test individual coefficients.

Test for Significance of Regression

The test for significance of regression in the case of multiple linear regression analysis is carried out using the analysis of variance. The test is used to check if a linear statistical relationship exists between the response variable and at least one of the predictor variables. The statements for the hypotheses are:

$$H_0: \quad \beta_1 = \beta_2 = \ldots = \beta_k = 0$$
$$H_1: \quad \beta_j \neq 0 \text{ for at least one } j$$

The test for H_0 is carried out using the following statistic:

$$F_0 = \frac{MS_R}{MS_E}$$

where MS_R is the regression mean square and MS_E is the error mean square. If the null hypothesis, H_0, is true then the statistic F_0 follows the F distribution with k degrees of freedom in the numerator and $n-$ ($k+1$) degrees of freedom in the denominator. The null hypothesis, H_0, is rejected if the calculated statistic, F_0, is such that:

$$F_0 > f_{\alpha,k,n-(k+1)}$$

Calculation of the Statistic *F_0*

To calculate the statistic F_0, the mean squares MS_R and MS_E must be known. As explained in Simple Linear Regression Analysis, the mean squares are obtained by dividing the sum of squares by their degrees of freedom. For example, the total mean square, MS_T, is obtained as follows:

$$MS_T = \frac{SS_T}{dof(SS_T)}$$

where SS_T is the total sum of squares and $dof(SS_T)$ is the number of degrees of freedom associated with SS_T. In multiple linear regression, the following equation is used to calculate SS_T :

$$SS_T = y' \left[I - (\frac{1}{n})J \right] y$$

where n is the total number of observations, y is the vector of observations (that was defined in Estimating Regression Models Using Least Squares), I is the identity matrix of order n and J represents an $n \times n$ square matrix of ones. The number of degrees of freedom associated with SS_T, $dof(SS_T)$, is ($n-1$). Knowing SS_T and $dof(SS_T)$ the total mean square, MS_T, can be calculated.

The regression mean square, MS_R, is obtained by dividing the regression sum of squares, SS_R, by the respective degrees of freedom, $dof(SS_R)$, as follows:

$$MS_R = \frac{SS_R}{dof(SS_R)}$$

The regression sum of squares, SS_R, is calculated using the following equation:

$$SS_R = y' \left[H - (\frac{1}{n})J \right] y$$

where n is the total number of observations, y is the vector of observations, H is the hat matrix and J represents an $n \times n$ square matrix of ones. The number of degrees of freedom associated with SS_R, $dof(SS_E)$, is k, where k is the number of predictor variables in the model. Knowing SS_R and $dof(SS_R)$ the regression mean square, MS_R, can be calculated. The error mean square, MS_E, is obtained by dividing the error sum of squares, SS_E, by the respective degrees of freedom, $dof(SS_E)$, as follows:

$$MS_E = \frac{SS_E}{dof(SS_E)}$$

The error sum of squares, SS_E, is calculated using the following equation:

$$SS_E = y'(I - H)y$$

where y is the vector of observations, I is the identity matrix of order n and H is the hat matrix. The number of degrees of freedom associated with SS_E, $dof(SS_E)$, is $n-(k+1)$, where n is the total number of observations and k is the number of predictor variables in the model. Knowing SS_E and $dof(SS_E)$, the error mean square, MS_E, can be calculated. The error mean square is an estimate of the variance, σ^2, of the random error terms, ϵ_i.

$$\hat{\sigma}^2 = MS_E$$

Example

The test for the significance of regression, for the regression model obtained for the data in the table (see Estimating Regression Models Using Least Squares), is illustrated in this example. The null hypothesis for the model is:

$$H_0 : \beta_1 = \beta_2 = 0$$

The statistic to test H_0 is:

$$F_0 = \frac{MS_R}{MS_E}$$

To calculate F_0, first the sum of squares are calculated so that the mean squares can be obtained. Then the mean squares are used to calculate the statistic F_0 to carry out the significance test. The regression sum of squares, SS_R, can be obtained as:

$$SS_R = y' \left[H - (\frac{1}{n})J \right] y$$

The hat matrix, H is calculated as follows using the design matrix X from the previous example:

$$H = X(X'X)^{-1}X'$$

$$= \begin{bmatrix} 0.27552 & 0.25154 & . & . & -0.04030 \\ 0.25154 & 0.23021 & . & . & -0.029120 \\ . & . & . & . & . \\ . & . & . & . & . \\ -0.04030 & -0.02920 & . & . & 0.30115 \end{bmatrix}$$

Knowing y, H and J, the regression sum of squares, SS_R, can be calculated:

$$SS_R = y' \left[H - (\frac{1}{n})J \right] y$$

$$= 12816.35$$

The degrees of freedom associated with SS_R is k, which equals to a value of two since there are two predictor variables in the data in the table (see Multiple Linear Regression Analysis). Therefore, the regression mean square is:

$$\begin{aligned} MS_R &= \frac{SS_R}{dof(SS_R)} \\ &= \frac{12816.35}{2} \\ &= 6408.17 \end{aligned}$$

Similarly to calculate the error mean square, MS_E, the error sum of squares, SS_E, can be obtained as:

$$\begin{aligned} SS_E &= y'[I-H]y \\ &= 423.37 \end{aligned}$$

The degrees of freedom associated with SS_E is $n-(k+1)$. Therefore, the error mean square, MS_E, is:

$$\begin{aligned} MS_E &= \frac{SS_E}{dof(SS_E)} \\ &= \frac{SS_E}{(n-(k+1))} \\ &= \frac{423.37}{(17-(2+1))} \\ &= 30.24 \end{aligned}$$

The statistic to test the significance of regression can now be calculated as:

$$\begin{aligned} f_0 &= \frac{MS_R}{MS_E} \\ &= \frac{6408.17}{423.37/(17-3)} \\ &= 211.9 \end{aligned}$$

The critical value for this test, corresponding to a significance level of 0.1, is:

$$\begin{aligned} f_{\alpha,k,n-(k+1)} &= f_{0.1,2,14} \\ &= 2.726 \end{aligned}$$

Since $f_0 > f_{0.1,2,14}$, $H_0 : \beta_1 = \beta_2 = 0$ is rejected and it is concluded that at least one coefficient out of β_1 and β_2 is significant. In other words, it is concluded that a regression model exists between yield and either one or both of the factors in the table. The analysis of variance is summarized in the following table.

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	*F* Statistic	*P* Value
Regression	2	12816.35	6408.17	211.9	0.00
Error	14	423.37	30.24		
Total	16	13239.72			

Test on Individual Regression Coefficients (*t*Test)

The t test is used to check the significance of individual regression coefficients in the multiple linear regression model. Adding a significant variable to a regression model makes the model more effective, while adding an unimportant variable may make the model worse. The hypothesis statements to test the significance of a particular regression coefficient, β_j, are:

$$\begin{aligned} H_0 &: \quad \beta_j = 0 \\ H_1 &: \quad \beta_j \neq 0 \end{aligned}$$

The test statistic for this test is based on the t distribution (and is similar to the one used in the case of simple linear regression models in Simple Linear Regression Anaysis):

$$T_0 = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$$

where the standard error, $se(\hat{\beta}_j)$, is obtained. The analyst would fail to reject the null hypothesis if the test statistic lies in the acceptance region:

$$-t_{\alpha/2,n-2} < T_0 < t_{\alpha/2,n-2}$$

This test measures the contribution of a variable while the remaining variables are included in the model. For the model $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$, if the test is carried out for β_1, then the test will check the significance of including the variable x_1 in the model that contains x_2 and x_3 (i.e., the model $\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$). Hence the test is also referred to as partial or marginal test. In DOE folios, this test is displayed in the Regression Information table.

Example

The test to check the significance of the estimated regression coefficients for the data is illustrated in this example. The null hypothesis to test the coefficient β_2 is:

$$H_0 : \beta_2 = 0$$

The null hypothesis to test β_1 can be obtained in a similar manner. To calculate the test statistic, T_0, we need to calculate the standard error. In the example, the value of the error mean square, MS_E, was obtained as 30.24. The error mean square is an estimate of the variance, σ^2.

Therefore:

$$\begin{aligned} \hat{\sigma}^2 &= MS_E \\ &= 30.24 \end{aligned}$$

The variance-covariance matrix of the estimated regression coefficients is:

$$
\begin{aligned}
C &= \hat{\sigma}^2 (X'X)^{-1} \\
&= 30.24 \begin{bmatrix} 336.5 & 1.2 & -13.1 \\ 1.2 & 0.005 & -0.049 \\ -13.1 & -0.049 & 0.5 \end{bmatrix} \\
&= \begin{bmatrix} 10176.75 & 37.145 & -395.83 \\ 37.145 & 0.1557 & -1.481 \\ -395.83 & -1.481 & 15.463 \end{bmatrix}
\end{aligned}
$$

From the diagonal elements of C, the estimated standard error for $\hat{\beta}_1$ and $\hat{\beta}_2$ is:

$$
\begin{aligned}
se(\hat{\beta}_1) &= \sqrt{0.1557} = 0.3946 \\
se(\hat{\beta}_2) &= \sqrt{15.463} = 3.93
\end{aligned}
$$

The corresponding test statistics for these coefficients are:

$$
\begin{aligned}
(t_0)_{\hat{\beta}_1} &= \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} = \frac{1.24}{0.3946} = 3.1393 \\
(t_0)_{\hat{\beta}_2} &= \frac{\hat{\beta}_2}{se(\hat{\beta}_2)} = \frac{12.08}{3.93} = 3.0726
\end{aligned}
$$

The critical values for the present t test at a significance of 0.1 are:

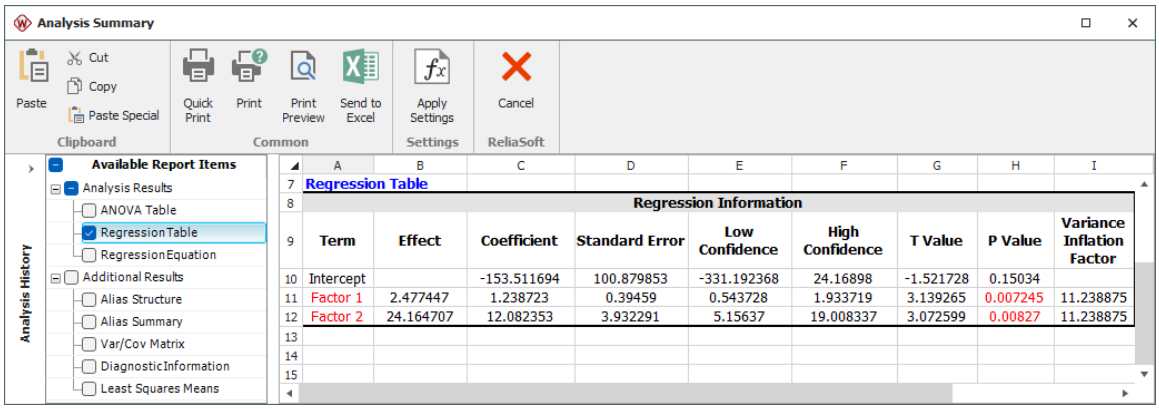
$$
\begin{aligned}
t_{\alpha/2,n-(k+1)} &= t_{0.05,14} = 1.761 \\
-t_{\alpha/2,n-(k+1)} &= -t_{0.05,14} = -1.761
\end{aligned}
$$

Considering $\hat{\beta}_2$, it can be seen that $(t_0)_{\hat{\beta}_2}$ does not lie in the acceptance region of $-t_{0.05,14} < t_0 < t_{0.05,14}$. The null hypothesis, $H_0 : \beta_2 = 0$, is rejected and it is concluded that β_2 is significant at $\alpha = 0.1$. This conclusion can also be arrived at using the p value noting that the hypothesis is two-sided. The p value corresponding to the test statistic, $(t_0)_{\hat{\beta}_2} = 3.0726$, based on the t distribution with 14 degrees of freedom is:

$$
\begin{aligned}
p\ value &= 2 \times (1 - P(T \leq |t_0|) \\
&= 2 \times (1 - 0.9959) \\
&= 0.0083
\end{aligned}
$$

Since the p value is less than the significance, $\alpha = 0.1$, it is concluded that β_2 is significant. The hypothesis test on β_1 can be carried out in a similar manner.

As explained in Simple Linear Regression Analysis, in DOE folios, the information related to the t test is displayed in the Regression Information table as shown in the figure below.



Regression Table

Regression Information

Term	Effect	Coefficient	Standard Error	Low Confidence	High Confidence	T Value	P Value	Variance Inflation Factor
Intercept		-153.511694	100.879853	-331.192368	24.16898	-1.521728	0.15034	
Factor 1	2.477447	1.238723	0.39459	0.543728	1.933719	3.139265	0.007245	11.238875
Factor 2	24.164707	12.082353	3.932291	5.15637	19.008337	3.072599	0.00827	11.238875

In this table, the t test for β_2 is displayed in the row for the term Factor 2 because β_2 is the coefficient that represents this factor in the regression model. Columns labeled Standard Error, T Value and P Value represent the standard error, the test statistic for the t test and the p value for the t test, respectively. These values have been calculated for β_2 in this example. The Coefficient column represents the estimate of regression coefficients. These values are calculated as shown in this example. The Effect column represents values obtained by multiplying the coefficients by a factor of 2. This value is useful in the case of two factor experiments and is explained in Two-Level Factorial Experiments. Columns labeled Low Confidence and High Confidence represent the limits of the confidence intervals for the regression coefficients and are explained in Confidence Intervals in Multiple Linear Regression. The Variance Inflation Factor column displays values that give a measure of *multicollinearity*. This is explained in Multicollinearity.

Test on Subsets of Regression Coefficients (Partial *F* Test)

This test can be considered to be the general form of the t test mentioned in the previous section. This is because the test simultaneously checks the significance of including many (or even one) regression coefficients in the multiple linear regression model. Adding a variable to a model increases the regression sum of squares, SS_R. The test is based on this increase in the regression sum of squares. The increase in the regression sum of squares is called the *extra sum of squares*. Assume that the vector of the regression coefficients, β, for the multiple linear regression model, $y = X\beta + \epsilon$, is partitioned into two vectors with the second vector, θ_2, containing the last r regression coefficients, and the first vector, θ_1, containing the first ($k+1-r$) coefficients as follows:

$$\beta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$$

with:

$$\theta_1 = [\beta_0, \beta_1 \ldots \beta_{k-r}]' \text{ and } \theta_2 = [\beta_{k-r+1}, \beta_{k-r+2} \ldots \beta_k]'$$

The hypothesis statements to test the significance of adding the regression coefficients in θ_2 to a model containing the regression coefficients in θ_1 may be written as:

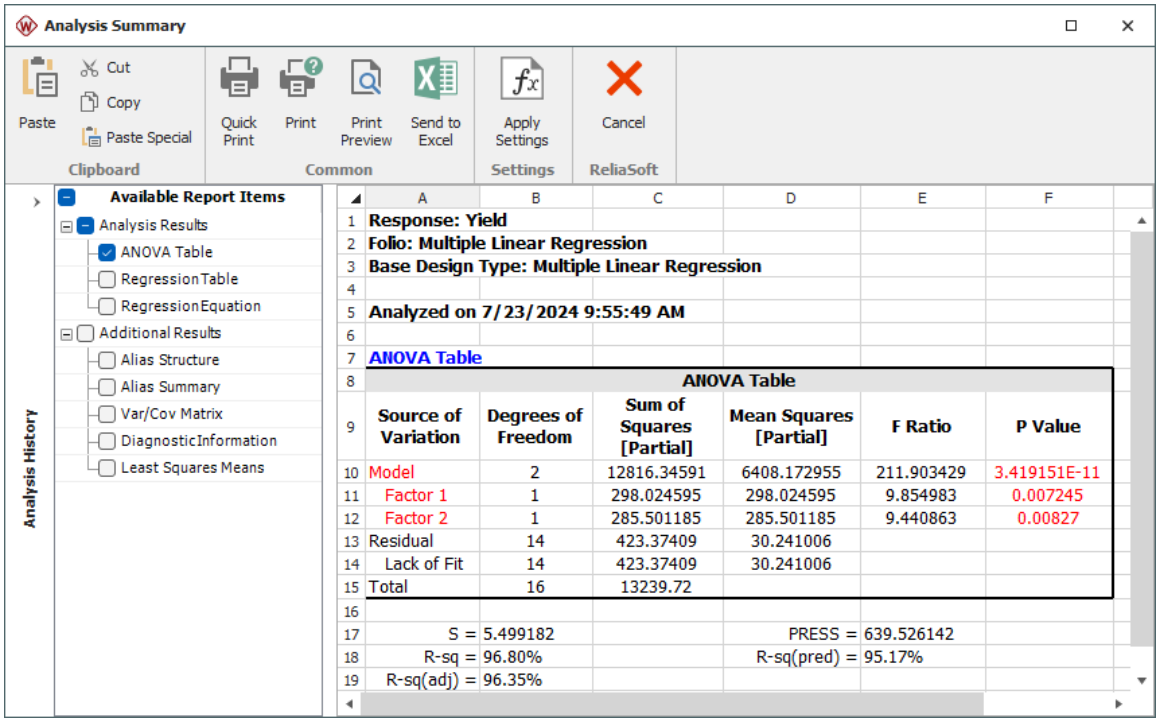
$$H_0: \quad \theta_2 = 0$$
$$H_1: \quad \theta_2 \neq 0$$

The test statistic for this test follows the F distribution and can be calculated as follows:

$$F_0 = \frac{SS_R(\theta_2|\theta_1)/r}{MS_E}$$

where $SS_R(\theta_2|\theta_1)$ is the the increase in the regression sum of squares when the variables corresponding to the coefficients in θ_2 are added to a model already containing θ_1, and MS_E is obtained from the equation given in Simple Linear Regression Analysis. The value of the extra sum of squares is obtained as explained in the next section.

The null hypothesis, H_0, is rejected if $F_0 > f_{\alpha,r,n-(k+1)}$. Rejection of H_0 leads to the conclusion that at least one of the variables in x_{k-r+1}, x_{k-r+2} ... x_k contributes significantly to the regression model. In a DOE folio, the results from the partial F test are displayed in the ANOVA table.



Response: Yield
Folio: Multiple Linear Regression
Base Design Type: Multiple Linear Regression

Analyzed on 7/23/2024 9:55:49 AM

ANOVA Table

ANOVA Table					
Source of Variation	Degrees of Freedom	Sum of Squares [Partial]	Mean Squares [Partial]	F Ratio	P Value
Model	2	12816.34591	6408.172955	211.903429	3.419151E-11
Factor 1	1	298.024595	298.024595	9.854983	0.007245
Factor 2	1	285.501185	285.501185	9.440863	0.00827
Residual	14	423.37409	30.241006		
Lack of Fit	14	423.37409	30.241006		
Total	16	13239.72			

S = 5.499182 PRESS = 639.526142
R-sq = 96.80% R-sq(pred) = 95.17%
R-sq(adj) = 96.35%

Types of Extra Sum of Squares

The extra sum of squares can be calculated using either the partial (or adjusted) sum of squares or the sequential sum of squares. The type of extra sum of squares used affects the calculation

of the test statistic for the partial F test described above. In DOE folios, selection for the type of extra sum of squares is available as shown in the figure below. The partial sum of squares is used as the default setting. The reason for this is explained in the following section on the partial sum of squares.

Partial Sum of Squares

The partial sum of squares for a term is the extra sum of squares when all terms, except the term under consideration, are included in the model. For example, consider the model:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \epsilon$$

The sum of squares of regression of this model is denoted by $SS_R(\beta_0, \beta_1, \beta_2, \beta_{12})$. Assume that we need to know the partial sum of squares for β_2. The partial sum of squares for β_2 is the increase in the regression sum of squares when β_2 is added to the model. This increase is the difference in the regression sum of squares for the full model of the equation given above and the model that includes all terms except β_2. These terms are β_0, β_1 and β_{12}. The model that contains these terms is:

$$Y = \beta_0 + \beta_1 x_1 + \beta_{12} x_1 x_2 + \epsilon$$

The sum of squares of regression of this model is denoted by $SS_R(\beta_0, \beta_1, \beta_{12})$. The partial sum of squares for β_2 can be represented as $SS_R(\beta_2|\beta_0, \beta_1, \beta_{12})$ and is calculated as follows:

$$SS_R(\beta_2|\beta_0, \beta_1, \beta_{12}) = \quad SS_R(\beta_0, \beta_1, \beta_2, \beta_{12}) - SS_R(\beta_0, \beta_1, \beta_{12})$$

For the present case, $\theta_2 = [\beta_2]'$ and $\theta_1 = [\beta_0, \beta_1, \beta_{12}]'$. It can be noted that for the partial sum of squares β_1 contains all coefficients other than the coefficient being tested.

A Weibull++ DOE folio has the partial sum of squares as the default selection. This is because the t test is a partial test, i.e., the t test on an individual coefficient is carried by assuming that all the remaining coefficients are included in the model (similar to the way the partial sum of squares is calculated). The results from the t test are displayed in the Regression Information table. The results from the partial F test are displayed in the ANOVA table. To keep the results in the two tables consistent with each other, the partial sum of squares is used as the default selection for the results displayed in the ANOVA table. The partial sum of squares for all terms of a model may not add up to the regression sum of squares for the full model when the regression coefficients are correlated. If it is preferred that the extra sum of squares for all terms in the model always add up to the regression sum of squares for the full model then the sequential sum of squares should be used.

EXAMPLE

This example illustrates the F test using the partial sum of squares. The test is conducted for the coefficient β_1 corresponding to the predictor variable x_1 for the data. The regression model used for this data set in the example is:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

The null hypothesis to test the significance of β_1 is:

$$H_0 : \beta_1 = 0$$

The statistic to test this hypothesis is:

$$F_0 = \frac{SS_R(\beta_1|\beta_2)/r}{MS_E}$$

where $SS_R(\beta_1|\beta_2)$ represents the partial sum of squares for β_1, r represents the number of degrees of freedom for $SS_R(\beta_1|\beta_2)$ (which is one because there is just one coefficient, β_1, being tested) and MS_E is the error mean square and has been calculated in the second example as 30.24.

The partial sum of squares for β_1 is the difference between the regression sum of squares for the full model, $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$, and the regression sum of squares for the model excluding β_1, $Y = \beta_0 + \beta_2 x_2 + \epsilon$. The regression sum of squares for the full model has been calculated in the second example as 12816.35. Therefore:

$$SS_R(\beta_0, \beta_1, \beta_2) = 12816.35$$

The regression sum of squares for the model $Y = \beta_0 + \beta_2 x_2 + \epsilon$ is obtained as shown next. First the design matrix for this model, X_{β_0,β_2}, is obtained by dropping the second column in the design matrix of the full model, X (the full design matrix, X, was obtained in the example). The second column of X corresponds to the coefficient β_1 which is no longer in the model. Therefore, the design matrix for the model, $Y = \beta_0 + \beta_2 x_2 + \epsilon$, is:

$$X_{\beta_0,\beta_2} = \begin{bmatrix} 1 & 29.1 \\ 1 & 29.3 \\ . & . \\ . & . \\ 1 & 32.9 \end{bmatrix}$$

The hat matrix corresponding to this design matrix is H_{β_0,β_2} . It can be calculated using $H_{\beta_0,\beta_2} = X_{\beta_0,\beta_2}(X'_{\beta_0,\beta_2}X_{\beta_0,\beta_2})^{-1}X'_{\beta_0,\beta_2}$. Once H_{β_0,β_2} is known, the regression sum of squares for the model $Y = \beta_0 + \beta_2 x_2 + \epsilon$, can be calculated as:

$$\begin{aligned} SS_R(\beta_0,\beta_2) &= y'\left[H_{\beta_0,\beta_2} - (\frac{1}{n})J\right]y \\ &= 12518.32 \end{aligned}$$

Therefore, the partial sum of squares for β_1 is:

$$\begin{aligned} SS_R(\beta_1|\beta_2) &= SS_R(\beta_0,\beta_1,\beta_2) - SS_R(\beta_0,\beta_2) \\ &= 12816.35 - 12518.32 \\ &= 298.03 \end{aligned}$$

Knowing the partial sum of squares, the statistic to test the significance of β_1 is:

$$\begin{aligned} f_0 &= \frac{SS_R(\beta_1|\beta_2)/r}{MS_E} \\ &= \frac{298.03/1}{30.24} \\ &= 9.855 \end{aligned}$$

The p value corresponding to this statistic based on the F distribution with 1 degree of freedom in the numerator and 14 degrees of freedom in the denominator is:

$$\begin{aligned} p\,value &= 1 - P(F \leq f_0) \\ &= 1 - 0.9928 \\ &= 0.0072 \end{aligned}$$

Assuming that the desired significance is 0.1, since p value < 0.1, $H_0 : \beta_1 = 0$ is rejected and it can be concluded that β_1 is significant. The test for β_2 can be carried out in a similar manner. In the results obtained from the DOE folio, the calculations for this test are displayed in the ANOVA table as shown in the following figure. Note that the conclusion obtained in this example can also be obtained using the t test as explained in the example in Test on Individual Regression Coefficients (t Test). The ANOVA and Regression Information tables in the DOE folio represent two different ways to test for the significance of the variables included in the multiple linear regression model.

Sequential Sum of Squares

The sequential sum of squares for a coefficient is the extra sum of squares when coefficients are added to the model in a sequence. For example, consider the model:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_3 x_3 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3 + \epsilon$$

The sequential sum of squares for β_{13} is the increase in the sum of squares when β_{13} is added to the model observing the sequence of the equation given above. Therefore this extra sum of squares can be obtained by taking the difference between the regression sum of squares for the model after β_{13} was added and the regression sum of squares for the model before β_{13} was added to the model. The model after β_{13} is added is as follows:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_3 x_3 + \beta_{13} x_1 x_3 + \epsilon$$

This is because to maintain the sequence all coefficients preceding β_{13} must be included in the model. These are the coefficients β_0 , β_1 , β_2 , β_{12} and β_3 . Similarly the model before β_{13} is added must contain all coefficients of the equation given above except β_{13} . This model can be obtained as follows:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_3 x_3 + \epsilon$$

The sequential sum of squares for β_{13} can be calculated as follows:

$$SS_R(\beta_{13}|\beta_0, \beta_1, \beta_2, \beta_{12}, \beta_3) = \quad SS_R(\beta_0, \beta_1, \beta_2, \beta_{12}, \beta_3, \beta_{13}) - SS_R(\beta_0, \beta_1, \beta_2, \beta_{12}, \beta_3)$$

For the present case, $\theta_2 = [\beta_{13}]'$ and $\theta_1 = [\beta_0, \beta_1, \beta_2, \beta_{12}, \beta_3]'$. It can be noted that for the sequential sum of squares β_1 contains all coefficients proceeding the coefficient being tested.

The sequential sum of squares for all terms will add up to the regression sum of squares for the full model, but the sequential sum of squares are order dependent.

EXAMPLE

This example illustrates the partial F test using the sequential sum of squares. The test is conducted for the coefficient β_1 corresponding to the predictor variable x_1 for the data. The regression model used for this data set in the example is:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

The null hypothesis to test the significance of β_1 is:

$$H_0 : \beta_1 = 0$$

The statistic to test this hypothesis is:

$$F_0 = \frac{SS_R(\beta_0, \beta_1)/r}{MS_E}$$

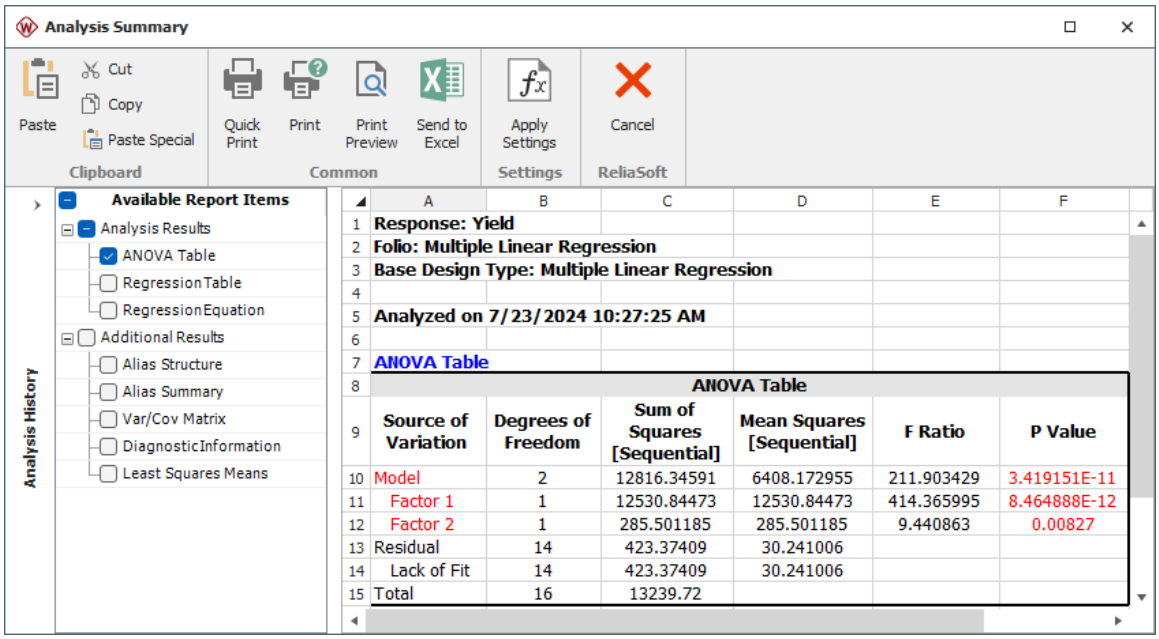
where $SS_R(\beta_0, \beta_1)$ represents the sequential sum of squares for β_1, r represents the number of degrees of freedom for $SS_R(\beta_0, \beta_1)$ (which is one because there is just one coefficient, β_1, being tested) and MS_E is the error mean square and has been calculated in the second example as 30.24.

The sequential sum of squares for β_1 is the difference between the regression sum of squares for the model after adding β_1, $Y = \beta_0 + \beta_1 x_1 + \epsilon$, and the regression sum of squares for the model before adding β_1, $Y = \beta_0 + \epsilon$. The regression sum of squares for the model $Y = \beta_0 + \beta_1 x_1 + \epsilon$ is obtained as shown next. First the design matrix for this model, X_{β_0,β_1}, is obtained by dropping the third column in the design matrix for the full model, X (the full design matrix, X, was obtained in the example). The third column of X corresponds to coefficient β_2 which is no longer used in the present model. Therefore, the design matrix for the model, $Y = \beta_0 + \beta_1 x_1 + \epsilon$, is:

$$X_{\beta_0,\beta_1} = \begin{bmatrix} 1 & 41.9 \\ 1 & 43.4 \\ . & . \\ . & . \\ 1 & 77.8 \end{bmatrix}$$

The hat matrix corresponding to this design matrix is H_{β_0,β_1}. It can be calculated using $H_{\beta_0,\beta_1} = X_{\beta_0,\beta_1}(X'_{\beta_0,\beta_1} X_{\beta_0,\beta_1})^{-1} X'_{\beta_0,\beta_1}$. Once H_{β_0,β_1} is known, the regression sum of squares for the model $Y = \beta_0 + \beta_1 x_1 + \epsilon$ can be calculated as:

$$\begin{aligned} SS_R(\beta_0, \beta_1) &= y' \left[H_{\beta_0,\beta_1} - (\frac{1}{n}) J \right] y \\ &= 12530.85 \end{aligned}$$



	A	B	C	D	E	F
1	**Response: Yield**					
2	**Folio: Multiple Linear Regression**					
3	**Base Design Type: Multiple Linear Regression**					
4						
5	**Analyzed on 7/23/2024 10:27:25 AM**					
6						
7	**ANOVA Table**					
8	**ANOVA Table**					
9	**Source of Variation**	**Degrees of Freedom**	**Sum of Squares [Sequential]**	**Mean Squares [Sequential]**	**F Ratio**	**P Value**
10	Model	2	12816.34591	6408.172955	211.903429	3.419151E-11
11	Factor 1	1	12530.84473	12530.84473	414.365995	8.464888E-12
12	Factor 2	1	285.501185	285.501185	9.440863	0.00827
13	Residual	14	423.37409	30.241006		
14	Lack of Fit	14	423.37409	30.241006		
15	Total	16	13239.72			

The regression sum of squares for the model $Y = \beta_0 + \epsilon$ is equal to zero since this model does not contain any variables. Therefore:

$$SS_R(\beta_0) = 0$$

The sequential sum of squares for β_1 is:

$$\begin{aligned} SS_R(\beta_1|\beta_0) &= SS_R(\beta_0, \beta_1) - SS_R(\beta_0) \\ &= 12530.85 - 0 \\ &= 12530.85 \end{aligned}$$

Knowing the sequential sum of squares, the statistic to test the significance of β_1 is:

$$\begin{aligned} f_0 &= \frac{SS_R(\beta_0, \beta_1)/r}{MS_E} \\ &= \frac{12530.85/1}{30.24} \\ &= 414.366 \end{aligned}$$

The p value corresponding to this statistic based on the F distribution with 1 degree of freedom in the numerator and 14 degrees of freedom in the denominator is:

$$\begin{aligned} p\ value &= 1 - P(F \leq f_0) \\ &= 1 - 0.999999 \\ &= 8.46 \times 10^{-12} \end{aligned}$$

Assuming that the desired significance is 0.1, since p value < 0.1, $H_0 : \beta_1 = 0$ is rejected and it can be concluded that β_1 is significant. The test for β_2 can be carried out in a similar manner. This result is shown in the following figure.

Confidence Intervals in Multiple Linear Regression

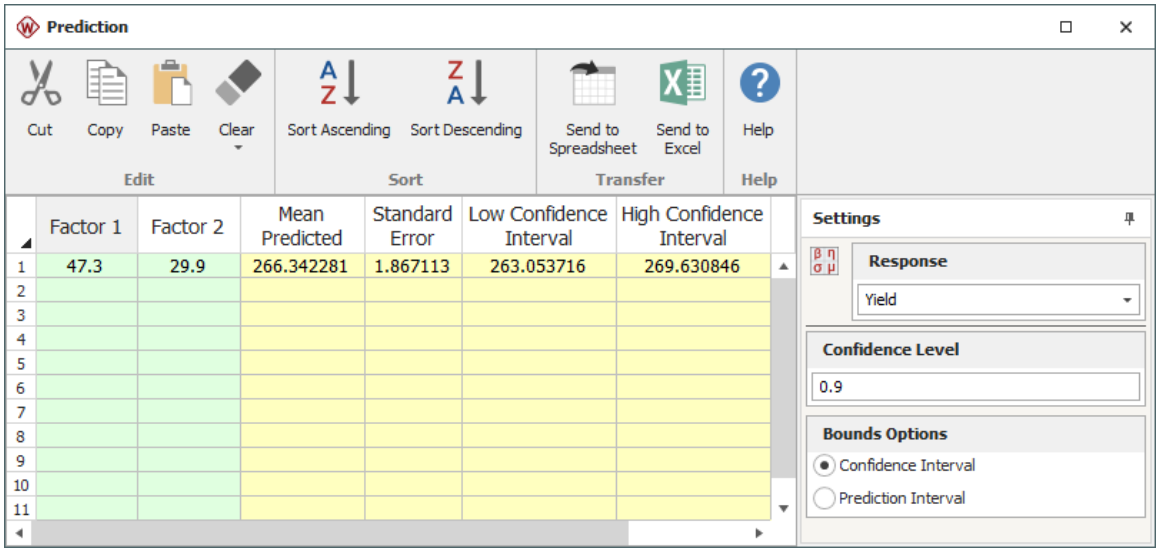
Calculation of confidence intervals for multiple linear regression models are similar to those for simple linear regression models explained in Simple Linear Regression Analysis.

Confidence Interval on Regression Coefficients

A 100 $(1-\alpha)$ percent confidence interval on the regression coefficient, β_j, is obtained as follows:

$$\hat{\beta}_j \pm t_{\alpha/2,n-(k+1)}\sqrt{C_{jj}}$$

The confidence interval on the regression coefficients are displayed in the Regression Information table under the Low Confidence and High Confidence columns as shown in the following figure.



Confidence Interval on Fitted Values, $\hat{y}_i$ A 100 $(1-\alpha)$ percent confidence interval on any fitted value, $\hat{y}_i$, is given by:

$$\hat{y}_i \pm t_{\alpha/2,n-(k+1)}\sqrt{\hat{\sigma}^2 x_i'(X'X)^{-1}x_i}$$

where:

$$x_i = \begin{bmatrix} 1 \\ x_{i1} \\ . \\ . \\ . \\ x_{ik} \end{bmatrix}$$

In the above example, the fitted value corresponding to the fifth observation was calculated as $\hat{y}_5 = 266.3$. The 90% confidence interval on this value can be obtained as shown in the figure below. The values of 47.3 and 29.9 used in the figure are the values of the predictor variables corresponding to the fifth observation the table.

Confidence Interval on New Observations

As explained in Simple Linear Regression Analysis, the confidence interval on a new observation is also referred to as the prediction interval. The prediction interval takes into account both the error from the fitted model and the error associated with future observations. A 100 ($1-\alpha$) percent confidence interval on a new observation, $\hat{y}_p$, is obtained as follows:

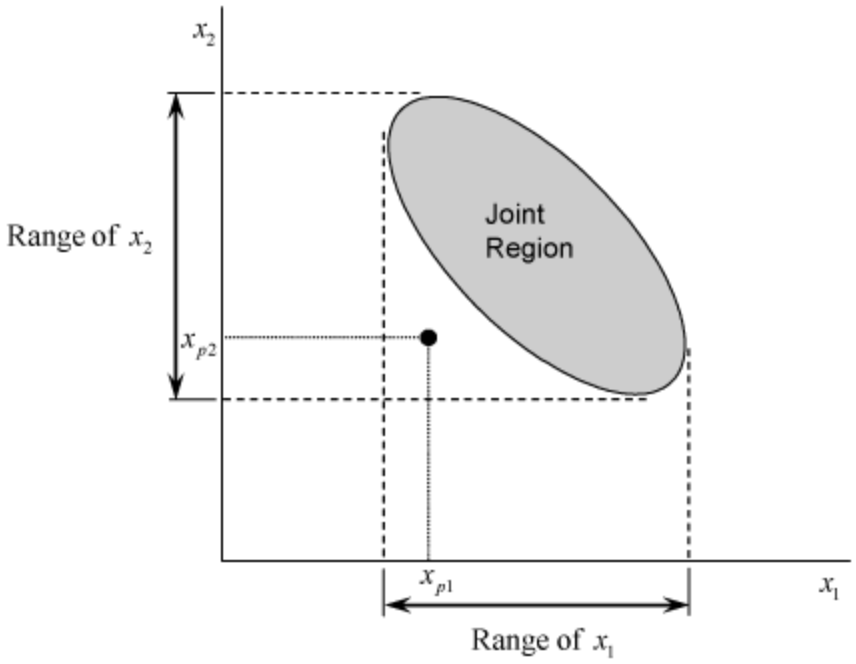
$$\hat{y}_p \pm t_{\alpha/2,n-(k+1)} \sqrt{\hat{\sigma}^2 (1 + x_p' (X'X)^{-1} x_p)}$$

where:

$$x_p = \begin{bmatrix} 1 \\ x_{p1} \\ . \\ . \\ . \\ x_{pk} \end{bmatrix}$$

x_{p1} ,..., x_{pk} are the levels of the predictor variables at which the new observation, $\hat{y}_p$, needs to be obtained.

In multiple linear regression, prediction intervals should only be obtained at the levels of the predictor variables where the regression model applies. In the case of multiple linear regression it is easy to miss this. Having values lying within the range of the predictor variables does not necessarily mean that the new observation lies in the region to which the model is applicable. For example, consider the next figure where the shaded area shows the region to which a two variable regression model is applicable. The point corresponding to p th level of first predictor variable, x_1 , and p th level of the second predictor variable, x_2 , does not lie in the shaded area, although both of these levels are within the range of the first and second predictor variables respectively. In this case, the regression model is not applicable at this point.



Measures of Model Adequacy

As in the case of simple linear regression, analysis of a fitted multiple linear regression model is important before inferences based on the model are undertaken. This section presents some techniques that can be used to check the appropriateness of the multiple linear regression model.

Coefficient of Multiple Determination, *R*^2

The coefficient of multiple determination is similar to the coefficient of determination used in the case of simple linear regression. It is defined as:

$$R^2 = \frac{SS_R}{SS_T}$$
$$= 1 - \frac{SS_E}{SS_T}$$

R^2 indicates the amount of total variability explained by the regression model. The positive square root of R^2 is called the multiple correlation coefficient and measures the linear association between Y and the predictor variables, x_1 , x_2 ... x_k .

The value of R^2 increases as more terms are added to the model, even if the new term does not contribute significantly to the model. An increase in the value of R^2 cannot be taken as a sign to conclude that the new model is superior to the older model. A better statistic to use is the adjusted R^2 statistic defined as follows:

$$\begin{aligned} R^2_{adj} &= 1 - \frac{MS_E}{MS_T} \\ &= 1 - \frac{SS_E/(n-(k+1))}{SS_T/(n-1)} \\ &= 1 - (\frac{n-1}{n-(k+1)})(1-R^2) \end{aligned}$$

The adjusted R^2 only increases when significant terms are added to the model. Addition of unimportant terms may lead to a decrease in the value of R^2_{adj}.

In a DOE folio, R^2 and R^2_{adj} values are displayed as R-sq and R-sq(adj), respectively. Other values displayed along with these values are S, PRESS and R-sq(pred). As explained in Simple Linear Regression Analysis, the value of S is the square root of the error mean square, MS_E, and represents the "standard error of the model."

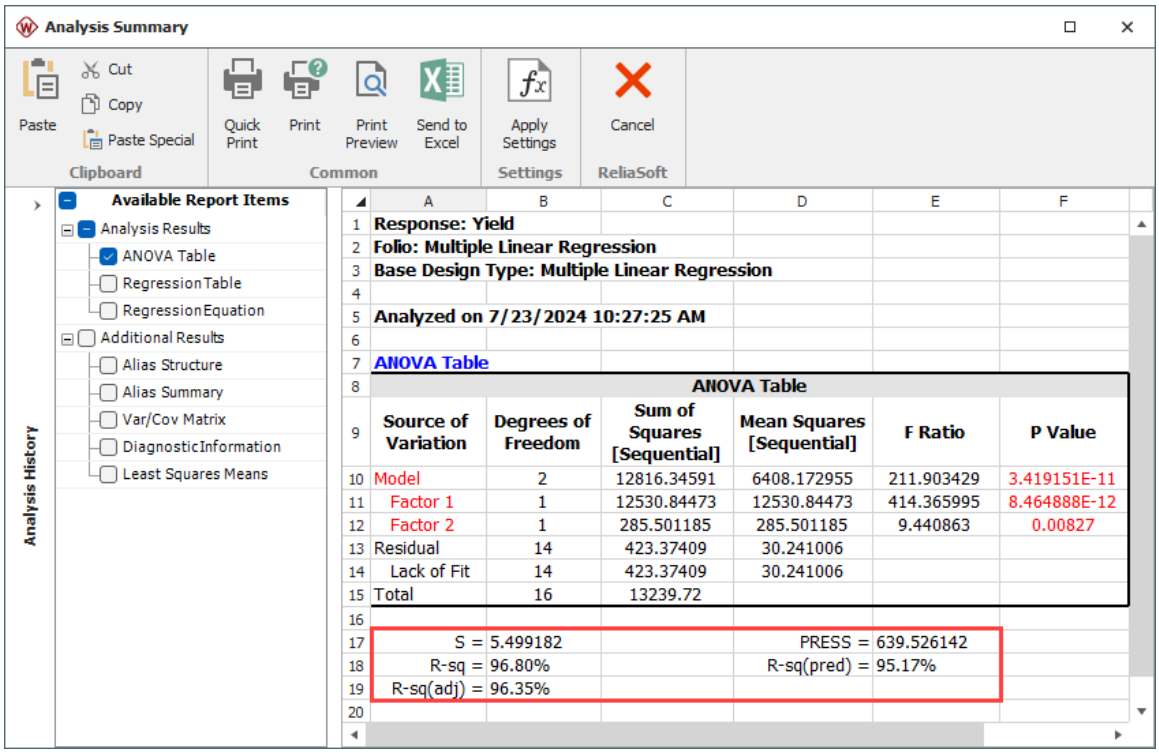
PRESS is an abbreviation for prediction error sum of squares. It is the error sum of squares calculated using the PRESS residuals in place of the residuals, e_i, in the equation for the error sum of squares. The PRESS residual, $e_{(i)}$, for a particular observation, y_i, is obtained by fitting the regression model to the remaining observations. Then the value for a new observation, $\hat{y}_p$, corresponding to the observation in question, y_i, is obtained based on the new regression model. The difference between y_i and $\hat{y}_p$ gives $e_{(i)}$. The PRESS residual, $e_{(i)}$, can also be obtained using h_{ii}, the diagonal element of the hat matrix, H, as follows:

$$e_{(i)} = \frac{e_i}{1-h_{ii}}$$

R-sq(pred), also referred to as prediction R^2, is obtained using PRESS as shown next:

$$R^2_{pred} = 1 - \frac{PRESS}{SS_T}$$

The values of R-sq, R-sq(adj) and S are indicators of how well the regression model fits the observed data. The values of PRESS and R-sq(pred) are indicators of how well the regression model predicts new observations. For example, higher values of PRESS or lower values of R-sq (pred) indicate a model that predicts poorly. The figure below shows these values for the data. The values indicate that the regression model fits the data well and also predicts well.



	A	B	C	D	E	F
1	**Response: Yield**					
2	**Folio: Multiple Linear Regression**					
3	**Base Design Type: Multiple Linear Regression**					
4						
5	**Analyzed on 7/23/2024 10:27:25 AM**					
6						
7	**ANOVA Table**					
8			**ANOVA Table**			
9	**Source of Variation**	**Degrees of Freedom**	**Sum of Squares [Sequential]**	**Mean Squares [Sequential]**	**F Ratio**	**P Value**
10	Model	2	12816.34591	6408.172955	211.903429	3.419151E-11
11	Factor 1	1	12530.84473	12530.84473	414.365995	8.464888E-12
12	Factor 2	1	285.501185	285.501185	9.440863	0.00827
13	Residual	14	423.37409	30.241006		
14	Lack of Fit	14	423.37409	30.241006		
15	Total	16	13239.72			
16						
17	S =	5.499182		PRESS =	639.526142	
18	R-sq =	96.80%		R-sq(pred) =	95.17%	
19	R-sq(adj) =	96.35%				
20						

Residual Analysis

Plots of residuals, e_i , similar to the ones discussed in Simple Linear Regression Analysis for simple linear regression, are used to check the adequacy of a fitted multiple linear regression model. The residuals are expected to be normally distributed with a mean of zero and a constant variance of σ^2 . In addition, they should not show any patterns or trends when plotted against any variable or in a time or run-order sequence. Residual plots may also be obtained using standardized and studentized residuals. Standardized residuals, d_i , are obtained using the following equation:

$$d_i = \frac{e_i}{\sqrt{\hat{\sigma}^2}}$$

$$= \frac{e_i}{\sqrt{MS_E}}$$

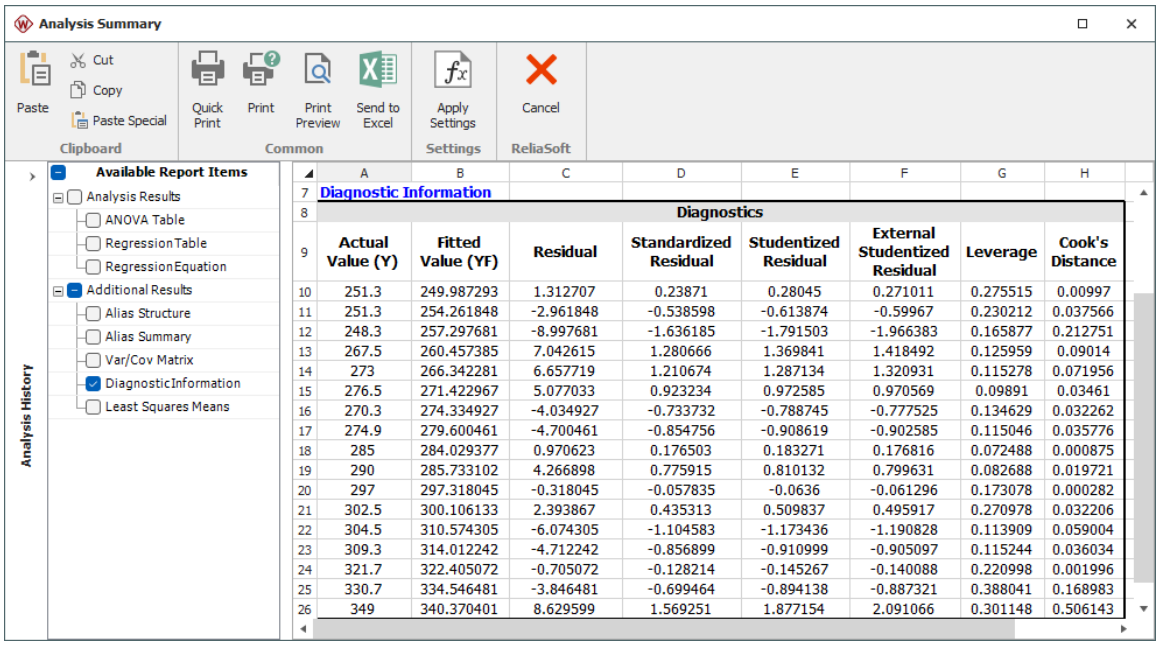
Standardized residuals are scaled so that the standard deviation of the residuals is approximately equal to one. This helps to identify possible outliers or unusual observations. However, standardized residuals may understate the true residual magnitude, hence studentized residuals, r_i , are used in their place. Studentized residuals are calculated as follows:

$$r_i = \frac{e_i}{\sqrt{\hat{\sigma}^2(1-h_{ii})}} = \frac{e_i}{\sqrt{MS_E(1-h_{ii})}}$$

where h_{ii} is the i th diagonal element of the hat matrix, H. External studentized (or the studentized deleted) residuals may also be used. These residuals are based on the PRESS residuals mentioned in Coefficient of Multiple Determination, *R*2. The reason for using the external studentized residuals is that if the i th observation is an outlier, it may influence the fitted model. In this case, the residual e_i will be small and may not disclose that i th observation is an outlier. The external studentized residual for the i th observation, t_i, is obtained as follows:

$$t_i = e_i \left[\frac{n-k}{SS_E(1-h_{ii}) - e_i^2}\right]^{0.5}$$

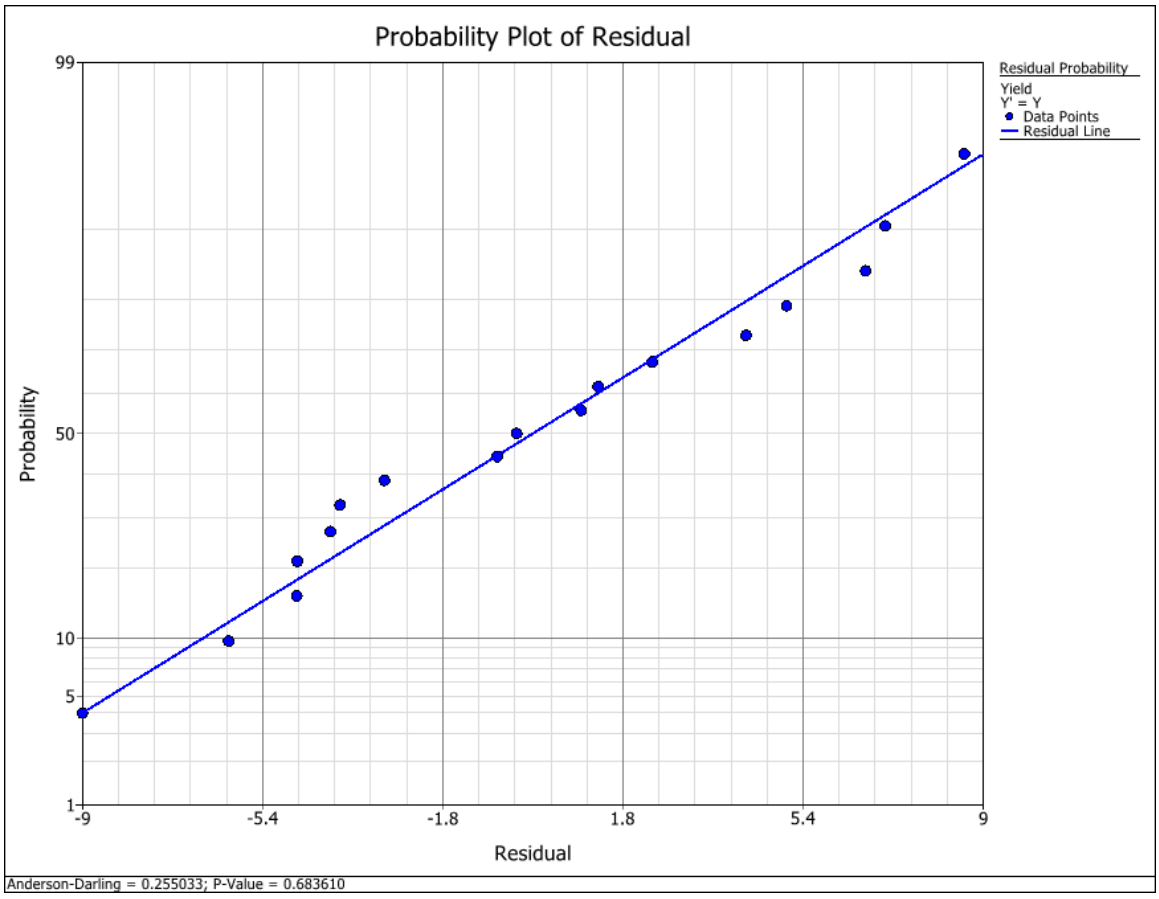
Residual values for the data are shown in the figure below. Standardized residual plots for the data are shown in next two figures. The Weibull++ DOE folio compares the residual values to the critical values on the t distribution for studentized and external studentized residuals.



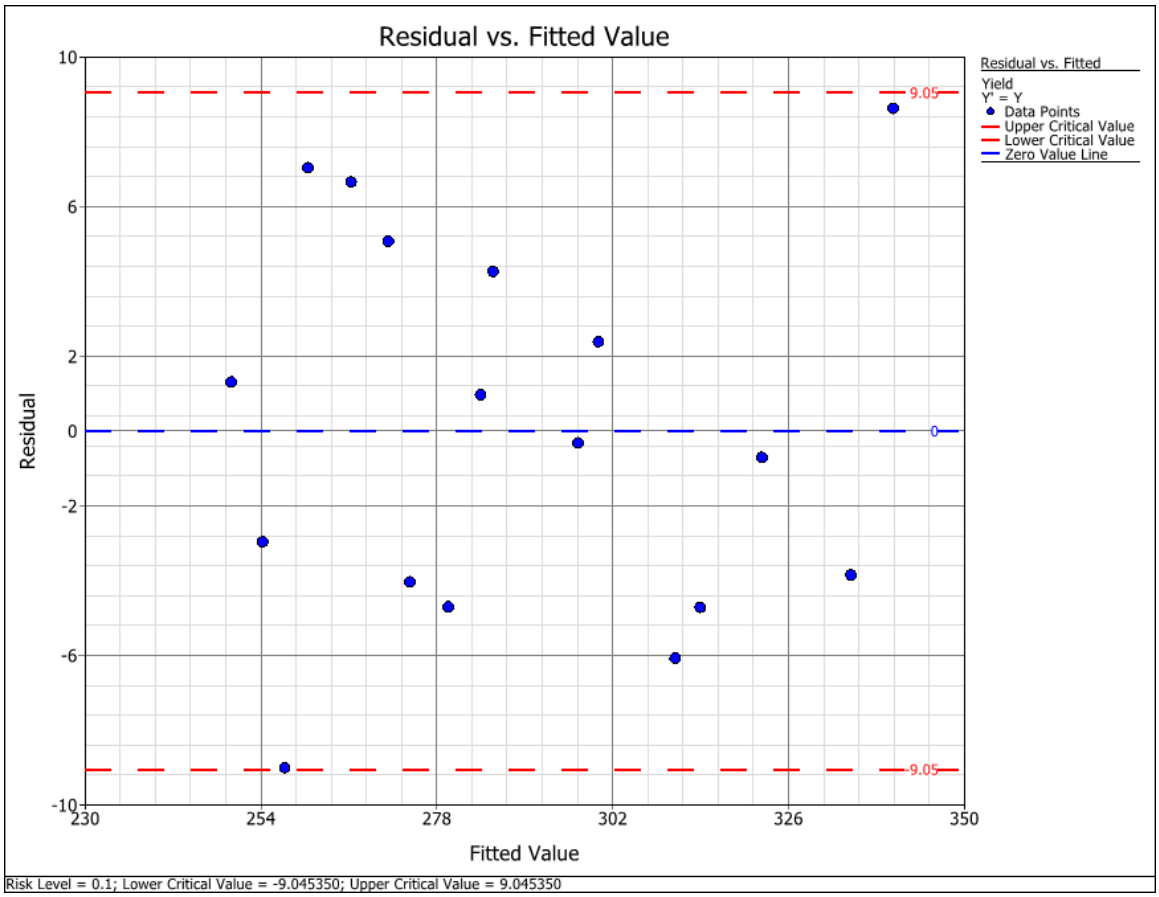
Diagnostic Information

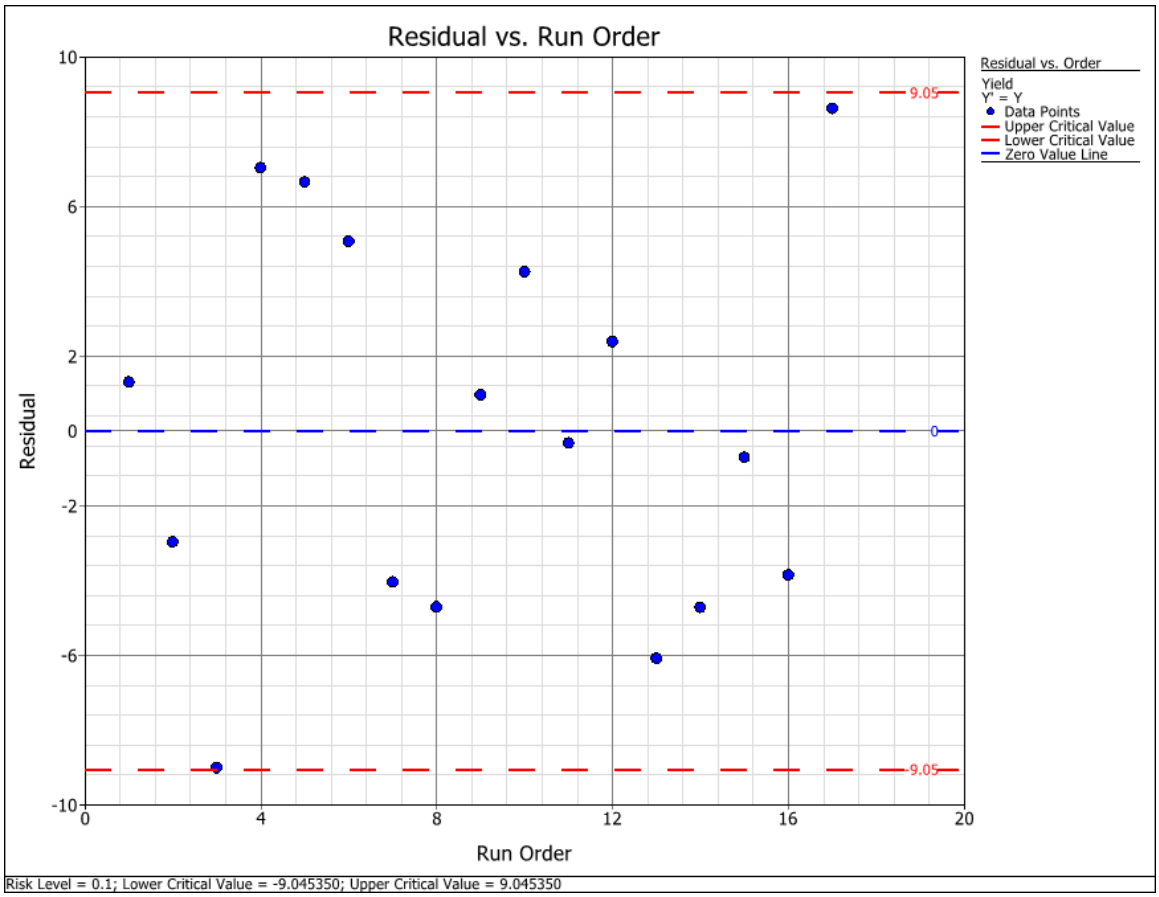
Diagnostics

Actual Value (Y)	Fitted Value (YF)	Residual	Standardized Residual	Studentized Residual	External Studentized Residual	Leverage	Cook's Distance
251.3	249.987293	1.312707	0.23871	0.28045	0.271011	0.275515	0.00997
251.3	254.261848	-2.961848	-0.538598	-0.613874	-0.59967	0.230212	0.037566
248.3	257.297681	-8.997681	-1.636185	-1.791503	-1.966383	0.165877	0.212751
267.5	260.457385	7.042615	1.280666	1.369841	1.418492	0.125959	0.09014
273	266.342281	6.657719	1.210674	1.287134	1.320931	0.115278	0.071956
276.5	271.422967	5.077033	0.923234	0.972585	0.970569	0.09891	0.03461
270.3	274.334927	-4.034927	-0.733732	-0.788745	-0.777525	0.134629	0.032262
274.9	279.600461	-4.700461	-0.854756	-0.908619	-0.902585	0.115046	0.035776
285	284.029377	0.970623	0.176503	0.183271	0.176816	0.072488	0.000875
290	285.733102	4.266898	0.775915	0.810132	0.799631	0.082688	0.019721
297	297.318045	-0.318045	-0.057835	-0.0636	-0.061296	0.173078	0.000282
302.5	300.106133	2.393867	0.435313	0.509837	0.495917	0.270978	0.032206
304.5	310.574305	-6.074305	-1.104583	-1.173436	-1.190828	0.113909	0.059004
309.3	314.012242	-4.712242	-0.856899	-0.910999	-0.905097	0.115244	0.036034
321.7	322.405072	-0.705072	-0.128214	-0.145267	-0.140088	0.220998	0.001996
330.7	334.546481	-3.846481	-0.699464	-0.894138	-0.887321	0.388041	0.168983
349	340.370401	8.629599	1.569251	1.877154	2.091066	0.301148	0.506143



For other residuals the normal distribution is used. For example, for the data, the critical values on the t distribution at a significance of 0.1 are $t_{0.05,14} = 1.761$ and $-t_{0.05,14} = -1.761$ (as calculated in the example, Test on Individual Regression Coefficients (t Test)). The studentized residual values corresponding to the 3rd and 17th observations lie outside the critical values. Therefore, the 3rd and 17th observations are outliers. This can also be seen on the residual plots in the next two figures.


Residual vs. Fitted Value
Residual vs. Fitted
Yield
Y' = Y
Data Points
Upper Critical Value
Lower Critical Value
Zero Value Line
9.05
0
-9.05
Residual
Fitted Value
10
6
2
0
-2
-6
-10
230
254
278
302
326
350
Risk Level = 0.1; Lower Critical Value = -9.045350; Upper Critical Value = 9.045350



Outlying *x* Observations

Residuals help to identify outlying y observations. Outlying x observations can be detected using leverage. Leverage values are the diagonal elements of the hat matrix, h_{ii} . The h_{ii} values always lie between 0 and 1. Values of h_{ii} greater than $2(k+1)/n$ are considered to be indicators of outlying x observations.

Influential Observations Detection

Once an outlier is identified, it is important to determine if the outlier has a significant effect on the regression model. One measure to detect influential observations is Cook's distance measure which is computed as follows:

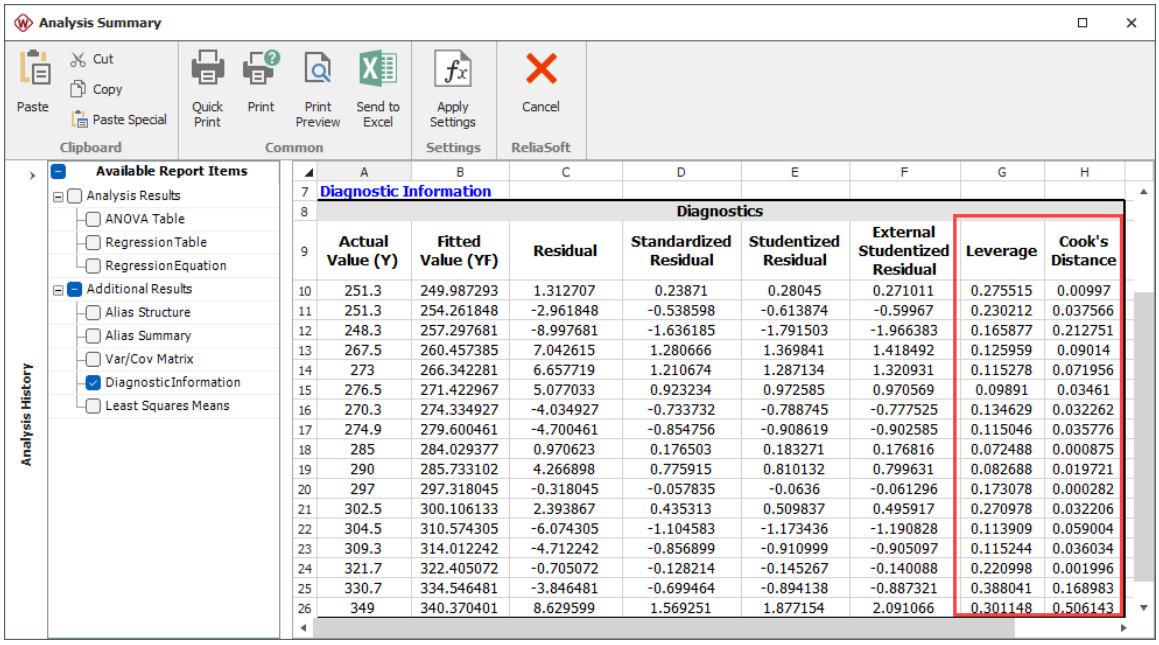
$$D_i = \frac{r_i^2}{(k+1)} \left[\frac{h_{ii}}{(1-h_{ii})} \right]$$

To use Cook's distance measure, the D_i values are compared to percentile values on the F distribution with $(k+1, n-(k+1))$ degrees of freedom. If the percentile value is less than 10 or 20 percent, then the i th case has little influence on the fitted values. However, if the percentile

value is close to 50 percent or greater, the i th case is influential, and fitted values with and without the i th case will differ substantially.

Example

Cook's distance measure can be calculated as shown next. The distance measure is calculated for the first observation of the data. The remaining values along with the leverage values are shown in the figure below (displaying Leverage and Cook's distance measure for the data).



Diagnostic Information

Diagnostics

Actual Value (Y)	Fitted Value (YF)	Residual	Standardized Residual	Studentized Residual	External Studentized Residual	Leverage	Cook's Distance
251.3	249.987293	1.312707	0.23871	0.28045	0.271011	0.275515	0.00997
251.3	254.261848	-2.961848	-0.538598	-0.613874	-0.59967	0.230212	0.037566
248.3	257.297681	-8.997681	-1.636185	-1.791503	-1.966383	0.165877	0.212751
267.5	260.457385	7.042615	1.280666	1.369841	1.418492	0.125959	0.09014
273	266.342281	6.657719	1.210674	1.287134	1.320931	0.115278	0.071956
276.5	271.422967	5.077033	0.923234	0.972585	0.970569	0.09891	0.03461
270.3	274.334927	-4.034927	-0.733732	-0.788745	-0.777525	0.134629	0.032262
274.9	279.600461	-4.700461	-0.854756	-0.908619	-0.902585	0.115046	0.035776
285	284.029377	0.970623	0.176503	0.183271	0.176816	0.072488	0.000875
290	285.733102	4.266898	0.775915	0.810132	0.799631	0.082688	0.019721
297	297.318045	-0.318045	-0.057835	-0.0636	-0.061296	0.173078	0.000282
302.5	300.106133	2.393867	0.435313	0.509837	0.495917	0.270978	0.032206
304.5	310.574305	-6.074305	-1.104583	-1.173436	-1.190828	0.113909	0.059004
309.3	314.012242	-4.712242	-0.856899	-0.910999	-0.905097	0.115244	0.036034
321.7	322.405072	-0.705072	-0.128214	-0.145267	-0.140088	0.220998	0.001996
330.7	334.546481	-3.846481	-0.699464	-0.894138	-0.887321	0.388041	0.168983
349	340.370401	8.629599	1.569251	1.877154	2.091066	0.301148	0.506143

The standardized residual corresponding to the first observation is:

$$\begin{aligned} r_1 &= \frac{e_1}{\sqrt{MS_E(1-h_{11})}} \\ &= \frac{1.3127}{\sqrt{30.3(1-0.2755)}} \\ &= 0.2804 \end{aligned}$$

Cook's distance measure for the first observation can now be calculated as:

$$\begin{aligned} D_1 &= \frac{r_1^2}{(k+1)}\left[\frac{h_{11}}{(1-h_{11})}\right] \\ &= \frac{0.2804^2}{(2+1)}\left[\frac{0.2755}{(1-0.2755)}\right] \\ &= 0.01 \end{aligned}$$

The 50th percentile value for $F_{3,14}$ is 0.83. Since all D_i values are less than this value there are no influential observations.

Lack-of-Fit Test

The lack-of-fit test for simple linear regression discussed in Simple Linear Regression Analysis may also be applied to multiple linear regression to check the appropriateness of the fitted response surface and see if a higher order model is required. Data for m replicates may be collected as follows for all n levels of the predictor variables:

$$\begin{aligned} y_{11}, y_{12}, \ldots, y_{1m} &\ m \text{ repeated observations at the first level} \\ y_{21}, y_{22}, \ldots, y_{2m} &\ m \text{ repeated observations at the second level} \\ &\ldots \\ y_{i1}, y_{i2}, \ldots, y_{im} &\ m \text{ repeated observations at the } i\text{th level} \\ &\ldots \\ y_{n1}, y_{n2}, \ldots, y_{nm} &\ m \text{ repeated observations at the } n\text{th level} \end{aligned}$$

The sum of squares due to pure error, SS_{PE}, can be obtained as discussed in the Simple Linear Regression Analysis as:

$$SS_{PE} = \sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij} - \bar{y}_i)^2$$

The number of degrees of freedom associated with SS_{PE} are:

$$dof(SS_{PE}) = nm - n$$

Knowing SS_{PE}, sum of squares due to lack-of-fit, SS_{LOF}, can be obtained as:

$$SS_{LOF} = SS_E - SS_{PE}$$

The number of degrees of freedom associated with SS_{LOF} are:

$$\begin{aligned} dof(SS_{LOF}) = &\quad dof(SS_E) - dof(SS_{PE}) \\ = &\quad n - (k+1) - (nm - n) \end{aligned}$$

The test statistic for the lack-of-fit test is:

$$\begin{aligned} F_0 = &\quad \frac{SS_{LOF}/dof(SS_{LOF})}{SS_{PE}/dof(SS_{PE})} \\ = &\quad \frac{MS_{LOF}}{MS_{PE}} \end{aligned}$$

Other Topics in Multiple Linear Regression

Polynomial Regression Models

Polynomial regression models are used when the response is curvilinear. The equation shown next presents a second order polynomial regression model with one predictor variable:

$$Y = \beta_0 + \beta_1 x_1 + \beta_{11} x_1^2 + \epsilon$$

Usually, coded values are used in these models. Values of the variables are coded by centering or expressing the levels of the variable as deviations from the mean value of the variable and then scaling or dividing the deviations obtained by half of the range of the variable.

$$coded\ value = \frac{actual\ value - mean}{half\ of\ range}$$

The reason for using coded predictor variables is that many times x and x^2 are highly correlated and, if uncoded values are used, there may be computational difficulties while calculating the $(X'X)^{-1}$ matrix to obtain the estimates, $\hat{\beta}$, of the regression coefficients using the equation for the F distribution given in Statistics Background on DOE.

Qualitative Factors

The multiple linear regression model also supports the use of qualitative factors. For example, gender may need to be included as a factor in a regression model. One of the ways to include qualitative factors in a regression model is to employ indicator variables. Indicator variables take on values of 0 or 1. For example, an indicator variable may be used with a value of 1 to indicate female and a value of 0 to indicate male.

$$x_1 = \begin{cases} 1 \text{ Female} \\ 0 \text{ Male} \end{cases}$$

In general ($n-1$) indicator variables are required to represent a qualitative factor with n levels. As an example, a qualitative factor representing three types of machines may be represented as follows using two indicator variables:

$$\begin{aligned} x_1 &= 1,\ x_2 = 0 \text{ Machine Type I} \\ x_1 &= 0,\ x_2 = 1 \text{ Machine Type II} \\ x_1 &= 0,\ x_2 = 0 \text{ Machine Type III} \end{aligned}$$

An alternative coding scheme for this example is to use a value of -1 for all indicator variables when representing the last level of the factor:

$$\begin{aligned} x_1 = \quad & 1, \; x_2 = \quad 0 \text{ Machine Type I} \\ x_1 = \quad & 0, \; x_2 = \quad 1 \text{ Machine Type II} \\ x_1 = \quad & -1, \; x_2 = \quad -1 \text{ Machine Type III} \end{aligned}$$

Indicator variables are also referred to as dummy variables or binary variables.

Example

Consider data from two types of reactors of a chemical process shown where the yield values are recorded for various levels of factor x_1 . Assuming there are no interactions between the reactor type and x_1 , a regression model can be fitted to this data as shown next.

Since the reactor type is a qualitative factor with two levels, it can be represented by using one indicator variable. Let x_2 be the indicator variable representing the reactor type, with 0 representing the first type of reactor and 1 representing the second type of reactor.

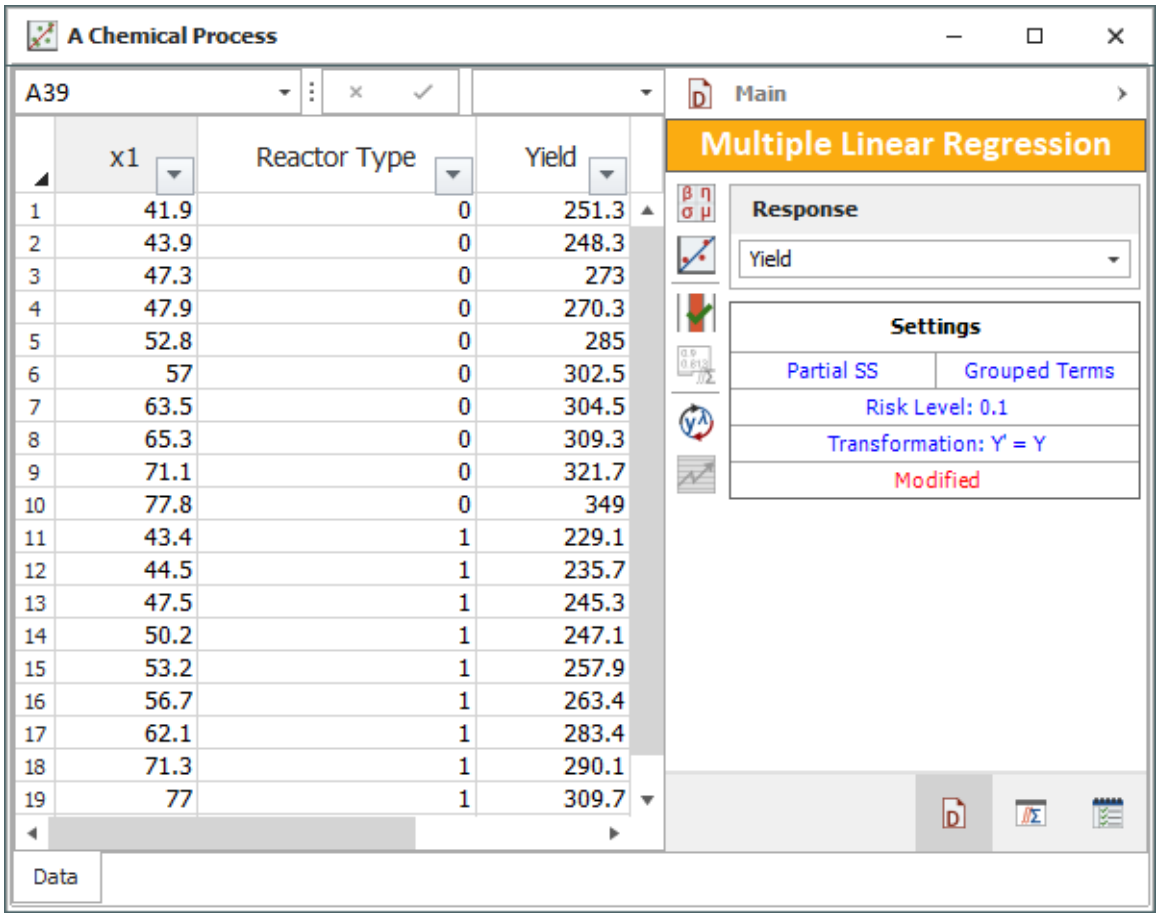
$$x_2 = \begin{cases} 0 \text{ Reactor Type I} \\ 1 \text{ Reactor Type II} \end{cases}$$

Observation Number	Factor (x_{i1})	Reactor Type	Yield (y_i)
1	41.9	I	251.3
2	43.9	I	248.3
3	47.3	I	273.0
4	47.9	I	270.3
5	52.8	I	285.0
6	57.0	I	302.5
7	63.5	I	304.5
8	65.3	I	309.3
9	71.1	I	321.7
10	77.8	I	349.0
11	43.4	II	229.1
12	44.5	II	235.7
13	47.5	II	245.3
14	50.2	II	247.1
15	53.2	II	257.9
16	56.7	II	263.4
17	62.1	II	283.4
18	71.3	II	290.1
19	77.0	II	309.7

Data entry in the DOE folio for this example is shown in the figure after the table below. The regression model for this data is:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

The X and y matrices for the given data are:



	x1	Reactor Type	Yield
1	41.9	0	251.3
2	43.9	0	248.3
3	47.3	0	273
4	47.9	0	270.3
5	52.8	0	285
6	57	0	302.5
7	63.5	0	304.5
8	65.3	0	309.3
9	71.1	0	321.7
10	77.8	0	349
11	43.4	1	229.1
12	44.5	1	235.7
13	47.5	1	245.3
14	50.2	1	247.1
15	53.2	1	257.9
16	56.7	1	263.4
17	62.1	1	283.4
18	71.3	1	290.1
19	77	1	309.7

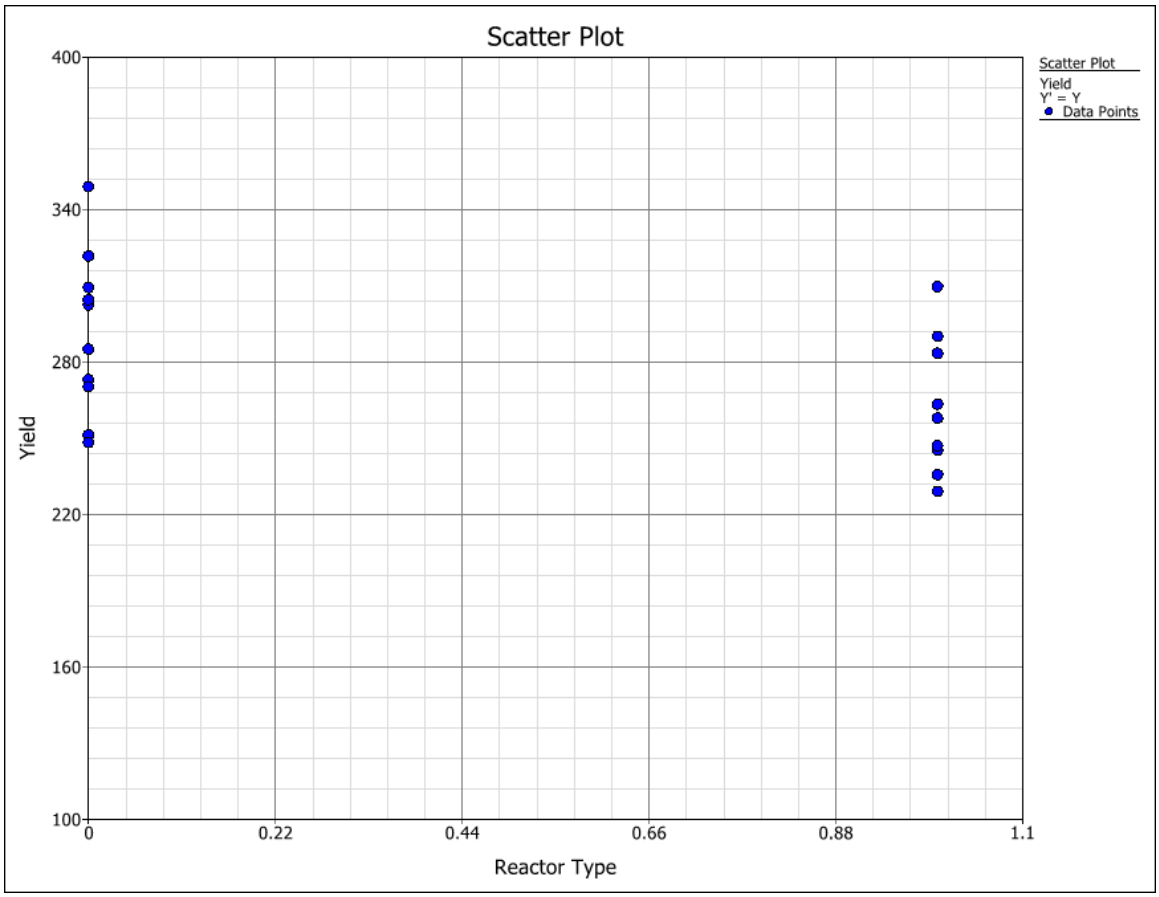
The estimated regression coefficients for the model can be obtained as:

$$\hat{\beta} = (X'X)^{-1}X'y = \begin{bmatrix} 153.7 \\ 2.4 \\ -27.5 \end{bmatrix}$$

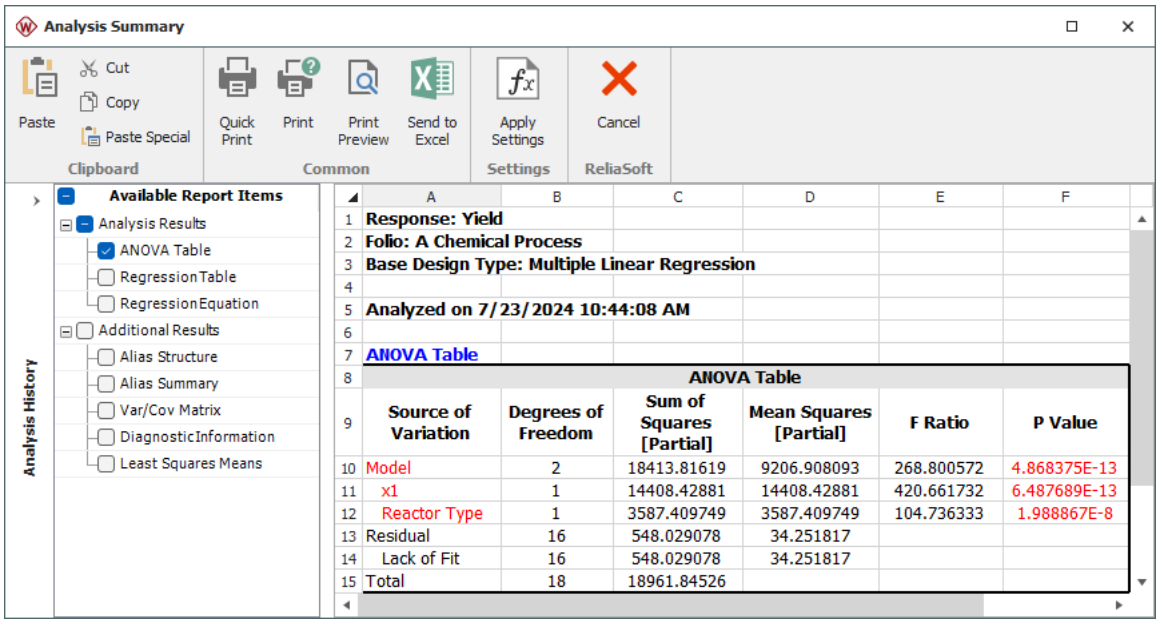
Therefore, the fitted regression model is:

$$\hat{y} = 153.7 + 2.4x_1 - 27.5x_2$$

Note that since x_2 represents a qualitative predictor variable, the fitted regression model cannot be plotted simultaneously against x_1 and x_2 in a two-dimensional space (because the resulting surface plot will be meaningless for the dimension in x_2). To illustrate this, a scatter plot of the data against x_2 is shown in the following figure.



It can be noted that, in the case of qualitative factors, the nature of the relationship between the response (yield) and the qualitative factor (reactor type) cannot be categorized as linear, or quadratic, or cubic, etc. The only conclusion that can be arrived at for these factors is to see if these factors contribute significantly to the regression model. This can be done by employing the partial F test discussed in Multiple Linear Regression Analysis (using the extra sum of squares of the indicator variables representing these factors). The results of the test for the present example are shown in the ANOVA table. The results show that x_2 (reactor type) contributes significantly to the fitted regression model.



Response: Yield
Folio: A Chemical Process
Base Design Type: Multiple Linear Regression

Analyzed on 7/23/2024 10:44:08 AM

ANOVA Table

ANOVA Table

Source of Variation	Degrees of Freedom	Sum of Squares [Partial]	Mean Squares [Partial]	F Ratio	P Value
Model	2	18413.81619	9206.908093	268.800572	4.868375E-13
x1	1	14408.42881	14408.42881	420.661732	6.487689E-13
Reactor Type	1	3587.409749	3587.409749	104.736333	1.988867E-8
Residual	16	548.029078	34.251817		
Lack of Fit	16	548.029078	34.251817		
Total	18	18961.84526			

Multicollinearity

At times the predictor variables included in a multiple linear regression model may be found to be dependent on each other. Multicollinearity is said to exist in a multiple regression model with strong dependencies between the predictor variables. Multicollinearity affects the regression coefficients and the extra sum of squares of the predictor variables. In a model with multicollinearity the estimate of the regression coefficient of a predictor variable depends on what other predictor variables are included the model. The dependence may even lead to change in the sign of the regression coefficient. In a such models, an estimated regression coefficient may not be found to be significant individually (when using the t test on the individual coefficient or looking at the p value) even though a statistical relation is found to exist between the response variable and the set of the predictor variables (when using the F test for the set of predictor variables). Therefore, you should be careful while looking at individual predictor variables in models that have multicollinearity. Care should also be taken while looking at the extra sum of squares for a predictor variable that is correlated with other variables. This is because in models with multicollinearity the extra sum of squares is not unique and depends on the other predictor variables included in the model.

Multicollinearity can be detected using the variance inflation factor (abbreviated VIF). VIF for a coefficient β_j is defined as:

$$VIF = \frac{1}{(1 - R_j^2)}$$

where R_j^2 is the coefficient of multiple determination resulting from regressing the j th predictor variable, x_j , on the remaining k -1 predictor variables. Mean values of VIF considerably greater than 1 indicate multicollinearity problems. A few methods of dealing with multicollinearity include increasing the number of observations in a way designed to break up dependencies among predictor variables, combining the linearly dependent predictor variables into one variable, eliminating variables from the model that are unimportant or using coded variables.

Example

Variance inflation factors can be obtained for the data below.

Observation Number	Factor 1 (x_{i1})	Factor 2 (x_{i2})	Yield (y_i)
1	41.9	29.1	251.3
2	43.4	29.3	251.3
3	43.9	29.5	248.3
4	44.5	29.7	267.5
5	47.3	29.9	273.0
6	47.5	30.3	276.5
7	47.9	30.5	270.3
8	50.2	30.7	274.9
9	52.8	30.8	285.0
10	53.2	30.9	290.0
11	56.7	31.5	297.0
12	57.0	31.7	302.5
13	63.5	31.9	304.5
14	65.3	32.0	309.3
15	71.1	32.1	321.7
16	77.0	32.5	330.7
17	77.8	32.9	349.0

To calculate the variance inflation factor for x_1 , R_1^2 has to be calculated.R_1^2 is the coefficient of determination for the model when x_1 is regressed on the remaining variables. In the case of this example there is just one remaining variable which is x_2 . If a regression model is fit to the data, taking x_1 as the response variable and x_2 as the predictor variable, then the design matrix and the vector of observations are:

$$X_{R_1} = \begin{bmatrix} 1 & 29.1 \\ 1 & 29.3 \\ . & . \\ . & . \\ . & . \\ 1 & 32.9 \end{bmatrix} \quad y_{R_1} = \begin{bmatrix} 41.9 \\ 43.4 \\ . \\ . \\ . \\ 77.8 \end{bmatrix}$$

The regression sum of squares for this model can be obtained as:

$$\begin{aligned} SS_R &= y'_{R_1} \left[H_{R_1} - (\frac{1}{n})J \right] y_{R_1} \\ &= 1988.6 \end{aligned}$$

where H_{R_1} is the hat matrix (and is calculated using $H_{R_1} = X_{R_1}(X'_{R_1}X_{R_1})^{-1}X'_{R_1}$) and J is the matrix of ones. The total sum of squares for the model can be calculated as:

$$\begin{aligned} SS_T &= y' \left[I - (\frac{1}{n})J \right] y \\ &= 2182.9 \end{aligned}$$

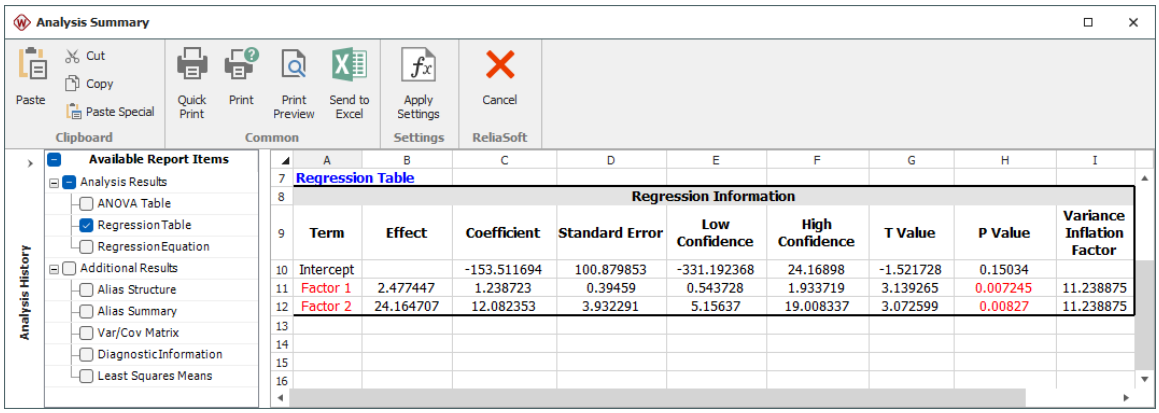
where I is the identity matrix. Therefore:

$$\begin{aligned} R_1^2 &= \frac{SS_R}{SS_T} \\ &= \frac{1988.6}{2182.9} \\ &= 0.911 \end{aligned}$$

Then the variance inflation factor for x_1 is:

$$\begin{aligned} VIF_1 &= \frac{1}{(1 - R_1^2)} \\ &= \frac{1}{1 - 0.911} \\ &= 11.2 \end{aligned}$$

The variance inflation factor for x_2 , VIF_2 , can be obtained in a similar manner. In the DOE folios, the variance inflation factors are displayed in the VIF column of the Regression Information table as shown in the following figure. Since the values of the variance inflation factors obtained are considerably greater than 1, multicollinearity is an issue for the data.



Regression Table

Regression Information								
Term	Effect	Coefficient	Standard Error	Low Confidence	High Confidence	T Value	P Value	Variance Inflation Factor
Intercept		-153.511694	100.879853	-331.192368	24.16898	-1.521728	0.15034	
Factor 1	2.477447	1.238723	0.39459	0.543728	1.933719	3.139265	0.007245	11.238875
Factor 2	24.164707	12.082353	3.932291	5.15637	19.008337	3.072599	0.00827	11.238875

One Factor Designs

IN THIS CHAPTER

Fitting ANOVA Models **112**

Treat Numerical Factors as Qualitative or Quantitative? **114**
Expression of the ANOVA Model as Y = XB + ε 114

Hypothesis Test in Single Factor Experiments **116**
Calculation of the Statistic F_0 116

Confidence Interval on the i^th Treatment Mean **119**

Confidence Interval on the Difference in Two Treatment Means **120**

Residual Analysis **123**

Box-Cox Method **124**
Example 126

As explained in Simple Linear Regression Analysis and Multiple Linear Regression Analysis, the analysis of observational studies involves the use of regression models. The analysis of experimental studies involves the use of analysis of variance (ANOVA) models. For a comparison of the two models see Fitting ANOVA Models. In single factor experiments, ANOVA models are used to compare the mean response values at different levels of the factor. Each level of the factor is investigated to see if the response is significantly different from the response at other levels of the factor. The analysis of single factor experiments is often referred to as *one-way ANOVA*.

To illustrate the use of ANOVA models in the analysis of experiments, consider a single factor experiment where the analyst wants to see if the surface finish of certain parts is affected by the speed of a lathe machine. Data is collected for three speeds (or three treatments). Each treatment is replicated four times. Therefore, this experiment design is balanced. Surface finish values recorded using randomization are shown in the following table.

Lathe Speed (levels)	Observations (replicates)			
	1	2	3	4
500	6	13	7	8
700	13	16	10	14
900	23	20	16	18

The ANOVA model for this experiment can be stated as follows:

$$Y_{ij} = \mu_i + \epsilon_{ij}$$

The ANOVA model assumes that the response at each factor level, i, is the sum of the mean response at the ith level, μ_i, and a random error term, ϵ_{ij}. The subscript i denotes the factor level while the subscript j denotes the replicate. If there are n_a levels of the factor and m replicates at each level then $i = 1, 2, \ldots, n_a$ and $j = 1, 2, \ldots, m$. The random error terms, ϵ_{ij}, are assumed to be normally and independently distributed with a mean of zero and variance of σ^2. Therefore, the response at each level can be thought of as a normally distributed population with a mean of μ_i and constant variance of σ^2. The equation given above is referred to as the *means model*.

The ANOVA model of the means model can also be written using $\mu_i = \mu + \tau_i$, where μ represents the overall mean and τ_i represents the effect due to the ith treatment.

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

Such an ANOVA model is called the *effects model*. In the effects models the treatment effects, τ_i, represent the deviations from the overall mean, μ. Therefore, the following constraint exists on the τ_i s:

$$\sum_{i=1}^{n_a} \tau_i = 0$$

Fitting ANOVA Models

To fit ANOVA models and carry out hypothesis testing in single factor experiments, it is convenient to express the effects model of the effects model in the form $y = X\beta + \epsilon$ (that was used for multiple linear regression models in Multiple Linear Regression Analysis). This can be done as shown next. Using the effects model, the ANOVA model for the single factor experiment in the first table can be expressed as:

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

where μ represents the overall mean and τ_i represents the ith treatment effect. There are three treatments in the first table (500, 600 and 700). Therefore, there are three treatment effects, τ_1,

and τ_3 . The following constraint exists for these effects:

$$\sum_{i=1}^{3} \tau_i = 0$$
$$\text{or } \tau_1 + \tau_2 + \tau_3 = 0$$

For the first treatment, the ANOVA model for the single factor experiment in the above table can be written as:

$$Y_{1j} = \mu + \tau_1 + 0 \cdot \tau_2 + 0 \cdot \tau_3 + \epsilon_{1j}$$

Using $\tau_3 = -(\tau_1 + \tau_2)$, the model for the first treatment is:

$$Y_{1j} = \mu + \tau_1 + 0 \cdot \tau_2 - 0 \cdot (\tau_1 + \tau_2) + \epsilon_{1j}$$
$$\text{or } Y_{1j} = \mu + \tau_1 + 0 \cdot \tau_2 + \epsilon_{1j}$$

Models for the second and third treatments can be obtained in a similar way. The models for the three treatments are:

$$\text{First Treatment : } Y_{1j} = 1 \cdot \mu + 1 \cdot \tau_1 + 0 \cdot \tau_2 + \epsilon_{1j}$$
$$\text{Second Treatment : } Y_{2j} = 1 \cdot \mu + 0 \cdot \tau_1 + 1 \cdot \tau_2 + \epsilon_{2j}$$
$$\text{Third Treatment : } Y_{3j} = 1 \cdot \mu - 1 \cdot \tau_1 - 1 \cdot \tau_2 + \epsilon_{3j}$$

The coefficients of the treatment effects τ_1 and τ_2 can be expressed using two indicator variables, x_1 and x_2 , as follows:

$$\text{Treatment Effect } \tau_1 : x_1 = 1,\ x_2 = 0$$
$$\text{Treatment Effect } \tau_2 : x_1 = 0,\ x_2 = 1$$
$$\text{Treatment Effect } \tau_3 : x_1 = -1,\ x_2 = -1$$

Using the indicator variables x_1 and x_2 , the ANOVA model for the data in the first table now becomes:

$$Y = \mu + x_1 \cdot \tau_1 + x_2 \cdot \tau_2 + \epsilon$$

The equation can be rewritten by including subscripts i (for the level of the factor) and j (for the replicate number) as:

$$Y_{ij} = \mu + x_{i1} \cdot \tau_1 + x_{i2} \cdot \tau_2 + \epsilon_{ij}$$

The equation given above represents the "regression version" of the ANOVA model.

Treat Numerical Factors as Qualitative or Quantitative?

It can be seen from the equation given above that in an ANOVA model each factor is treated as a qualitative factor. In the present example the factor, lathe speed, is a quantitative factor with three levels. But the ANOVA model treats this factor as a qualitative factor with three levels. Therefore, two indicator variables, x_1 and x_2, are required to represent this factor.

Note that in a regression model a variable can either be treated as a quantitative or a qualitative variable. The factor, lathe speed, would be used as a quantitative factor and represented with a single predictor variable in a regression model. For example, if a first order model were to be fitted to the data in the first table, then the regression model would take the form $Y_{ij} = \beta_0 + \beta_1 x_{i1} + \epsilon_{ij}$. If a second order regression model were to be fitted, the regression model would be $Y_{ij} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i1}^2 + \epsilon_{ij}$. Notice that unlike these regression models, the regression version of the ANOVA model does not make any assumption about the nature of relationship between the response and the factor being investigated.

The choice of treating a particular factor as a quantitative or qualitative variable depends on the objective of the experimenter. In the case of the data of the first table, the objective of the experimenter is to compare the levels of the factor to see if change in the levels leads to a significant change in the response. The objective is not to make predictions on the response for a given level of the factor. Therefore, the factor is treated as a qualitative factor in this case. If the objective of the experimenter were prediction or optimization, the experimenter would focus on aspects such as the nature of relationship between the factor, lathe speed, and the response, surface finish, so that the factor should be modeled as a quantitative factor to make accurate predictions.

Expression of the ANOVA Model as $Y = XB + \varepsilon$

The regression version of the ANOVA model can be expanded for the three treatments and four replicates of the data in the first table as follows:

$$
\begin{aligned}
Y_{11} &= 6 = \mu + 1 \cdot \tau_1 + 0 \cdot \tau_2 + \epsilon_{11} \text{ Level 1, Replicate 1} \\
Y_{21} &= 13 = \mu + 0 \cdot \tau_1 + 1 \cdot \tau_2 + \epsilon_{21} \text{ Level 2, Replicate 1} \\
Y_{31} &= 23 = \mu - 1 \cdot \tau_1 - 1 \cdot \tau_2 + \epsilon_{31} \text{ Level 3, Replicate 1} \\
Y_{12} &= 13 = \mu + 1 \cdot \tau_1 + 0 \cdot \tau_2 + \epsilon_{12} \text{ Level 1, Replicate 2} \\
Y_{22} &= 16 = \mu + 0 \cdot \tau_1 + 1 \cdot \tau_2 + \epsilon_{22} \text{ Level 2, Replicate 2} \\
Y_{32} &= 20 = \mu - 1 \cdot \tau_1 - 1 \cdot \tau_2 + \epsilon_{32} \text{ Level 3, Replicate 2} \\
&\dots \\
Y_{34} &= 18 = \mu - 1 \cdot \tau_1 - 1 \cdot \tau_2 + \epsilon_{34} \text{ Level 3, Replicate 4}
\end{aligned}
$$

The corresponding matrix notation is:

$$y = X\beta + \epsilon$$

where

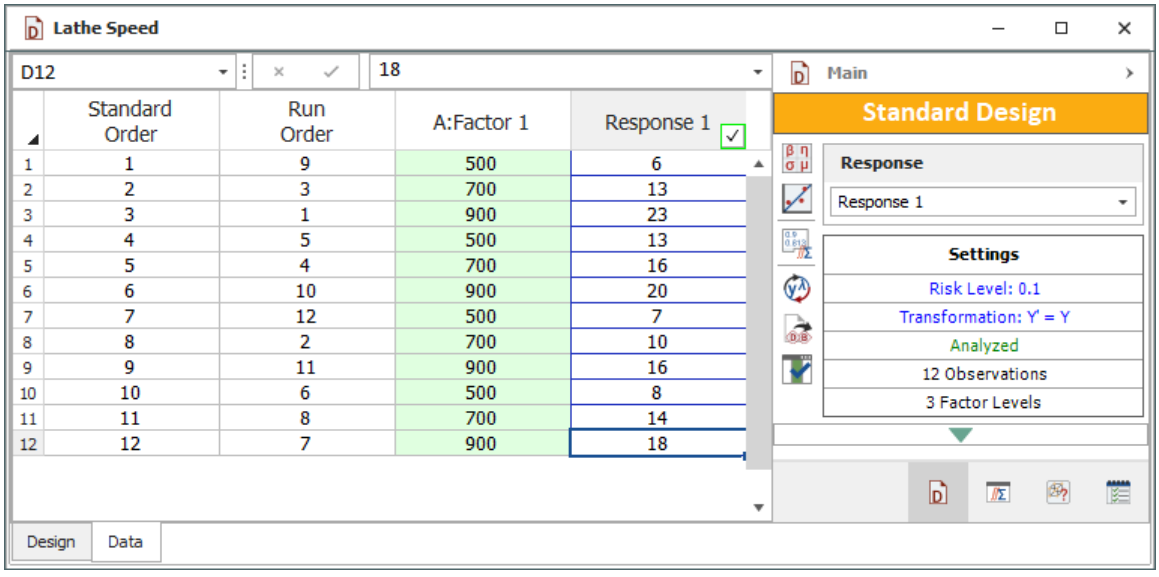
$$y = \begin{bmatrix} Y_{11} \\ Y_{21} \\ Y_{31} \\ Y_{12} \\ Y_{22} \\ . \\ . \\ . \\ Y_{34} \end{bmatrix} = X\beta + \epsilon = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ . & . & . \\ . & . & . \\ . & . & . \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{21} \\ \epsilon_{31} \\ \epsilon_{12} \\ \epsilon_{22} \\ . \\ . \\ . \\ \epsilon_{34} \end{bmatrix}$$

Thus:

$$y = X\beta + \epsilon$$

$$\begin{bmatrix} 6 \\ 13 \\ 23 \\ 13 \\ 16 \\ . \\ . \\ . \\ 18 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ . & . & . \\ . & . & . \\ . & . & . \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \end{bmatrix} + \begin{bmatrix} \epsilon_{11} \\ \epsilon_{21} \\ \epsilon_{31} \\ \epsilon_{12} \\ \epsilon_{22} \\ . \\ . \\ . \\ \epsilon_{34} \end{bmatrix}$$

The matrices y, X and β are used in the calculation of the sum of squares in the next section. The data in the first table can be entered into the DOE folio as shown in the figure below.



	Standard Order	Run Order	A:Factor 1	Response 1
1	1	9	500	6
2	2	3	700	13
3	3	1	900	23
4	4	5	500	13
5	5	4	700	16
6	6	10	900	20
7	7	12	500	7
8	8	2	700	10
9	9	11	900	16
10	10	6	500	8
11	11	8	700	14
12	12	7	900	18

Hypothesis Test in Single Factor Experiments

The hypothesis test in single factor experiments examines the ANOVA model to see if the response at any level of the investigated factor is significantly different from that at the other levels. If this is not the case and the response at all levels is not significantly different, then it can be concluded that the investigated factor does not affect the response. The test on the ANOVA model is carried out by checking to see if any of the treatment effects, τ_i, are non-zero. The test is similar to the test of significance of regression mentioned in Simple Linear Regression Analysis and Multiple Linear Regression Analysis in the context of regression models. The hypotheses statements for this test are:

$$H_0: \quad \tau_1 = \tau_2 = \ldots = \tau_{n_a} = 0$$
$$H_1: \quad \tau_i \neq 0 \text{ for at least one } i$$

The test for H_0 is carried out using the following statistic:

$$F_0 = \frac{MS_{TR}}{MS_E}$$

where MS_{TR} represents the mean square for the ANOVA model and MS_E is the error mean square. Note that in the case of ANOVA models we use the notation MS_{TR} (treatment mean square) for the model mean square and SS_{TR} (treatment sum of squares) for the model sum of squares (instead of MS_R, regression mean square, and SS_R, regression sum of squares, used in Simple Linear Regression Analysis and Multiple Linear Regression Analysis). This is done to indicate that the model under consideration is the ANOVA model and not the regression model. The calculations to obtain MS_{TR} and SS_{TR} are identical to the calculations to obtain MS_R and SS_R explained in Multiple Linear Regression Analysis.

Calculation of the Statistic F_0

The sum of squares to obtain the statistic F_0 can be calculated as explained in Multiple Linear Regression Analysis. Using the data in the first table, the model sum of squares, SS_{TR}, can be calculated as:

$$SS_{TR} = y'[H - (\frac{1}{n_a \cdot m})J]y$$
$$= \begin{bmatrix} 6 \\ 13 \\ . \\ . \\ 18 \end{bmatrix}' \begin{bmatrix} 0.1667 & -0.0833 & . & . & -0.0833 \\ -0.0833 & 0.1667 & . & . & -0.0833 \\ . & . & . & . & . \\ . & . & . & . & . \\ -0.0833 & -0.0833 & . & . & 0.1667 \end{bmatrix} \begin{bmatrix} 6 \\ 13 \\ . \\ . \\ 18 \end{bmatrix}$$
$$= 232.1667$$

In the previous equation, n_a represents the number of levels of the factor, m represents the replicates at each level, y represents the vector of the response values, H represents the hat matrix and J represents the matrix of ones. (For details on each of these terms, refer to Multiple Linear Regression Analysis.) Since two effect terms, τ_1 and τ_2, are used in the regression version of the ANOVA model, the degrees of freedom associated with the model sum of squares, SS_{TR}, is two.

$$dof(SS_{TR}) = 2$$

The total sum of squares, SS_T, can be obtained as follows:

$$\begin{aligned} SS_T &= y'[I - (\frac{1}{n_a \cdot m})J]y \\ &= \begin{bmatrix} 6 \\ 13 \\ . \\ . \\ 18 \end{bmatrix}' \begin{bmatrix} 0.9167 & -0.0833 & . & . & -0.0833 \\ -0.0833 & 0.9167 & . & . & -0.0833 \\ . & . & . & . & . \\ . & . & . & . & . \\ -0.0833 & -0.0833 & . & . & 0.9167 \end{bmatrix} \begin{bmatrix} 6 \\ 13 \\ . \\ . \\ 18 \end{bmatrix} \\ &= 306.6667 \end{aligned}$$

In the previous equation, I is the identity matrix. Since there are 12 data points in all, the number of degrees of freedom associated with SS_T is 11.

$$dof(SS_T) = 11$$

Knowing SS_T and SS_{TR}, the error sum of squares is:

$$\begin{aligned} SS_E &= SS_T - SS_{TR} \\ &= 306.6667 - 232.1667 \\ &= 74.5 \end{aligned}$$

The number of degrees of freedom associated with SS_E is:

$$\begin{aligned} dof(SS_E) &= dof(SS_T) - dof(SS_{TR}) \\ &= 11 - 2 \\ &= 9 \end{aligned}$$

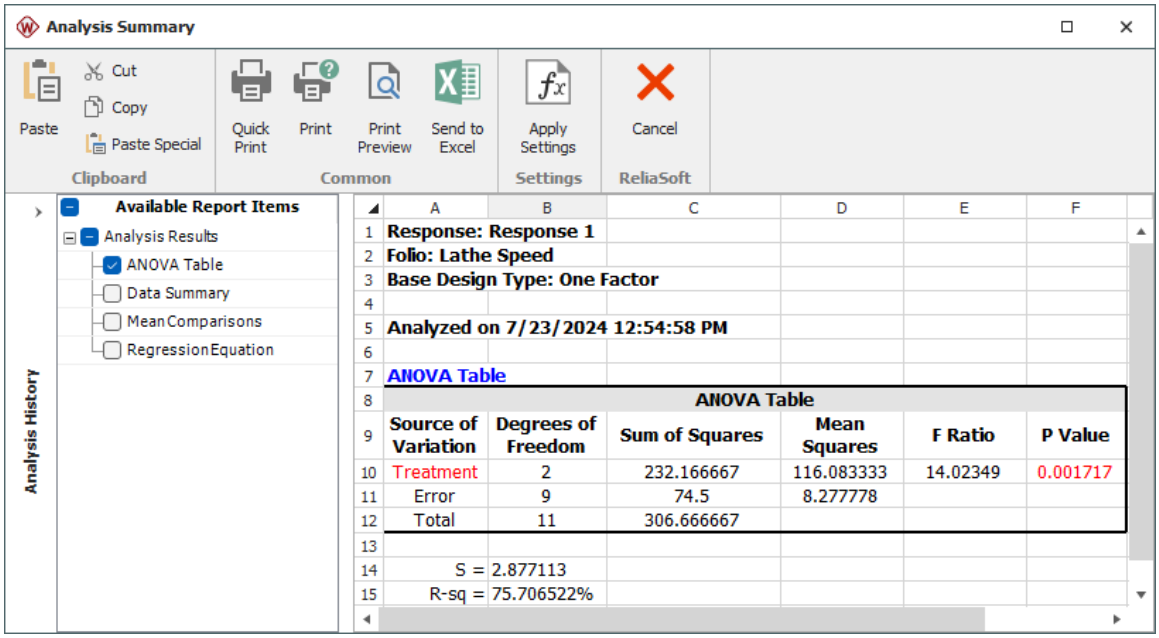
The test statistic can now be calculated using the equation given in Hypothesis Test in Single Factor Experiments as:

$$
\begin{aligned}
f_0 &= \frac{MS_{TR}}{MS_E} \\
&= \frac{SS_{TR}/dof(SS_{TR})}{SS_E/dof(SS_E)} \\
&= \frac{232.1667/2}{74.5/9} \\
&= 14.0235
\end{aligned}
$$

The p value for the statistic based on the F distribution with 2 degrees of freedom in the numerator and 9 degrees of freedom in the denominator is:

$$
\begin{aligned}
p\,value &= 1 - P(F \leq f_0) \\
&= 1 - 0.9983 \\
&= 0.0017
\end{aligned}
$$

Assuming that the desired significance level is 0.1, since p value < 0.1, H_0 is rejected and it is concluded that change in the lathe speed has a significant effect on the surface finish. The Weibull++ DOE folio displays these results in the ANOVA table, as shown in the figure below. The values of S and R-sq are the standard error and the coefficient of determination for the model, respectively. These values are explained in Multiple Linear Regression Analysis and indicate how well the model fits the data. The values in the figure below indicate that the fit of the ANOVA model is fair.



Confidence Interval on the *i*^th Treatment Mean

The response at each treatment of a single factor experiment can be assumed to be a normal population with a mean of μ_i and variance of σ^2 provided that the error terms can be assumed to be normally distributed. A point estimator of μ_i is the average response at each treatment, $\bar{y}_{i.}$. Since this is a sample average, the associated variance is σ^2/m_i, where m_i is the number of replicates at the ith treatment. Therefore, the confidence interval on μ_i is based on the t distribution. Recall from Statistical Background on DOE (inference on population mean when variance is unknown) that:

$$\begin{aligned} T_0 &= \frac{\bar{y}_{i.} - \mu_i}{\sqrt{\hat{\sigma}^2/m_i}} \\ &= \frac{\bar{y}_{i.} - \mu_i}{\sqrt{MS_E/m_i}} \end{aligned}$$

has a t distribution with degrees of freedom $= dof(SS_E)$. Therefore, a 100 $(1-\alpha)$ percent confidence interval on the ith treatment mean, μ_i, is:

$$\bar{y}_{i.} \pm t_{\alpha/2,dof(SS_E)}\sqrt{\frac{MS_E}{m_i}}$$

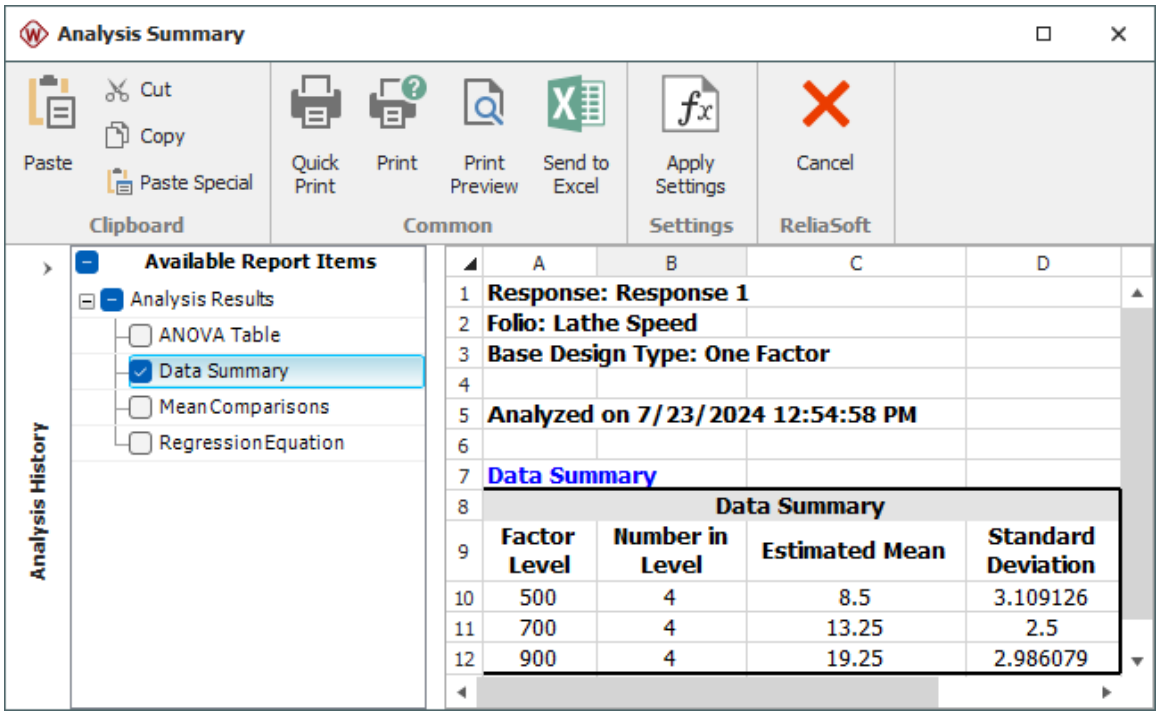
For example, for the first treatment of the lathe speed we have:

$$\begin{aligned} \hat{\mu}_1 &= \bar{y}_{1.} \\ &= \frac{6+13+7+8}{4} \\ &= 8.5 \end{aligned}$$

In the DOE folio, this value is displayed as the Estimated Mean for the first level, as shown in the Data Summary table in the figure below. The value displayed as the standard deviation for this level is simply the sample standard deviation calculated using the observations corresponding to this level. The 90% confidence interval for this treatment is:

$$\begin{aligned} &= \bar{y}_{1.} \pm t_{\alpha/2,dof(SS_E)}\sqrt{\frac{MS_E}{m_i}} \\ &= \bar{y}_{1.} \pm t_{0.05,9}\sqrt{\frac{MS_E}{4}} \\ &= 8.5 \pm 1.833\sqrt{\frac{(74.5/9)}{4}} \\ &= 8.5 \pm 1.833(1.44) \\ &= 8.5 \pm 2.64 \end{aligned}$$

The 90% limits on μ_1 are 5.9 and 11.1, respectively.



Confidence Interval on the Difference in Two Treatment Means

The confidence interval on the difference in two treatment means, $\mu_i - \mu_j$, is used to compare two levels of the factor at a given significance. If the confidence interval does not include the value of zero, it is concluded that the two levels of the factor are significantly different. The point estimator of $\mu_i - \mu_j$ is $\bar{y}_{i\cdot} - \bar{y}_{j\cdot}$. The variance for $\bar{y}_{i\cdot} - \bar{y}_{j\cdot}$ is:

$$\begin{aligned} var(\bar{y}_{i\cdot} - \bar{y}_{j\cdot}) &= var(\bar{y}_{i\cdot}) + var(\bar{y}_{j\cdot}) \\ &= \sigma^2/m_i + \sigma^2/m_j \end{aligned}$$

For balanced designs all $m_i = m$. Therefore:

$$var(\bar{y}_{i\cdot} - \bar{y}_{j\cdot}) = 2\sigma^2/m$$

The standard deviation for $\bar{y}_{i\cdot} - \bar{y}_{j\cdot}$ can be obtained by taking the square root of $var(\bar{y}_{i\cdot} - \bar{y}_{j\cdot})$ and is referred to as the pooled standard error:

$$\begin{aligned} Pooled\ Std.\ Error &= \sqrt{var(\bar{y}_{i\cdot} - \bar{y}_{j\cdot})} \\ &= \sqrt{2\hat{\sigma}^2/m} \end{aligned}$$

The t statistic for the difference is:

$$T_0 = \frac{\bar{y}_{i.} - \bar{y}_{j.} - (\mu_i - \mu_j)}{\sqrt{2\hat{\sigma}^2/m}}$$
$$= \frac{\bar{y}_{i.} - \bar{y}_{j.} - (\mu_i - \mu_j)}{\sqrt{2MS_E/m}}$$

Then a 100 (1- α) percent confidence interval on the difference in two treatment means, $\mu_i - \mu_j$, is:

$$\bar{y}_{i.} - \bar{y}_{j.} \pm t_{\alpha/2,dof(SS_E)}\sqrt{\frac{2MS_E}{m}}$$

For example, an estimate of the difference in the first and second treatment means of the lathe speed, $\mu_1 - \mu_2$, is:

$$\hat{\mu}_1 - \hat{\mu}_2 = \bar{y}_{1.} - \bar{y}_{2.}$$
$$= 8.5 - 13.25$$
$$= -4.75$$

The pooled standard error for this difference is:

$$Pooled\ Std.\ Error = \sqrt{var(\bar{y}_{1.} - \bar{y}_{2.})}$$
$$= \sqrt{2\hat{\sigma}^2/m}$$
$$= \sqrt{2MS_E/m}$$
$$= \sqrt{\frac{2(74.5/9)}{4}}$$
$$= 2.0344$$

To test $H_0 : \mu_1 - \mu_2 = 0$, the t statistic is:

$$t_0 = \frac{\bar{y}_{1.} - \bar{y}_{2.} - (\mu_1 - \mu_2)}{\sqrt{2MS_E/m}}$$
$$= \frac{-4.75 - (0)}{\sqrt{\frac{2(74.5/9)}{4}}}$$
$$= \frac{-4.75}{2.0344}$$
$$= -2.3348$$

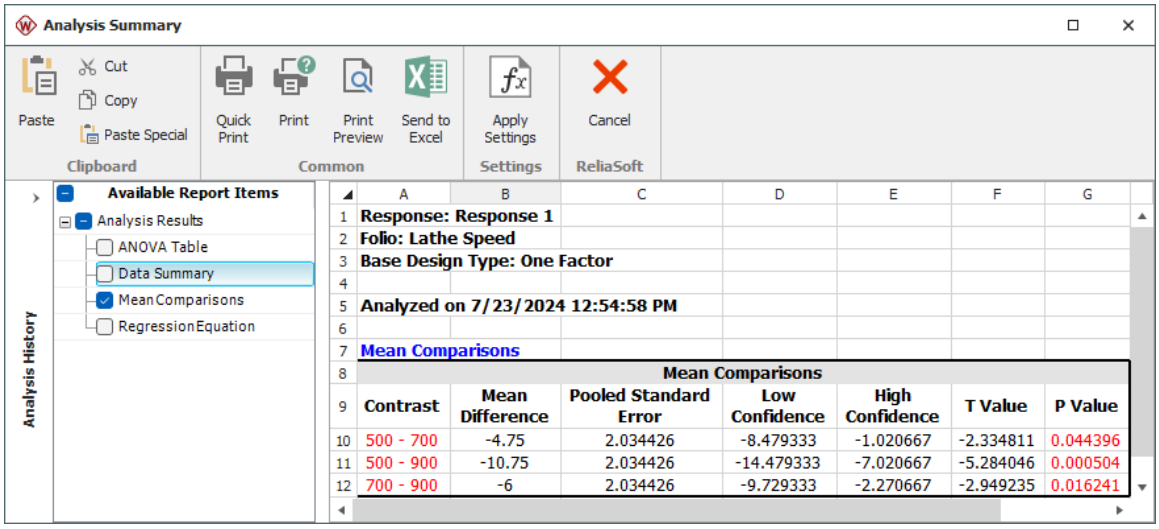
In the DOE folio, the value of the statistic is displayed in the Mean Comparisons table under the column T Value as shown in the figure below. The 90% confidence interval on the difference $\mu_1 - \mu_2$ is:

$$
\begin{aligned}
&= \bar{y}_{1.} - \bar{y}_{2.} \pm t_{\alpha/2, dof(SS_E)} \sqrt{\frac{2MS_E}{m}} \\
&= \bar{y}_{1.} - \bar{y}_{2.} \pm t_{0.05,9} \sqrt{\frac{2MS_E}{m}} \\
&= -4.75 \pm 1.833 \sqrt{\frac{2(74.5/9)}{4}} \\
&= -4.75 \pm 1.833(2.0344) \\
&= -4.75 \pm 3.729
\end{aligned}
$$

Hence the 90% limits on $\mu_1 - \mu_2$ are -8.479 and -1.021, respectively. These values are displayed under the Low CI and High CI columns in the following figure. Since the confidence interval for this pair of means does not included zero, it can be concluded that these means are significantly different at 90% confidence. This conclusion can also be arrived at using the p value noting that the hypothesis is two-sided. The p value corresponding to the statistic $t_0 = -2.3348$, based on the t distribution with 9 degrees of freedom is:

$$
\begin{aligned}
p\ value &= 2 \times (1 - P(T \leq |t_0|)) \\
&= 2 \times (1 - P(T \leq 2.3348)) \\
&= 2 \times (1 - 0.9778) \\
&= 0.0444
\end{aligned}
$$

Since p value < 0.1, the means are significantly different at 90% confidence. Bounds on the difference between other treatment pairs can be obtained in a similar manner and it is concluded that all treatments are significantly different.



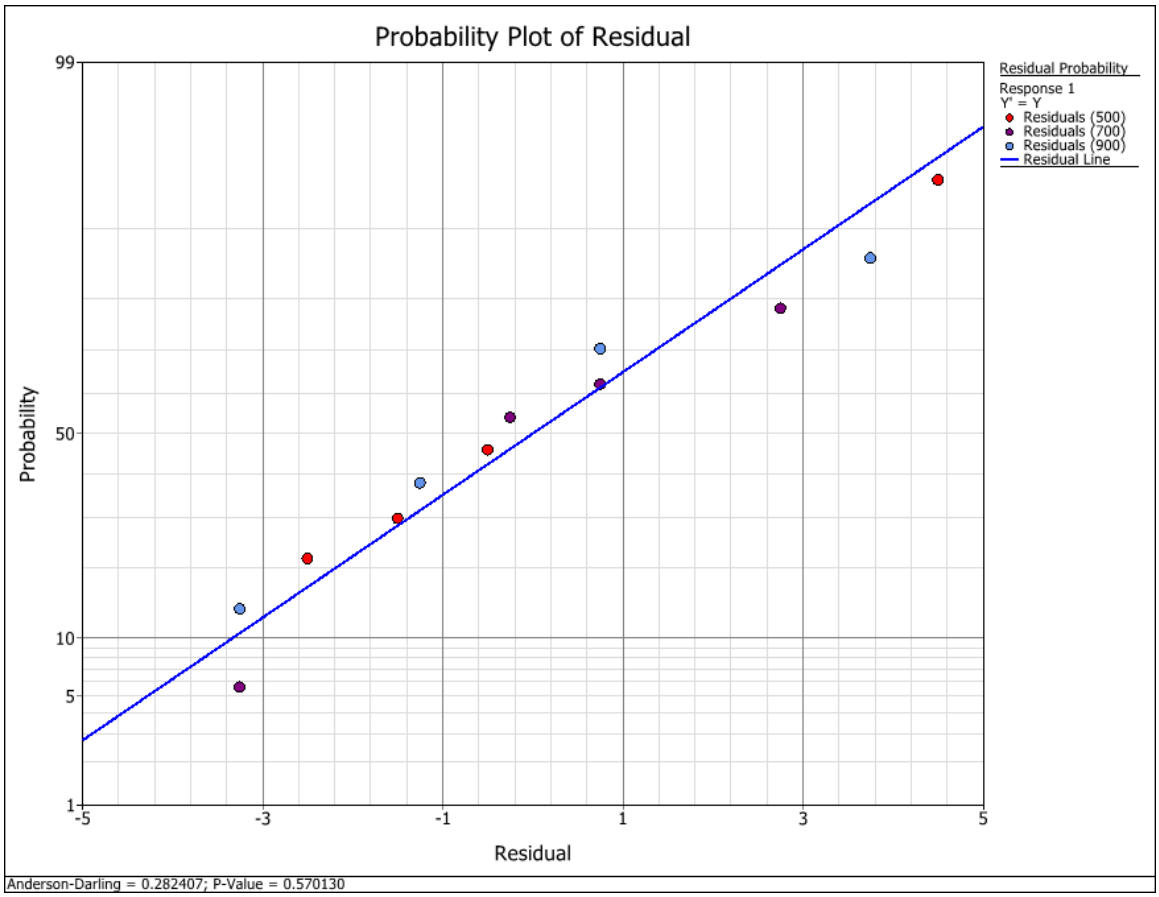
Mean Comparisons

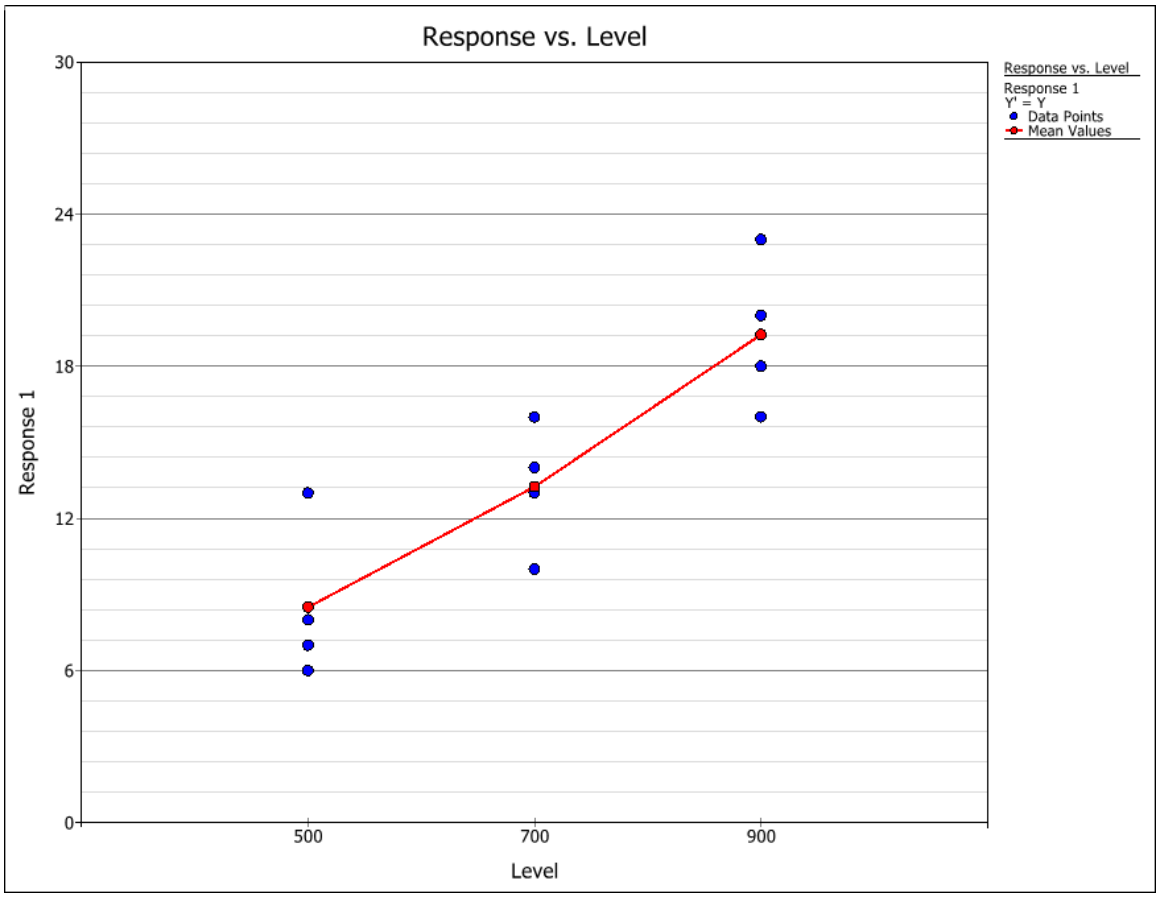
Contrast	Mean Difference	Pooled Standard Error	Low Confidence	High Confidence	T Value	P Value
500 - 700	-4.75	2.034426	-8.479333	-1.020667	-2.334811	0.044396
500 - 900	-10.75	2.034426	-14.479333	-7.020667	-5.284046	0.000504
700 - 900	-6	2.034426	-9.729333	-2.270667	-2.949235	0.016241

Residual Analysis

Plots of residuals, e_{ij} , similar to the ones discussed in the previous chapters on regression, are used to ensure that the assumptions associated with the ANOVA model are not violated. The ANOVA model assumes that the random error terms, ϵ_{ij} , are normally and independently distributed with the same variance for each treatment. The normality assumption can be checked by obtaining a normal probability plot of the residuals.

Equality of variance is checked by plotting residuals against the treatments and the treatment averages, $\bar{y}_{i.}$ (also referred to as fitted values), and inspecting the spread in the residuals. If a pattern is seen in these plots, then this indicates the need to use a suitable transformation on the response that will ensure variance equality. Box-Cox transformations are discussed in the next section. To check for independence of the random error terms residuals are plotted against time or run-order to ensure that a pattern does not exist in these plots. Residual plots for the given example are shown in the following two figures. The plots show that the assumptions associated with the ANOVA model are not violated.





Box-Cox Method

Transformations on the response may be used when residual plots for an experiment show a pattern. This indicates that the equality of variance does not hold for the residuals of the given model. The Box-Cox method can be used to automatically identify a suitable power transformation for the data based on the following relationship:

$$Y^* = Y^\lambda$$

λ is determined using the given data such that SS_E is minimized. The values of Y^λ are not used as is because of issues related to calculation or comparison of SS_E values for different values of λ. For example, for $\lambda = 0$ all response values will become 1. Therefore, the following relationship is used to obtain Y^λ :

$$Y^\lambda = \begin{cases} \frac{y^\lambda - 1}{\lambda \dot{y}^{\lambda-1}} & \lambda \neq 0 \\ \dot{y} \ln y & \lambda = 0 \end{cases}$$

where $\dot{y} = \exp\left[\frac{1}{n}\sum_{i=1}^{n}\ln(y_i)\right]$. Once all Y^{λ} values are obtained for a value of λ, the corresponding SS_E for these values is obtained using $y^{\lambda\prime}[I - H]y^{\lambda}$. The process is repeated for a number of λ values to obtain a plot of SS_E against λ. Then the value of λ corresponding to the minimum SS_E is selected as the required transformation for the given data. The DOE folio plots $\ln SS_E$ values against λ values because the range of SS_E values is large and if this is not done, all values cannot be displayed on the same plot. The range of search for the best λ value in the software is from -5 to 5, because larger values of of λ are usually not meaningful. The DOE folio also displays a recommended transformation based on the best λ value obtained as shown in the following table.

Best Lambda	Recommended Transformation	Equation
$-2.5 < \lambda \leq -1.5$	Power $\lambda = -2$	$Y^* = \frac{1}{Y^2}$
$-1.5 < \lambda \leq -0.75$	Reciprocal $\lambda = -1$	$Y^* = \frac{1}{Y}$
$-0.75 < \lambda \leq -0.25$	Reciprocal Square Root $\lambda = -0.5$	$Y^* = \frac{1}{\sqrt{Y}}$
$-0.25 < \lambda \leq 0.25$	Natural Log $\lambda = 0$	$Y^* = \ln Y$
$0.25 < \lambda \leq 0.75$	Square Root $\lambda = 0.5$	$Y^* = \sqrt{Y}$
$0.75 < \lambda \leq 1.5$	None $\lambda = 1$	$Y^* = Y$
$1.5 < \lambda \leq 2.5$	Power $\lambda = 2$	$Y^* = Y^2$

Confidence intervals on the selected λ values are also available. Let $SS_E(\lambda)$ be the value of SS_E corresponding to the selected value of λ. Then, to calculate the 100 (1- α) percent confidence intervals on λ, we need to calculate SS^* as shown next:

$$SS^* = SS_E(\lambda)\left(1 + \frac{t^2_{\alpha/2,dof(SS_E)}}{dof(SS_E)}\right)$$

The required limits for λ are the two values of λ corresponding to the value SS^* (on the plot of SS_E against λ). If the limits for λ do not include the value of one, then the transformation is

applicable for the given data. Note that the power transformations are not defined for response values that are negative or zero. The DOE folio deals with negative and zero response values using the following equations (that involve addition of a suitable quantity to all of the response values if a zero or negative response value is encountered).

$$y(i) = y(i) + |y_{\min}| \times 1.1 \quad \text{Negative Response}$$
$$y(i) = y(i) + 1 \quad \text{Zero Response}$$

Here $y_{\min}$ represents the minimum response value and $|y_{\min}|$ represents the absolute value of the minimum response.

Example

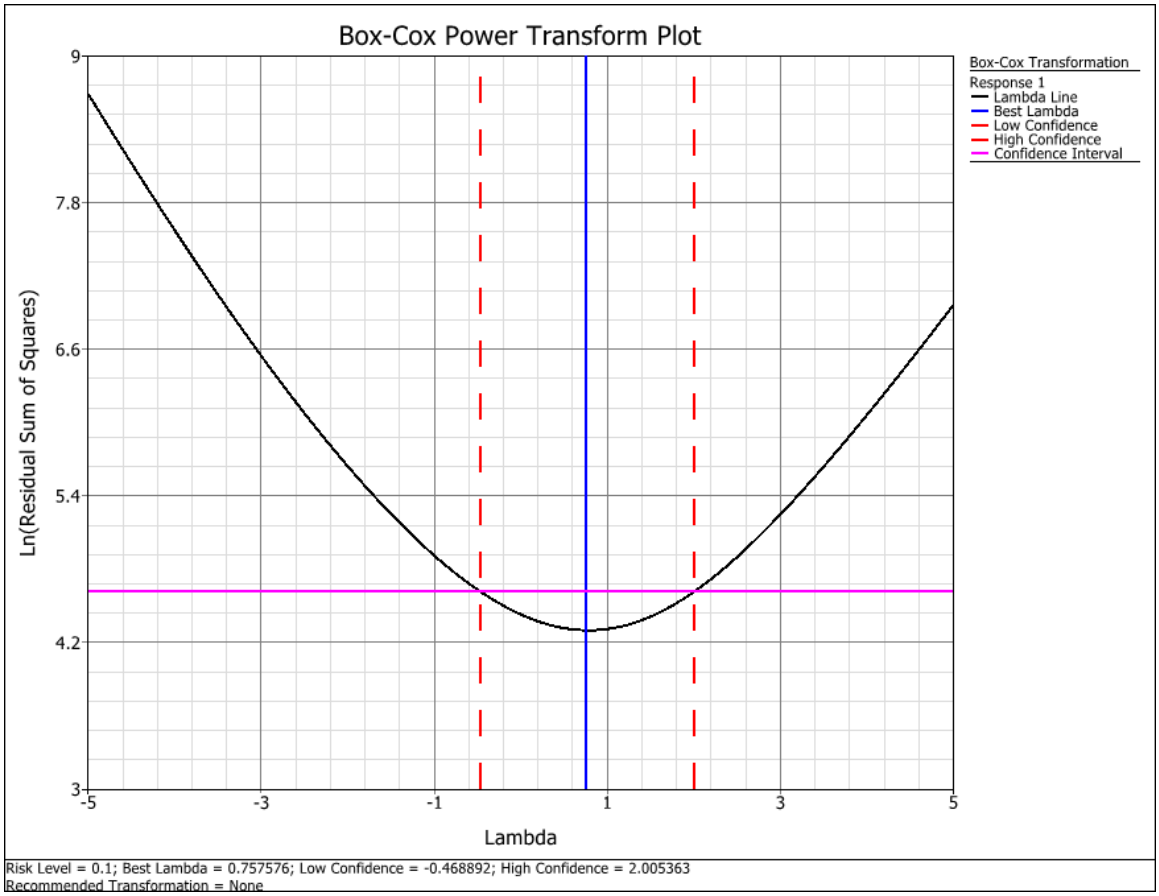
To illustrate the Box-Cox method, consider the experiment given in the first table. Transformed response values for various values of λ can be calculated using the equation for Y^{λ} given in Box-Cox Method. Knowing the hat matrix, H, SS_E values corresponding to each of these λ values can easily be obtained using $y^{\lambda}[I - H]y^{\lambda}$. SS_E values calculated for λ values between -5 and 5 for the given data are shown below:

λ	SS_E	$\ln SS_E$
−5	5947.8	8.6908
−4	1946.4	7.5737
−3	696.5	6.5461
−2	282.2	5.6425
−1	135.8	4.9114
0	83.9	4.4299
1	74.5	4.3108
2	101.0	4.6154
3	190.4	5.2491
4	429.5	6.0627
5	1057.6	−6.9638

A plot of $\ln SS_E$ for various λ values, as obtained from the DOE folio, is shown in the following figure. The value of λ that gives the minimum SS_E is identified as 0.7841. The SS_E value corresponding to this value of λ is 73.74. A 90% confidence interval on this λ value is calculated as follows. SS^* can be obtained as shown next:

$$\begin{aligned} SS^* &= SS_E(\lambda)\left(1+\frac{t^2_{\alpha/2,dof(SS_E)}}{dof(SS_E)}\right) \\ &= 73.74\left(1+\frac{t^2_{0.05,9}}{9}\right) \\ &= 73.74\left(1+\frac{1.833^2}{9}\right) \\ &= 101.27 \end{aligned}$$

Therefore, $\ln SS^* = 4.6178$. The λ values corresponding to this value from the following figure are -0.4689 and 2.0054. Therefore, the 90% confidence limits on are -0.4689 and 2.0054. Since the confidence limits include the value of 1, this indicates that a transformation is not required for the data in the first table.



General Full Factorial Designs

IN THIS CHAPTER

Investigating Factor Effects **129**

Investigating Interactions **130**

Analysis of General Factorial Experiments **132**
Hypothesis Tests in General Factorial Experiments 133
Least Squares Means 144
Residual Analysis 145

Experiments with two or more factors are encountered frequently. The best way to carry out such experiments is by using full factorial experiments. These are experiments in which all combinations of factors are investigated in each replicate of the experiment. Full factorial experiments are the only means to completely and systematically study interactions between factors in addition to identifying significant factors. One-factor-at-a-time experiments (where each factor is investigated separately by keeping all the remaining factors constant) do not reveal the interaction effects between the factors. Further, in one-factor-at-a-time experiments, full randomization is not possible.

To illustrate full factorial experiments, consider an experiment where the response is investigated for two factors, A and B. Assume that the response is studied at two levels of factor A with A_{low} representing the lower level of A and A_{high} representing the higher level. Similarly, let B_{low} and B_{high} represent the two levels of factor B that are being investigated in this experiment. Since there are two factors with two levels, a total of $2 \times 2 = 4$ combinations exist (A_{low} - B_{low}, A_{low} - B_{high}, A_{high} - B_{low}, A_{high} - B_{high}). Thus, four runs are required for each replicate if a factorial experiment is to be carried out in this case. Assume that the response values for each of these four possible combinations are obtained as shown in the next table.

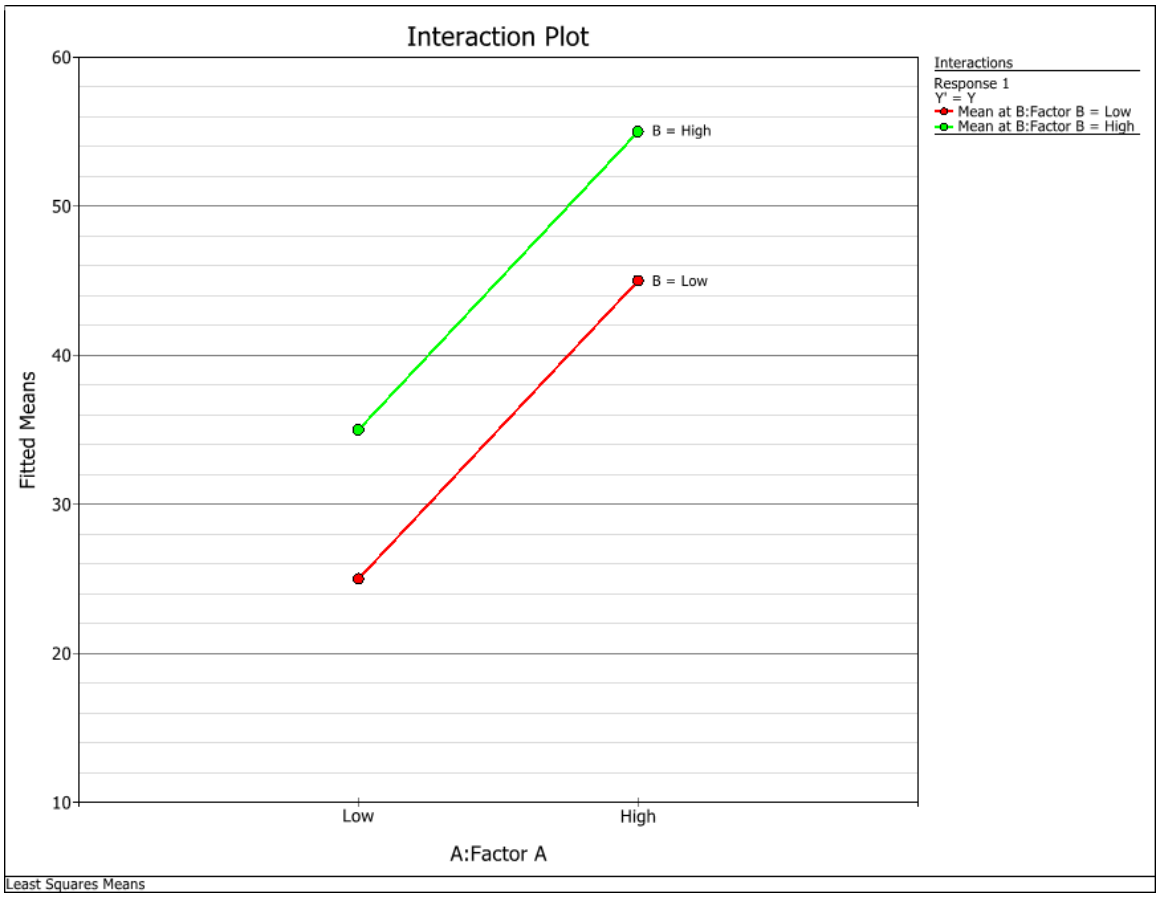
Factor A	Factor B	
	B_{low}	B_{high}
A_{low}	25	35
A_{high}	45	55

Investigating Factor Effects

The effect of factor A on the response can be obtained by taking the difference between the average response when A is high and the average response when A is low. The change in the response due to a change in the level of a factor is called the *main effect* of the factor. The main effect of A as per the response values in the third table is:

$$\begin{aligned} A &= Average\ response\ at\ A_{high} - Average\ response\ at\ A_{low} \\ &= \frac{45+55}{2} - \frac{25+35}{2} \\ &= 50 - 30 \\ &= 20 \end{aligned}$$

Therefore, when A is changed from the lower level to the higher level, the response increases by 20 units. A plot of the response for the two levels of A at different levels of B is shown next. The plot shows that change in the level of A leads to an increase in the response by 20 units regardless of the level of B. Therefore, no interaction exists in this case as indicated by the parallel lines on the plot.



The main effect of B can be obtained as:

$$\begin{aligned} B &= Average\ response\ at\ B_{high} - Average\ response\ at\ B_{low} \\ &= \frac{35+55}{2} - \frac{25+45}{2} \\ &= 45 - 35 \\ &= 10 \end{aligned}$$

Investigating Interactions

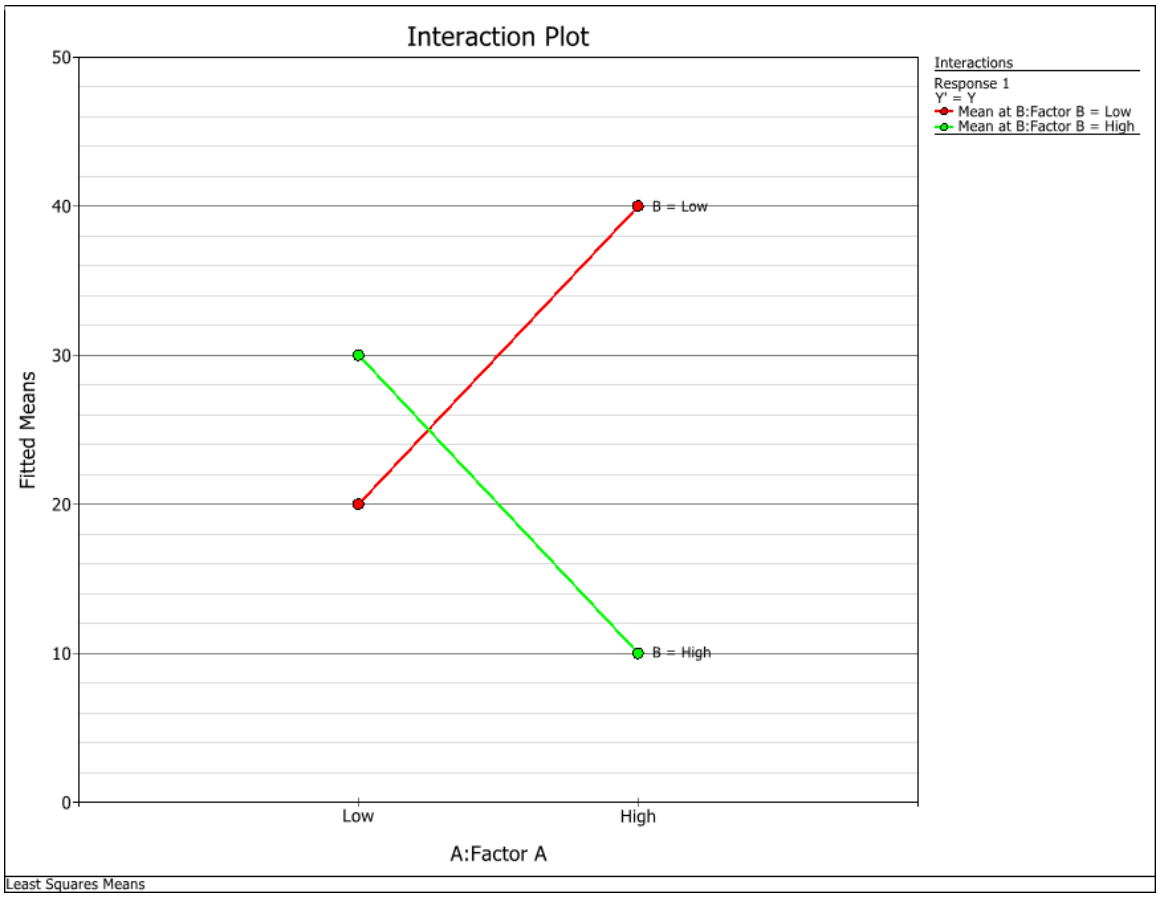
Now assume that the response values for each of the four treatment combinations were obtained as shown next.

Factor A	Factor B	
	B_{low}	B_{high}
A_{low}	20	30
A_{high}	40	10

The main effect of A in this case is:

$$\begin{aligned} A &= Average\ response\ at\ A_{high} - Average\ response\ at\ A_{low} \\ &= \frac{40+10}{2} - \frac{20+30}{2} \\ &= 0 \end{aligned}$$

It appears that A does not have an effect on the response. However, a plot of the response of A at different levels of B shows that the response does change with the levels of A but the effect of A on the response is dependent on the level of B (see the figure below).



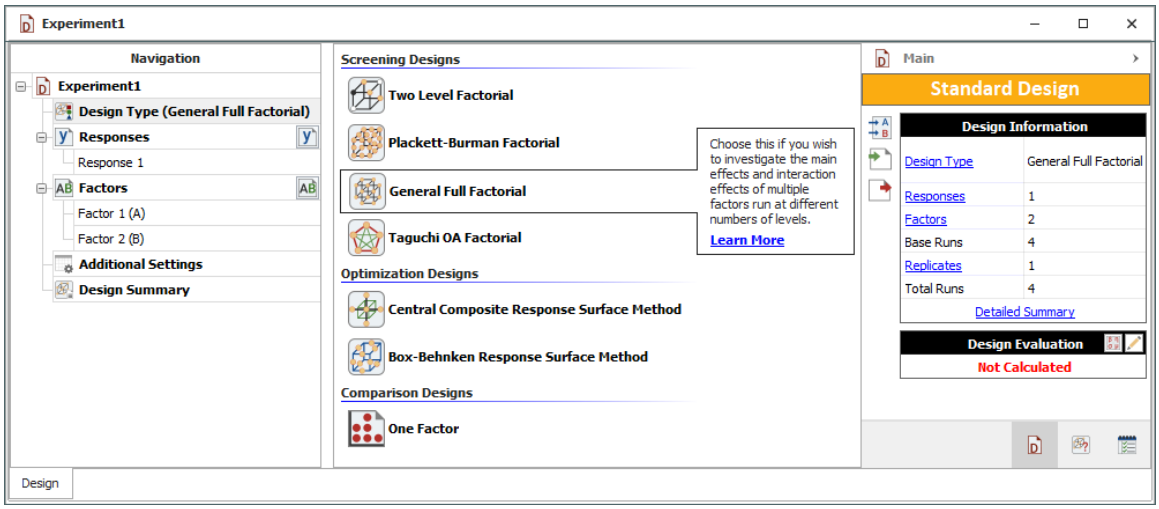
Therefore, an interaction between A and B exists in this case (as indicated by the non-parallel lines of the figure). The interaction effect between A and B can be calculated as follows:

$$\begin{aligned} AB &= Average\ response\ at\ A_{high}\text{-}B_{high}\ and\ A_{low}\text{-}B_{low} - \\ &\quad Average\ response\ at\ A_{low}\text{-}B_{high}\ and\ A_{high}\text{-}B_{low} \\ &= \frac{10+20}{2} - \frac{40+30}{2} \\ &= -20 \end{aligned}$$

Note that in this case, if a one-factor-at-a-time experiment were used to investigate the effect of factor A on the response, it would lead to incorrect conclusions. For example, if the response at factor A was studied by holding B constant at its lower level, then the main effect of A would be obtained as $40 - 20 = 20$, indicating that the response increases by 20 units when the level of A is changed from low to high. On the other hand, if the response at factor A was studied by holding B constant at its higher level than the main effect of A would be obtained as $10 - 30 = -20$, indicating that the response decreases by 20 units when the level of A is changed from low to high.

Analysis of General Factorial Experiments

In Weibull++ DOE folios, factorial experiments are referred to as *factorial designs*. The experiments explained in this section are referred to as *general factorial designs*. This is done to distinguish these experiments from the other factorial designs supported by Weibull++ DOE folios (see the figure below).



The other designs (such as the two level full factorial designs that are explained in Two Level Factorial Experiments) are special cases of these experiments in which factors are limited to a specified number of levels. The ANOVA model for the analysis of factorial experiments is formulated as shown next. Assume a factorial experiment in which the effect of two factors, A and B, on the response is being investigated. Let there be n_a levels of factor A and n_b levels of factor B. The ANOVA model for this experiment can be stated as:

$$Y_{ijk} = \mu + \tau_i + \delta_j + (\tau\delta)_{ij} + \epsilon_{ijk}$$

where:

- μ represents the overall mean effect
- τ_i is the effect of the ith level of factor A $(i = 1, 2, \ldots, n_a)$
- δ_j is the effect of the jth level of factor B $(j = 1, 2, \ldots, n_b)$
- $(\tau\delta)_{ij}$ represents the interaction effect between A and B
- ϵ_{ijk} represents the random error terms (which are assumed to be normally distributed with a mean of zero and variance of σ^2)
- and the subscript k denotes the m replicates $(k = 1, 2, \ldots, m)$

Since the effects τ_i, δ_j and $(\tau\delta)_{ij}$ represent deviations from the overall mean, the following constraints exist:

$$\sum_{i=1}^{n_a} \tau_i = 0$$

$$\sum_{j=1}^{n_b} \delta_j = 0$$

$$\sum_{i=1}^{n_a} (\tau\delta)_{ij} = 0$$

$$\sum_{j=1}^{n_b} (\tau\delta)_{ij} = 0$$

Hypothesis Tests in General Factorial Experiments

These tests are used to check whether each of the factors investigated in the experiment is significant or not. For the previous example, with two factors, A and B, and their interaction, AB, the statements for the hypothesis tests can be formulated as follows:

1. $H_0 : \tau_1 = \tau_2 = \ldots = \tau_{n_a} = 0$ (Main effect of A is absent)
   $H_1 : \tau_i \neq 0$ for at least one i
2. $H_0 : \delta_1 = \delta_2 = \ldots = \delta_{n_b} = 0$ (Main effect of B is absent)
   $H_1 : \delta_j \neq 0$ for at least one j
3. $H_0 : (\tau\delta)_{11} = (\tau\delta)_{12} = \ldots = (\tau\delta)_{n_a n_b} = 0$ (Interaction AB is absent)
   $H_1 : (\tau\delta)_{ij} \neq 0$ for at least one ij

The test statistics for the three tests are as follows:

1) $(F_0)_A = \frac{MS_A}{MS_E}$

where MS_A is the mean square due to factor A and MS_E is the error mean square.

2) $(F_0)_B = \frac{MS_B}{MS_E}$

where MS_B is the mean square due to factor B and MS_E is the error mean square.

3) $(F_0)_{AB} = \frac{MS_{AB}}{MS_E}$

where MS_{AB} is the mean square due to interaction AB and MS_E is the error mean square.

The tests are identical to the partial F test explained in Multiple Linear Regression Analysis. The sum of squares for these tests (to obtain the mean squares) are calculated by splitting the model sum of squares into the extra sum of squares due to each factor. The extra sum of squares calculated for each of the factors may either be partial or sequential. For the present example, if the extra sum of squares used is sequential, then the model sum of squares can be written as:

$$SS_{TR} = SS_A + SS_B + SS_{AB}$$

where SS_{TR} represents the model sum of squares, SS_A represents the sequential sum of squares due to factor A, SS_B represents the sequential sum of squares due to factor and SS_{AB} represents the sequential sum of squares due to the interaction AB. The mean squares are obtained by dividing the sum of squares by the associated degrees of freedom. Once the mean squares are known the test statistics can be calculated. For example, the test statistic to test the significance of factor A (or the hypothesis $H_0 : \tau_i = 0$) can then be obtained as:

$$\begin{aligned}(F_0)_A &= \frac{MS_A}{MS_E} \\ &= \frac{SS_A/dof(SS_A)}{SS_E/dof(SS_E)}\end{aligned}$$

Similarly the test statistic to test significance of factor B and the interaction AB can be respectively obtained as:

$$(F_0)_B = \frac{MS_B}{MS_E} = \frac{SS_B/dof(SS_B)}{SS_E/dof(SS_E)}$$

$$(F_0)_{AB} = \frac{MS_{AB}}{MS_E} = \frac{SS_{AB}/dof(SS_{AB})}{SS_E/dof(SS_E)}$$

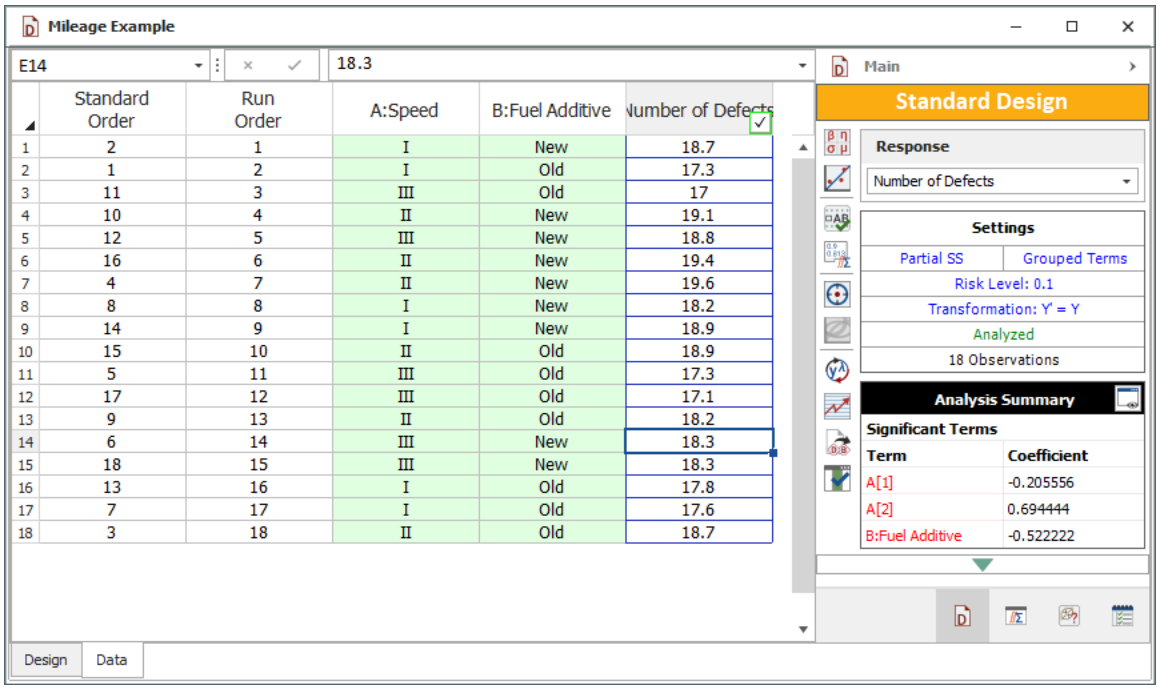
It is recommended to conduct the test for interactions before conducting the test for the main effects. This is because, if an interaction is present, then the main effect of the factor depends on the level of the other factors and looking at the main effect is of little value. However, if the interaction is absent then the main effects become important.

Example

Consider an experiment to investigate the effect of speed and type of fuel additive used on the mileage of a sport utility vehicle. Three speeds and two types of fuel additives are investigated. Each of the treatment combinations are replicated three times. The mileage values observed are displayed in the table below.

Speed	Fuel Additive	
	Type I	Type II
I	17.3, 17.8, 17.6	18.7, 18.2, 18.9
II	18.9, 18.2, 18.7	19.1, 19.4, 19.6
III	17.1, 17.0, 17.3	18.8, 18.3, 18.3

The experimental design for the data is shown in the figure below.



	Standard Order	Run Order	A:Speed	B:Fuel Additive	Number of Defects
1	2	1	I	New	18.7
2	1	2	I	Old	17.3
3	11	3	III	Old	17
4	10	4	II	New	19.1
5	12	5	III	New	18.8
6	16	6	II	New	19.4
7	4	7	II	New	19.6
8	8	8	I	New	18.2
9	14	9	I	New	18.9
10	15	10	II	Old	18.9
11	5	11	III	Old	17.3
12	17	12	III	Old	17.1
13	9	13	II	Old	18.2
14	6	14	III	New	18.3
15	18	15	III	New	18.3
16	13	16	I	Old	17.8
17	7	17	I	Old	17.6
18	3	18	II	Old	18.7

In the figure, the factor Speed is represented as factor A and the factor Fuel Additive is represented as factor B. The experimenter would like to investigate if speed, fuel additive or the interaction between speed and fuel additive affects the mileage of the sport utility vehicle. In other words, the following hypotheses need to be tested:

1. $H_0 : \tau_1 = \tau_2 = \tau_3 = 0$ (No main effect of factor A, speed)
   $H_1 : \tau_i \neq 0$ for at least one i
2. $H_0 : \delta_1 = \delta_2 = \delta_3 = 0$ (No main effect of factor B, fuel additive)
   $H_1 : \delta_j \neq 0$ for at least one j
3. $H_0 : (\tau\delta)_{11} = (\tau\delta)_{12} = \ldots = (\tau\delta)_{33} = 0$ (No interaction AB)
   $H_1 : (\tau\delta)_{ij} \neq 0$ for at least one ij

The test statistics for the three tests are:

1. $(F_0)_A = \dfrac{MS_A}{MS_E}$

   where MS_A is the mean square for factor A and MS_E is the error mean square

2. $(F_0)_B = \dfrac{MS_B}{MS_E}$

   where MS_B is the mean square for factor B and MS_E is the error mean square

3. $(F_0)_{AB} = \frac{MS_{AB}}{MS_E}$

where MS_{AB} is the mean square for interaction AB and MS_E is the error mean square

The ANOVA model for this experiment can be written as:

$$Y_{ijk} = \mu + \tau_i + \delta_j + (\tau\delta)_{ij} + \epsilon_{ijk}$$

where τ_i represents the ith treatment of factor A (speed) with i =1, 2, 3; δ_j represents the jth treatment of factor B (fuel additive) with j =1, 2; and $(\tau\delta)_{ij}$ represents the interaction effect. In order to calculate the test statistics, it is convenient to express the ANOVA model of the equation given above in the form $y = X\beta + \epsilon$. This can be done as explained next.

Expression of the ANOVA Model as *y = XB + ε*

Since the effects τ_i, δ_j and $(\tau\delta)_{ij}$ represent deviations from the overall mean, the following constraints exist. Constraints on τ_i are:

$$\sum_{i=1}^{3} \tau_i = 0$$
$$\text{or } \tau_1 + \tau_2 + \tau_3 = 0$$

Therefore, only two of the τ_i effects are independent. Assuming that τ_1 and τ_2 are independent, $\tau_3 = -(\tau_1 + \tau_2)$. (The null hypothesis to test the significance of factor A can be rewritten using only the independent effects as $H_0 : \tau_1 = \tau_2 = 0$.) The DOE folio displays only the independent effects because only these effects are important to the analysis. The independent effects, τ_1 and τ_2, are displayed as A[1] and A[2] respectively because these are the effects associated with factor A (speed). Constraints on δ_j are:

$$\sum_{j=1}^{2} \delta_j = 0$$
$$\text{or } \delta_1 + \delta_2 = 0$$

Therefore, only one of the δ_j effects are independent. Assuming that δ_1 is independent, $\delta_2 = -\delta_1$. (The null hypothesis to test the significance of factor B can be rewritten using only the independent effect as $H_0 : \delta_1 = 0$.) The independent effect δ_1 is displayed as B:B in the DOE folio. Constraints on $(\tau\delta)_{ij}$ are:

$$\sum_{i=1}^{3}(\tau\delta)_{ij} = 0$$
$$\text{and } \sum_{j=1}^{2}(\tau\delta)_{ij} = 0$$
$$\text{or } (\tau\delta)_{11} + (\tau\delta)_{21} + (\tau\delta)_{31} = 0$$
$$(\tau\delta)_{12} + (\tau\delta)_{22} + (\tau\delta)_{32} = 0$$
$$\text{and } (\tau\delta)_{11} + (\tau\delta)_{12} = 0$$
$$(\tau\delta)_{21} + (\tau\delta)_{22} = 0$$
$$(\tau\delta)_{31} + (\tau\delta)_{32} = 0$$

The last five equations given above represent four constraints, as only four of these five equations are independent. Therefore, only two out of the six $(\tau\delta)_{ij}$ effects are independent. Assuming that $(\tau\delta)_{11}$ and $(\tau\delta)_{21}$ are independent, the other four effects can be expressed in terms of these effects. (The null hypothesis to test the significance of interaction AB can be rewritten using only the independent effects as $H_0 : (\tau\delta)_{11} = (\tau\delta)_{21} = 0$.) The effects $(\tau\delta)_{11}$ and $(\tau\delta)_{21}$ are displayed as A[1]B and A[2]B respectively in the DOE folio.

The regression version of the ANOVA model can be obtained using indicator variables, similar to the case of the single factor experiment in Fitting ANOVA Models. Since factor A has three levels, two indicator variables, x_1 and x_2, are required which need to be coded as shown next:

$$\text{Treatment Effect } \tau_1 : x_1 = 1,\ x_2 = 0$$
$$\text{Treatment Effect } \tau_2 : x_1 = 0,\ x_2 = 1$$
$$\text{Treatment Effect } \tau_3 : x_1 = -1,\ x_2 = -1$$

Factor B has two levels and can be represented using one indicator variable, x_3, as follows:

$$\text{Treatment Effect } \delta_1 : x_3 = 1$$
$$\text{Treatment Effect } \delta_2 : x_3 = -1$$

The AB interaction will be represented by all possible terms resulting from the product of the indicator variables representing factors A and B. There are two such terms here - x_1x_3 and x_2x_3. The regression version of the ANOVA model can finally be obtained as:

$$Y = \mu + \tau_1 \cdot x_1 + \tau_2 \cdot x_2 + \delta_1 \cdot x_3 + (\tau\delta)_{11} \cdot x_1x_3 + (\tau\delta)_{21} \cdot x_2x_3 + \epsilon$$

In matrix notation this model can be expressed as:

$$y = X\beta + \epsilon$$

where:

$$y = \begin{bmatrix} Y_{111} \\ Y_{211} \\ Y_{311} \\ Y_{121} \\ Y_{221} \\ Y_{321} \\ Y_{112} \\ Y_{212} \\ . \\ . \\ Y_{323} \end{bmatrix} = X\beta + \epsilon = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & -1 & -1 & 1 & -1 & -1 \\ 1 & 1 & 0 & -1 & -1 & 0 \\ 1 & 0 & 1 & -1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & -1 & -1 & -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \delta_1 \\ (\tau\delta)_{11} \\ (\tau\delta)_{21} \end{bmatrix} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{211} \\ \epsilon_{311} \\ \epsilon_{121} \\ \epsilon_{221} \\ \epsilon_{321} \\ \epsilon_{112} \\ \epsilon_{212} \\ . \\ . \\ \epsilon_{323} \end{bmatrix}$$

The vector y can be substituted with the response values from the above table to get:

$$y = \begin{bmatrix} Y_{111} \\ Y_{211} \\ Y_{311} \\ Y_{121} \\ Y_{221} \\ Y_{321} \\ Y_{112} \\ Y_{212} \\ . \\ . \\ Y_{323} \end{bmatrix} = \begin{bmatrix} 17.3 \\ 18.9 \\ 17.1 \\ 18.7 \\ 19.1 \\ 18.8 \\ 17.8 \\ 18.2 \\ . \\ . \\ 18.3 \end{bmatrix}$$

Knowing y, X and β, the sum of squares for the ANOVA model and the extra sum of squares for each of the factors can be calculated. These are used to calculate the mean squares that are used to obtain the test statistics.

Calculation of Sum of Squares for the Model

The model sum of squares, SS_{TR}, for the regression version of the ANOVA model can be obtained as:

$$\begin{aligned} SS_{TR} &= y'[H - (\frac{1}{n_a \cdot n_b \cdot m})J]y \\ &= y'[H - (\frac{1}{18})J]y \\ &= 9.7311 \end{aligned}$$

where H is the hat matrix and J is the matrix of ones. Since five effect terms (τ_1, τ_2, δ_1, $(\tau\delta)_{11}$ and $(\tau\delta)_{21}$) are used in the model, the number of degrees of freedom associated with SS_{TR} is five ($dof(SS_{TR}) = 5$).

The total sum of squares, SS_T, can be calculated as:

$$\begin{aligned} SS_T &= y'[I - (\frac{1}{n_a \cdot n_b \cdot m})J]y \\ &= y'[I - (\frac{1}{18})J]y \\ &= 10.7178 \end{aligned}$$

Since there are 18 observed response values, the number of degrees of freedom associated with the total sum of squares is 17 ($dof(SS_T) = 17$). The error sum of squares can now be obtained:

$$\begin{aligned} SS_E &= SS_T - SS_{TR} \\ &= 10.7178 - 9.7311 \\ &= 0.9867 \end{aligned}$$

Since there are three replicates of the full factorial experiment, all of the error sum of squares is pure error. (This can also be seen from the preceding figure, where each treatment combination of the full factorial design is repeated three times.) The number of degrees of freedom associated with the error sum of squares is:

$$\begin{aligned} dof(SS_E) &= dof(SS_T) - dof(SS_{TR}) \\ &= 17 - 5 \\ &= 12 \end{aligned}$$

Calculation of Extra Sum of Squares for the Factors

The sequential sum of squares for factor A can be calculated as:

$$\begin{aligned} SS_A &= SS_{TR}(\mu, \tau_1, \tau_2) - SS_{TR}(\mu) \\ &= y'[H_{\mu,\tau_1,\tau_2} - (\frac{1}{18})J]y - 0 \end{aligned}$$

where $H_{\mu,\tau_1,\tau_2} = X_{\mu,\tau_1,\tau_2}(X'_{\mu,\tau_1,\tau_2}X_{\mu,\tau_1,\tau_2})^{-1}X'_{\mu,\tau_1,\tau_2}$ and X_{μ,τ_1,τ_2} is the matrix containing only the first three columns of the X matrix. Thus:

$$\begin{aligned} SS_A &= y'[H_{\mu,\tau_1,\tau_2} - (\frac{1}{18})J]y - 0 \\ &= 4.5811 - 0 \\ &= 4.5811 \end{aligned}$$

Since there are two independent effects (τ_1, τ_2) for factor A, the degrees of freedom associated with SS_A are two ($dof(SS_A) = 2$).

Similarly, the sum of squares for factor B can be calculated as:

$$\begin{aligned} SS_B &= SS_{TR}(\mu, \tau_1, \tau_2, \delta_1) - SS_{TR}(\mu, \tau_1, \tau_2) \\ &= y'[H_{\mu,\tau_1,\tau_2,\delta_1} - (\frac{1}{18})J]y - y'[H_{\mu,\tau_1,\tau_2} - (\frac{1}{18})J]y \\ &= 9.4900 - 4.5811 \\ &= 4.9089 \end{aligned}$$

Since there is one independent effect, δ_1, for factor B, the number of degrees of freedom associated with SS_B is one ($dof(SS_B) = 1$).

The sum of squares for the interaction AB is:

$$\begin{aligned} SS_{AB} &= SS_{TR}(\mu, \tau_1, \tau_2, \delta_1, (\tau\delta)_{11}, (\tau\delta)_{21}) - SS_{TR}(\mu, \tau_1, \tau_2, \delta_1) \\ &= SS_{TR} - SS_{TR}(\mu, \tau_1, \tau_2, \delta_1) \\ &= 9.7311 - 9.4900 \\ &= 0.2411 \end{aligned}$$

Since there are two independent interaction effects, $(\tau\delta)_{11}$ and $(\tau\delta)_{21}$, the number of degrees of freedom associated with SS_{AB} is two ($dof(SS_{AB}) = 2$).

Calculation of the Test Statistics

Knowing the sum of squares, the test statistic for each of the factors can be calculated. Analyzing the interaction first, the test statistic for interaction AB is:

$$\begin{aligned} (f_0)_{AB} &= \frac{MS_{AB}}{MS_E} \\ &= \frac{0.2411/2}{0.9867/12} \\ &= 1.47 \end{aligned}$$

The p value corresponding to this statistic, based on the F distribution with 2 degrees of freedom in the numerator and 12 degrees of freedom in the denominator, is:

$$\begin{aligned} p\ value &= 1 - P(F \leq (f_0)_{AB}) \\ &= 1 - 0.7307 \\ &= 0.2693 \end{aligned}$$

Assuming that the desired significance level is 0.1, since p value > 0.1, we fail to reject $H_0 : (\tau\delta)_{ij} = 0$ and conclude that the interaction between speed and fuel additive does not significantly affect the mileage of the sport utility vehicle. The DOE folio displays this result in the ANOVA table, as shown in the following figure. In the absence of the interaction, the analysis of main effects becomes important.

The test statistic for factor A is:

$$
\begin{aligned}
(f_0)_A &= \frac{MS_A}{MS_E} \\
&= \frac{SS_A/dof(SS_A)}{SS_E/dof(SS_E)} \\
&= \frac{4.5811/2}{0.9867/12} \\
&= 27.86
\end{aligned}
$$

The p value corresponding to this statistic based on the F distribution with 2 degrees of freedom in the numerator and 12 degrees of freedom in the denominator is:

$$
\begin{aligned}
p\,value &= 1 - P(F \leq (f_0)_A) \\
&= 1 - 0.99997 \\
&= 0.00003
\end{aligned}
$$

Since p value < 0.1, $H_0 : \tau_i = 0$ is rejected and it is concluded that factor A (or speed) has a significant effect on the mileage.

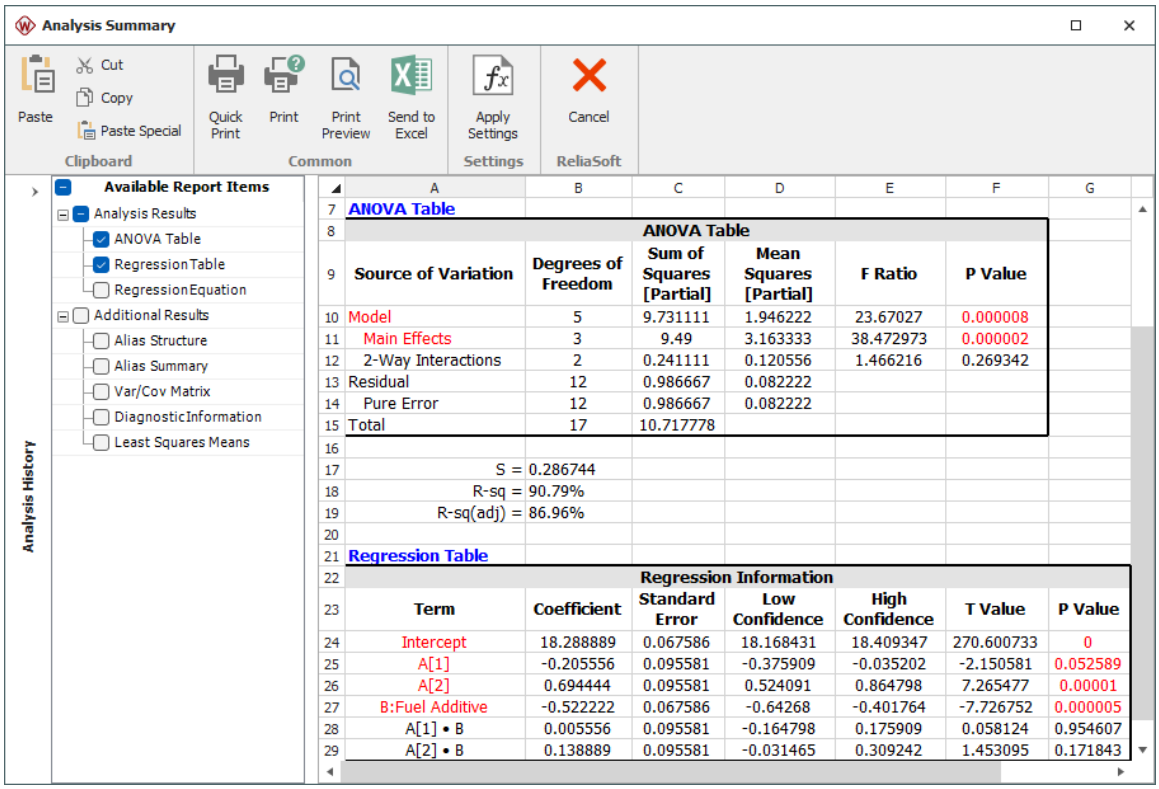
The test statistic for factor B is:

$$
\begin{aligned}
(f_0)_B &= \frac{MS_B}{MS_E} \\
&= \frac{4.9089/1}{0.9867/12} \\
&= 59.7
\end{aligned}
$$

The p value corresponding to this statistic based on the F distribution with 2 degrees of freedom in the numerator and 12 degrees of freedom in the denominator is:

$$
\begin{aligned}
p\,value &= 1 - P(F \leq (f_0)_B) \\
&= 1 - 0.999995 \\
&= 0.000005
\end{aligned}
$$

Since p value < 0.1, $H_0 : \delta_j = 0$ is rejected and it is concluded that factor B (or fuel additive type) has a significant effect on the mileage. Therefore, it can be concluded that speed and fuel additive type affect the mileage of the vehicle significantly. The results are displayed in the ANOVA table of the following figure.



ANOVA Table

Source of Variation	Degrees of Freedom	Sum of Squares [Partial]	Mean Squares [Partial]	F Ratio	P Value
Model	5	9.731111	1.946222	23.67027	0.000008
Main Effects	3	9.49	3.163333	38.472973	0.000002
2-Way Interactions	2	0.241111	0.120556	1.466216	0.269342
Residual	12	0.986667	0.082222		
Pure Error	12	0.986667	0.082222		
Total	17	10.717778			

S = 0.286744
R-sq = 90.79%
R-sq(adj) = 86.96%

Regression Table

Regression Information

Term	Coefficient	Standard Error	Low Confidence	High Confidence	T Value	P Value
Intercept	18.288889	0.067586	18.168431	18.409347	270.600733	0
A[1]	-0.205556	0.095581	-0.375909	-0.035202	-2.150581	0.052589
A[2]	0.694444	0.095581	0.524091	0.864798	7.265477	0.00001
B:Fuel Additive	-0.522222	0.067586	-0.64268	-0.401764	-7.726752	0.000005
A[1] • B	0.005556	0.095581	-0.164798	0.175909	0.058124	0.954607
A[2] • B	0.138889	0.095581	-0.031465	0.309242	1.453095	0.171843

Calculation of Effect Coefficients

Results for the effect coefficients of the model of the regression version of the ANOVA model are displayed in the Regression Information table in the following figure. Calculations of the results in this table are discussed next. The effect coefficients can be calculated as follows:

$$\hat{\beta} = (X'X)^{-1}X'y = \begin{bmatrix} 18.2889 \\ -0.2056 \\ 0.6944 \\ -0.5222 \\ 0.0056 \\ 0.1389 \end{bmatrix}$$

Therefore, $\hat{\mu} = 18.2889$, $\hat{\tau}_1 = -0.2056$, $\hat{\tau}_2 = 0.6944$ etc. As mentioned previously, these coefficients are displayed as Intercept, A[1] and A[2] respectively depending on the name of the factor used in the experimental design. The standard error for each of these estimates is obtained using the diagonal elements of the variance-covariance matrix C.

$$
\begin{aligned}
C =& \hat{\sigma}^2 (X'X)^{-1} \\
=& MS_E \cdot (X'X)^{-1} \\
=& \begin{bmatrix}
0.0046 & 0 & 0 & 0 & 0 & 0 \\
0 & 0.0091 & -0.0046 & 0 & 0 & 0 \\
0 & -0.0046 & 0.0091 & 0 & 0 & 0 \\
0 & 0 & 0 & 0.0046 & 0 & 0 \\
0 & 0 & 0 & 0 & 0.0091 & -0.0046 \\
0 & 0 & 0 & 0 & -0.0046 & 0.0091
\end{bmatrix}
\end{aligned}
$$

For example, the standard error for $\hat{\tau}_1$ is:

$$
\begin{aligned}
se(\hat{\tau}_1) =& \sqrt{C_{22}} \\
=& \sqrt{0.0091} \\
=& 0.0956
\end{aligned}
$$

Then the t statistic for $\hat{\tau}_1$ can be obtained as:

$$
\begin{aligned}
t_0 =& \frac{\hat{\tau}_1}{se(\hat{\tau}_1)} \\
=& \frac{-0.2056}{0.0956} \\
=& -2.1506
\end{aligned}
$$

The p value corresponding to this statistic is:

$$
\begin{aligned}
p\,value &= 2 \cdot (1 - P(T \leq |t_0|) \\
&= 2 \cdot 0.0263 \\
&= 0.0526
\end{aligned}
$$

Confidence intervals on τ_1 can also be calculated. The 90% limits on τ_1 are:

$$
\begin{aligned}
=& \hat{\tau}_1 \pm t_{\alpha/2, n-(k+1)} \sqrt{C_{22}} \\
=& \tau_1 \pm t_{0.05,12} \sqrt{C_{22}} \\
=& -0.2056 \pm 0.1704
\end{aligned}
$$

Thus, the 90% limits on τ_1 are -0.3760 and -0.0352 respectively. Results for other coefficients are obtained in a similar manner.

Least Squares Means

The estimated mean response corresponding to the ith level of any factor is obtained using the adjusted estimated mean which is also called the least squares mean. For example, the mean

response corresponding to the first level of factor A is $\mu+\tau_1$. An estimate of this is $\hat{\mu}+\hat{\tau}_1$ or ($18.2889+(-0.2056)=18.0833$). Similarly, the estimated response at the third level of factor A is $\hat{\mu}+\hat{\tau}_3$ or $\hat{\mu}+(-\hat{\tau}_1-\hat{\tau}_2)$ or ($18.2889+(0.2056-0.6944)=17.8001$).

Residual Analysis

As in the case of single factor experiments, plots of residuals can also be used to check for model adequacy in factorial experiments. Box-Cox transformations are also available in Weibull++ DOE folios for factorial experiments.

Randomization and Blocking in DOE

IN THIS CHAPTER

Randomization **146**

Blocking **147**
Example 147
Expression of the ANOVA Model as y = XB + ε 149
Calculation of the Sum of Squares for the Model 151
Calculation of the Extra Sum of Squares for the Factors 152
Calculation of the Test Statistics 153

Randomization

The aspect of recording observations in an experiment in a random order is referred to as *randomization*. Specifically, randomization is the process of assigning the various levels of the investigated factors to the experimental units in a random fashion. An experiment is said to be *completely randomized* if the probability of an experimental unit to be subjected to any level of a factor is equal for all the experimental units. The importance of randomization can be illustrated using an example. Consider an experiment where the effect of the speed of a lathe machine on the surface finish of a product is being investigated. In order to save time, the experimenter records surface finish values by running the lathe machine continuously and recording observations in the order of increasing speeds. The analysis of the experiment data shows that an increase in lathe speeds causes a decrease in the quality of surface finish. However the results of the experiment are disputed by the lathe operator who claims that he has been able to obtain better surface finish quality in the products by operating the lathe machine at higher speeds. It is later found that the faulty results were caused because of overheating of the tool used in the machine. Since the lathe was run continuously in the order of increased speeds the observations were recorded in the order of increased tool temperatures. This problem could have been avoided if the experimenter had randomized the experiment and taken reading at the various lathe speeds in a random fashion. This would require the experimenter to stop and restart the machine at every observation, thereby keeping the temperature of the tool within a reasonable range. Randomization would have ensured that the effect of heating of the machine tool is not included in the experiment.

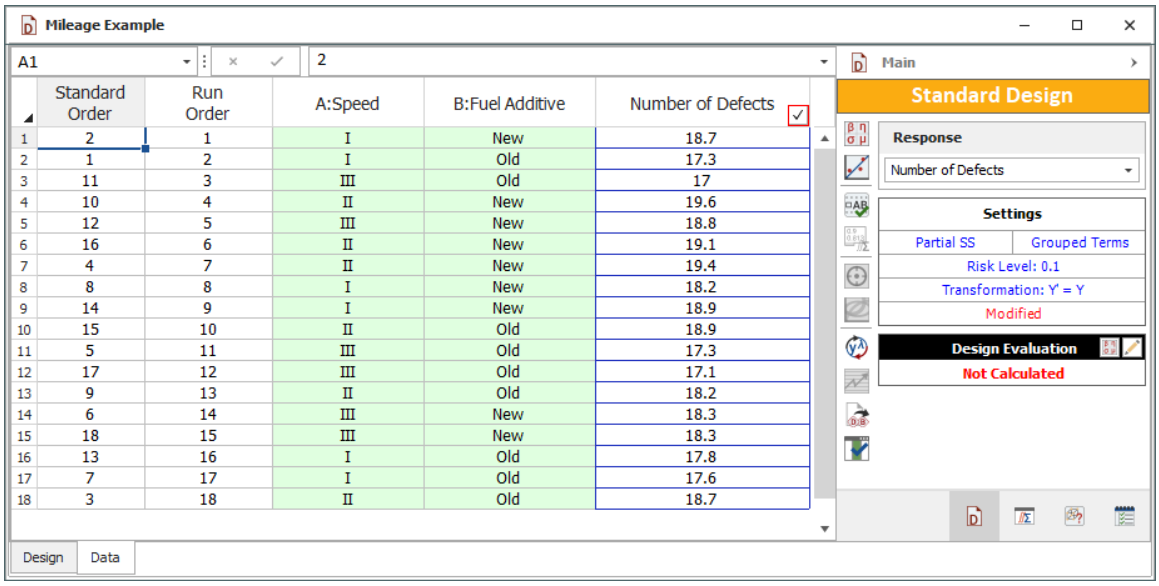
Blocking

Many times a factorial experiment requires so many runs that not all of them can be completed under homogeneous conditions. This may lead to inclusion of the effects of *nuisance factors* into the investigation. Nuisance factors are factors that have an effect on the response but are not of primary interest to the investigator. For example, two replicates of a two factor factorial experiment require eight runs. If four runs require the duration of one day to be completed, then the total experiment will require two days to be completed. The difference in the conditions on the two days may introduce effects on the response that are not the result of the two factors being investigated. Therefore, the day is a nuisance factor for this experiment. Nuisance factors can be accounted for using *blocking*. In blocking, experimental runs are separated based on levels of the nuisance factor. For the case of the two factor factorial experiment (where the day is a nuisance factor), separation can be made into two groups or *blocks*: runs that are carried out on the first day belong to block 1, and runs that are carried out on the second day belong to block 2. Thus, within each block conditions are the same with respect to the nuisance factor. As a result, each block investigates the effects of the factors of interest, while the difference in the blocks measures the effect of the nuisance factor. For the example of the two factor factorial experiment, a possible assignment of runs to the blocks could be as follows: one replicate of the experiment is assigned to block 1 and the second replicate is assigned to block 2 (now each block contains all possible treatment combinations). Within each block, runs are subjected to randomization (i.e., randomization is now restricted to the runs within a block). Such a design, where each block contains one complete replicate and the treatments within a block are subjected to randomization, is called *randomized complete block design*.

In summary, blocking should always be used to account for the effects of nuisance factors if it is not possible to hold the nuisance factor at a constant level through all of the experimental runs. Randomization should be used within each block to counter the effects of any unknown variability that may still be present.

Example

Consider the example discussed in General Full Factorial Design where the mileage of a sports utility vehicle was investigated for the effects of speed and fuel additive type. Now assume that the three replicates for this experiment were carried out on three different vehicles. To ensure that the variation from one vehicle to another does not have an effect on the analysis, each vehicle is considered as one block. See the experiment design in the following figure.



	Standard Order	Run Order	A:Speed	B:Fuel Additive	Number of Defects
1	2	1	I	New	18.7
2	1	2	I	Old	17.3
3	11	3	III	Old	17
4	10	4	II	New	19.6
5	12	5	III	New	18.8
6	16	6	II	New	19.1
7	4	7	II	New	19.4
8	8	8	I	New	18.2
9	14	9	I	New	18.9
10	15	10	II	Old	18.9
11	5	11	III	Old	17.3
12	17	12	III	Old	17.1
13	9	13	II	Old	18.2
14	6	14	III	New	18.3
15	18	15	III	New	18.3
16	13	16	I	Old	17.8
17	7	17	I	Old	17.6
18	3	18	II	Old	18.7

For the purpose of the analysis, the block is considered as a main effect except that it is assumed that interactions between the block and the other main effects do not exist. Therefore, there is one block main effect (having three levels - block 1, block 2 and block 3), two main effects (speed -having three levels; and fuel additive type - having two levels) and one interaction effect (speed-fuel additive interaction) for this experiment. Let ζ_i represent the block effects. The hypothesis test on the block main effect checks if there is a significant variation from one vehicle to the other. The statements for the hypothesis test are:

$$H_0: \quad \zeta_1 = \zeta_2 = \zeta_3 = 0 \text{ (no main effect of block)}$$
$$H_1: \quad \zeta_i \neq 0 \text{ for at least one } i$$

The test statistic for this test is:

$$F_0 = \frac{MS_{Block}}{MS_E}$$

where MS_{Block} represents the mean square for the block main effect and MS_E is the error mean square. The hypothesis statements and test statistics to test the significance of factors A (speed), B (fuel additive) and the interaction AB (speed-fuel additive interaction) can be obtained as explained in the example. The ANOVA model for this example can be written as:

$$Y_{ijk} = \mu + \zeta_i + \tau_j + \delta_k + (\tau\delta)_{jk} + \epsilon_{ijk}$$

where:

- μ represents the overall mean effect
- ζ_i is the effect of the $_i$th level of the block ($i = 1, 2, 3$)

- τ_j is the effect of the jth level of factor A $(j = 1, 2, 3)$
- δ_k is the effect of the kth level of factor B $(k = 1, 2)$
- $(\tau\delta)_{jk}$ represents the interaction effect between A and B
- and ϵ_{ijk} represents the random error terms (which are assumed to be normally distributed with a mean of zero and variance of σ^2)

In order to calculate the test statistics, it is convenient to express the ANOVA model of the equation given above in the form $y = X\beta + \epsilon$. This can be done as explained next.

Expression of the ANOVA Model as $y = XB + \varepsilon$

Since the effects ζ_i, τ_j, δ_k, and $(\tau\delta)_{jk}$ are defined as deviations from the overall mean, the following constraints exist.
Constraints on ζ_i are:

$$\sum_{i=1}^{3} \zeta_i = 0$$
$$\text{or } \zeta_1 + \zeta_2 + \zeta_3 = 0$$

Therefore, only two of the ζ_i effects are independent. Assuming that ζ_1 and ζ_2 are independent, $\zeta_3 = -(\zeta_1 + \zeta_2)$. (The null hypothesis to test the significance of the blocks can be rewritten using only the independent effects as $H_0 : \zeta_1 = \zeta_2 = 0$.) In DOE folios, the independent block effects, ζ_1 and ζ_2, are displayed as Block[1] and Block[2], respectively.

Constraints on τ_j are:

$$\sum_{j=1}^{3} \tau_j = 0$$
$$\text{or } \tau_1 + \tau_2 + \tau_3 = 0$$

Therefore, only two of the τ_j effects are independent. Assuming that τ_1 and τ_2 are independent, $\tau_3 = -(\tau_1 + \tau_2)$. The independent effects, τ_1 and τ_2, are displayed as A[1] and A[2], respectively. Constraints on δ_k are:

$$\sum_{k=1}^{2} \delta_k = 0$$
$$\text{or } \delta_1 + \delta_2 = 0$$

Therefore, only one of the δ_k effects is independent. Assuming that δ_1 is independent, $\delta_2 = -\delta_1$. The independent effect, δ_1, is displayed as B:B. Constraints on $(\tau\delta)_{jk}$ are:

$$\sum_{j=1}^{3}(\tau\delta)_{jk} = 0$$

$$\text{and } \sum_{k=1}^{2}(\tau\delta)_{jk} = 0$$

$$\text{or } (\tau\delta)_{11} + (\tau\delta)_{21} + (\tau\delta)_{31} = 0$$

$$(\tau\delta)_{12} + (\tau\delta)_{22} + (\tau\delta)_{32} = 0$$

$$\text{and } (\tau\delta)_{11} + (\tau\delta)_{12} = 0$$

$$(\tau\delta)_{21} + (\tau\delta)_{22} = 0$$

$$(\tau\delta)_{31} + (\tau\delta)_{32} = 0$$

The last five equations given above represent four constraints as only four of the five equations are independent. Therefore, only two out of the six $(\tau\delta)_{jk}$ effects are independent. Assuming that $(\tau\delta)_{11}$ and $(\tau\delta)_{21}$ are independent, we can express the other four effects in terms of these effects. The independent effects, $(\tau\delta)_{11}$ and $(\tau\delta)_{21}$, are displayed as A[1]B and A[2]B, respectively.

The regression version of the ANOVA model can be obtained using indicator variables. Since the block has three levels, two indicator variables, x_1 and x_2, are required, which need to be coded as shown next:

$$\text{Block 1}: \quad x_1 = 1,\ x_2 = 0$$

$$\text{Block 2}: \quad x_1 = 0,\ x_2 = 1$$

$$\text{Block 3}: \quad x_1 = -1,\ x_2 = -1$$

Factor A has three levels and two indicator variables, x_3 and x_4, are required:

$$\text{Treatment Effect } \tau_1: \quad x_3 = 1,\ x_4 = 0$$

$$\text{Treatment Effect } \tau_2: \quad x_3 = 0,\ x_4 = 1$$

$$\text{Treatment Effect } \tau_3: \quad x_3 = -1,\ x_4 = -1$$

Factor B has two levels and can be represented using one indicator variable, x_5, as follows:

$$\text{Treatment Effect } \delta_1: \quad x_5 = 1$$

$$\text{Treatment Effect } \delta_2: \quad x_5 = -1$$

The AB interaction will be represented by x_3x_5 and x_4x_5. The regression version of the ANOVA model can finally be obtained as:

$$Y = \mu + \zeta_1 \cdot x_1 + \zeta_2 \cdot x_2 + \tau_1 \cdot x_3 + \tau_2 \cdot x_4 + \delta_1 \cdot x_5 + (\tau\delta)_{11} \cdot x_3x_5 + (\tau\delta)_{21} \cdot x_4x_5 + \epsilon$$

In matrix notation this model can be expressed as:

$$y = X\beta + \epsilon$$

or:

$$\begin{bmatrix} 17.3 \\ 18.9 \\ 17.1 \\ 18.7 \\ 19.1 \\ 18.8 \\ 17.8 \\ 18.2 \\ . \\ . \\ 18.3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & -1 & -1 & 1 & -1 & -1 \\ 1 & 1 & 0 & 1 & 0 & -1 & -1 & 0 \\ 1 & 1 & 0 & 0 & 1 & -1 & 0 & -1 \\ 1 & 1 & 0 & -1 & -1 & -1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \zeta_1 \\ \zeta_2 \\ \tau_1 \\ \tau_2 \\ \delta_1 \\ (\tau\delta)_{11} \\ (\tau\delta)_{21} \end{bmatrix} + \begin{bmatrix} \epsilon_{111} \\ \epsilon_{121} \\ \epsilon_{131} \\ \epsilon_{112} \\ \epsilon_{122} \\ \epsilon_{132} \\ \epsilon_{211} \\ \epsilon_{221} \\ . \\ . \\ \epsilon_{332} \end{bmatrix}$$

Knowing y, X and β, the sum of squares for the ANOVA model and the extra sum of squares for each of the factors can be calculated. These are used to calculate the mean squares that are used to obtain the test statistics.

Calculation of the Sum of Squares for the Model

The model sum of squares, SS_{TR}, for the ANOVA model of this example can be obtained as:

$$\begin{aligned} SS_{TR} = & \quad y'[H - (\frac{1}{n_a \cdot n_b \cdot m})J]y \\ = & \quad y'[H - (\frac{1}{18})J]y \\ = & \quad 9.9256 \end{aligned}$$

Since seven effect terms (ζ_1, ζ_2, τ_1, τ_2, δ_1, $(\tau\delta)_{11}$ and $(\tau\delta)_{21}$) are used in the model the number of degrees of freedom associated with SS_{TR} is seven ($dof(SS_{TR}) = 7$).

The total sum of squares can be calculated as:

$$\begin{aligned} SS_T = & \quad y'[I - (\frac{1}{n_a \cdot n_b \cdot m})J]y \\ = & \quad y'[H - (\frac{1}{18})J]y \\ = & \quad 10.7178 \end{aligned}$$

Since there are 18 observed response values, the number of degrees of freedom associated with the total sum of squares is 17 ($dof(SS_T) = 17$). The error sum of squares can now be obtained:

$$\begin{aligned} SS_E =& SS_T - SS_{TR} \\ =& 10.7178 - 9.9256 \\ =& 0.7922 \end{aligned}$$

The number of degrees of freedom associated with the error sum of squares is:

$$\begin{aligned} dof(SS_E) &= dof(SS_T) - dof(SS_{TR}) \\ &= 17 - 7 \\ &= 10 \end{aligned}$$

Since there are no true replicates of the treatments (as can be seen from the design of the previous figure, where all of the treatments are seen to be run just once), all of the error sum of squares is the sum of squares due to lack of fit. The lack of fit arises because the model used is not a full model since it is assumed that there are no interactions between blocks and other effects.

Calculation of the Extra Sum of Squares for the Factors

The sequential sum of squares for the blocks can be calculated as:

$$\begin{aligned} SS_{Block} &= SS_{TR}(\mu, \zeta_1, \zeta_2) - SS_{TR}(\mu) \\ &= y'[H_{\mu,\zeta_1,\zeta_2} - (\frac{1}{18})J]y - 0 \end{aligned}$$

where J is the matrix of ones, H_{μ,ζ_1,ζ_2} is the hat matrix, which is calculated using $H_{\mu,\zeta_1,\zeta_2} = X_{\mu,\zeta_1,\zeta_2}(X'_{\mu,\zeta_1,\zeta_2}X_{\mu,\zeta_1,\zeta_2})^{-1}X'_{\mu,\zeta_1,\zeta_2}$, and X_{μ,ζ_1,ζ_2} is the matrix containing only the first three columns of the X matrix. Thus

$$\begin{aligned} SS_{Block} &= y'[H_{\mu,\zeta_1,\zeta_2} - (\frac{1}{18})J]y - 0 \\ &= 0.1944 - 0 \\ &= 0.1944 \end{aligned}$$

Since there are two independent block effects, and ζ_2 , the number of degrees of freedom associated with SS_{Blocks} is two ($dof(SS_{Blocks}) = 2$).

Similarly, the sequential sum of squares for factor A can be calculated as:

$$\begin{aligned} SS_A &= SS_{TR}(\mu, \zeta_1, \zeta_2, \tau_1, \tau_2) - SS_{TR}(\mu, \zeta_1, \zeta_2) \\ &= y'[H_{\mu,\zeta_1,\zeta_2,\tau_1,\tau_2} - (\frac{1}{18})J]y - y'[H_{\mu,\zeta_1,\zeta_2} - (\frac{1}{18})J]y \\ &= 4.7756 - 0.1944 \\ &= 4.5812 \end{aligned}$$

Sequential sum of squares for the other effects are obtained as $SS_B = 4.9089$ and $SS_{AB} = 0.2411$.

Calculation of the Test Statistics

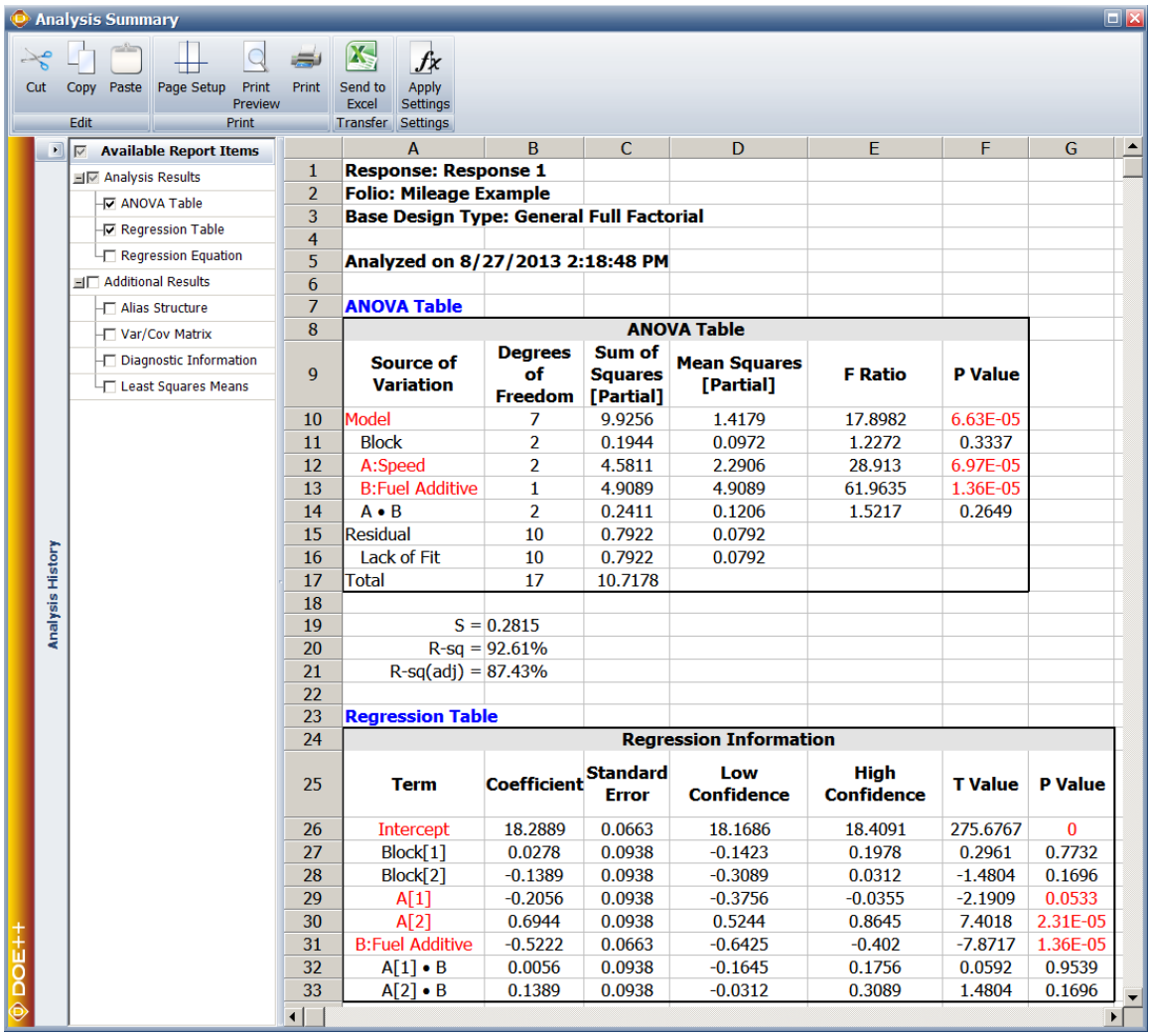
Knowing the sum of squares, the test statistics for each of the factors can be calculated. For example, the test statistic for the main effect of the blocks is:

$$\begin{aligned}(f_0)_{Block} &= \frac{MS_{Block}}{MS_E} \\ &= \frac{SS_{Block}/dof(SS_{Blocks})}{SS_E/dof(SS_E)} \\ &= \frac{0.1944/2}{0.7922/10} \\ &= 1.227\end{aligned}$$

The p value corresponding to this statistic based on the F distribution with 2 degrees of freedom in the numerator and 10 degrees of freedom in the denominator is:

$$\begin{aligned}p\ value &= 1 - P(F \leq (f_0)_{Block}) \\ &= 1 - 0.6663 \\ &= 0.3337\end{aligned}$$

Assuming that the desired significance level is 0.1, since p value > 0.1, we fail to reject $H_0 : \zeta_i = 0$ and conclude that there is no significant variation in the mileage from one vehicle to the other. Statistics to test the significance of other factors can be calculated in a similar manner. The complete analysis results obtained from the DOE folio for this experiment are presented in the following figure.



Response: Response 1
Folio: Mileage Example
Base Design Type: General Full Factorial

Analyzed on 8/27/2013 2:18:48 PM

ANOVA Table

Source of Variation	Degrees of Freedom	Sum of Squares [Partial]	Mean Squares [Partial]	F Ratio	P Value
Model	7	9.9256	1.4179	17.8982	6.63E-05
Block	2	0.1944	0.0972	1.2272	0.3337
A:Speed	2	4.5811	2.2906	28.913	6.97E-05
B:Fuel Additive	1	4.9089	4.9089	61.9635	1.36E-05
A • B	2	0.2411	0.1206	1.5217	0.2649
Residual	10	0.7922	0.0792		
Lack of Fit	10	0.7922	0.0792		
Total	17	10.7178			

S = 0.2815
R-sq = 92.61%
R-sq(adj) = 87.43%

Regression Table

Regression Information

Term	Coefficient	Standard Error	Low Confidence	High Confidence	T Value	P Value
Intercept	18.2889	0.0663	18.1686	18.4091	275.6767	0
Block[1]	0.0278	0.0938	-0.1423	0.1978	0.2961	0.7732
Block[2]	-0.1389	0.0938	-0.3089	0.0312	-1.4804	0.1696
A[1]	-0.2056	0.0938	-0.3756	-0.0355	-2.1909	0.0533
A[2]	0.6944	0.0938	0.5244	0.8645	7.4018	2.31E-05
B:Fuel Additive	-0.5222	0.0663	-0.6425	-0.402	-7.8717	1.36E-05
A[1] • B	0.0056	0.0938	-0.1645	0.1756	0.0592	0.9539
A[2] • B	0.1389	0.0938	-0.0312	0.3089	1.4804	0.1696

Two Level Factorial Experiments

IN THIS CHAPTER

2^k Designs **156**
The 2^2 Design 157
The 2^3 Design 159

Analysis of 2^k Designs **160**
Notation 160
Special Features 162

Replicated and Repeated Runs **172**

Unreplicated 2^k Designs **173**
Pooling Higher Order Interactions 173
Normal Probability Plot of Effects 173
Center Point Replicates 177
Using Center Point Replicates to Test Curvature 180

Blocking in 2^k Designs **183**
Unreplicated 2^k Designs in 2^p Blocks 188

Variability Analysis **191**

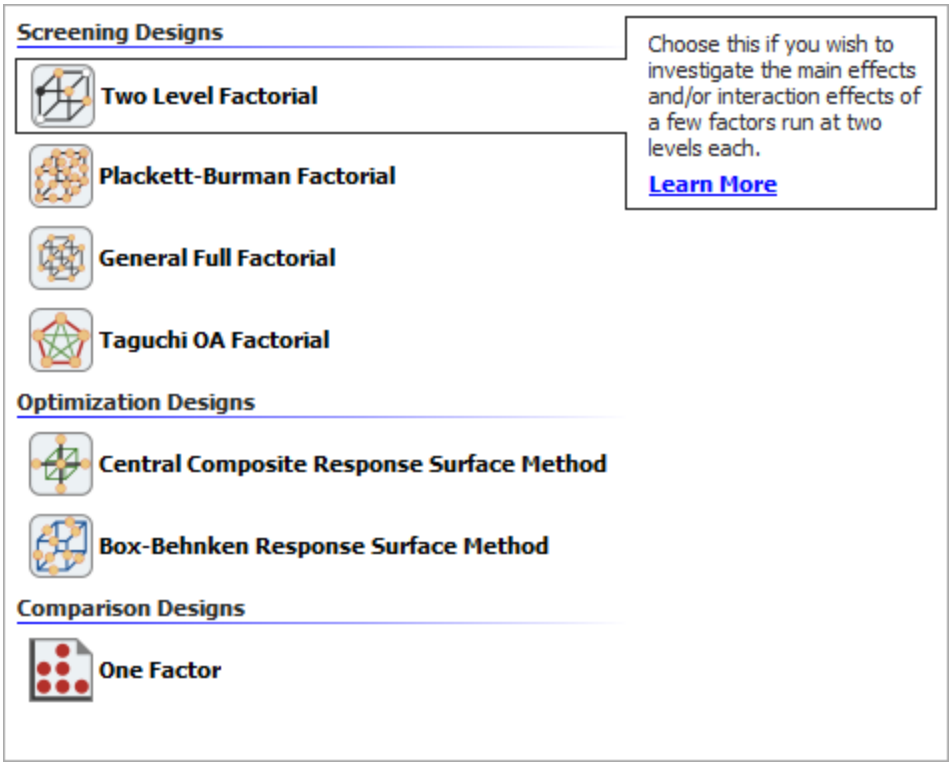
Two Level Fractional Factorial Designs **195**
Two Level Fractional Factorial Designs 195
Half-fraction Designs 195
Quarter and Smaller Fraction Designs 199
Design Resolution 201
Projection 207
Resolution III Designs 208
Alias Matrix 212

Two level factorial experiments are factorial experiments in which each factor is investigated at only two levels. The early stages of experimentation usually involve the investigation of a large number of potential factors to discover the "vital few" factors. Two level factorial experiments

are used during these stages to quickly filter out unwanted effects so that attention can then be focused on the important ones.

2^*k* Designs

The factorial experiments, where all combination of the levels of the factors are run, are usually referred to as *full factorial experiments*. Full factorial two level experiments are also referred to as 2^k designs where k denotes the number of factors being investigated in the experiment. In Weibull++ DOE folios, these designs are referred to as 2 Level Factorial Designs as shown in the figure below.



A full factorial two level design with k factors requires 2^k runs for a single replicate. For example, a two level experiment with three factors will require $2 \times 2 \times 2 = 2^3 = 8$ runs. The choice of the two levels of factors used in two level experiments depends on the factor; some factors naturally have two levels. For example, if gender is a factor, then male and female are the two levels. For other factors, the limits of the range of interest are usually used. For example, if temperature is a factor that varies from 45^oC to 90^oC, then the two levels used in the 2^k design for this factor would be 45^oC and 90^oC.

The two levels of the factor in the 2^k design are usually represented as -1 (for the first level) and 1 (for the second level). Note that this representation is reversed from the coding used in General Full Factorial Designs for the indicator variables that represent two level factors in

ANOVA models. For ANOVA models, the first level of the factor was represented using a value of 1 for the indicator variable, while the second level was represented using a value of -1. For details on the notation used for two level experiments refer to Notation.

The 2^2 Design

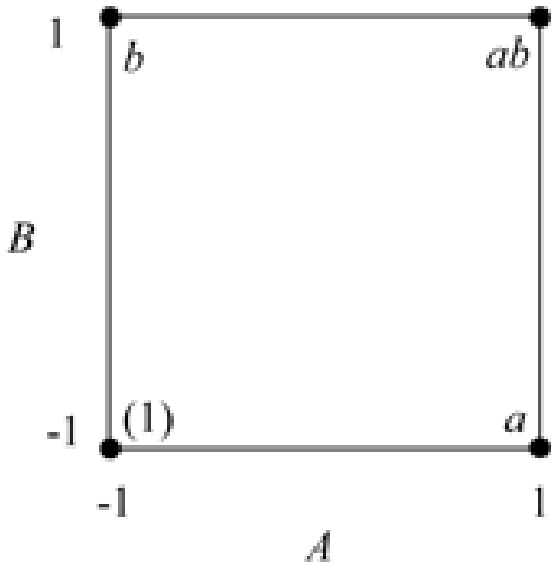
The simplest of the two level factorial experiments is the 2^2 design where two factors (say factor A and factor B) are investigated at two levels. A single replicate of this design will require four runs ($2^2 = 2 \times 2 = 4$) The effects investigated by this design are the two main effects, A and B, and the interaction effect AB. The treatments for this design are shown in figure (a) below. In figure (a), letters are used to represent the treatments. The presence of a letter indicates the high level of the corresponding factor and the absence indicates the low level. For example, (1) represents the treatment combination where all factors involved are at the low level or the level represented by -1 ; a represents the treatment combination where factor A is at the high level or the level of 1, while the remaining factors (in this case, factor B) are at the low level or the level of -1. Similarly, b represents the treatment combination where factor B is at the high level or the level of 1, while factor A is at the low level and ab represents the treatment combination where factors A and B are at the high level or the level of the 1. Figure (b) below shows the design matrix for the 2^2 design. It can be noted that the sum of the terms resulting from the product of any two columns of the design matrix is zero. As a result the 2^2 design is an *orthogonal design*. In fact, all 2^k designs are orthogonal designs. This property of the 2^k designs offers a great advantage in the analysis because of the simplifications that result from orthogonality. These simplifications are explained later on in this chapter. The 2^2 design can also be represented geometrically using a square with the four treatment combinations lying at the four corners, as shown in figure (c) below.

Treatment Name	Factors	
	A	*B*
(1)	-1	-1
a	1	-1
b	-1	1
ab	1	1

(a)

I	*A*	*B*	*AB*
1	-1	-1	1
1	1	-1	-1
1	-1	1	-1
1	1	1	1

(b)



(c)

The 2^3 Design

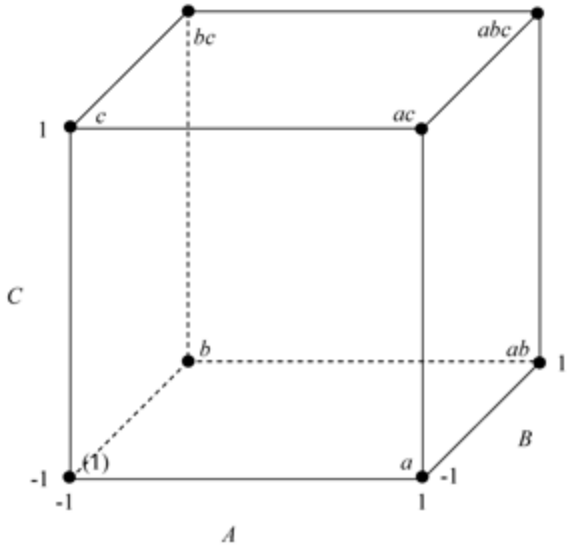
The 2^3 design is a two level factorial experiment design with three factors (say factors A, B and C). This design tests three ($k = 3$) main effects, A, B and C ; three ($\binom{k}{2} = \binom{3}{2} = 3$) two factor interaction effects, AB, BC, AC ; and one ($\binom{k}{3} = \binom{3}{3} = 1$) three factor interaction effect, ABC. The design requires eight runs per replicate. The eight treatment combinations corresponding to these runs are (1), a, b, ab, c, ac, bc and abc. Note that the treatment combinations are written in such an order that factors are introduced one by one with each new factor being combined with the preceding terms. This order of writing the treatments is called the *standard order* or *Yates' order*. The 2^3 design is shown in figure (a) below. The design matrix for the 2^3 design is shown in figure (b). The design matrix can be constructed by following the standard order for the treatment combinations to obtain the columns for the main effects and then multiplying the main effects columns to obtain the interaction columns.

Treatment Name	Factors		
	A	*B*	*C*
(1)	-1	-1	-1
a	1	-1	-1
b	-1	1	-1
ab	1	1	-1
c	-1	-1	1
ac	1	-1	1
bc	-1	1	1
abc	1	1	1

(a)

I	*A*	*B*	*AB*	*C*	*AC*	*BC*	*ABC*
1	-1	-1	1	-1	1	1	-1
1	1	-1	-1	-1	-1	1	1
1	-1	1	-1	-1	1	-1	1
1	1	1	1	-1	-1	-1	-1
1	-1	-1	1	1	-1	-1	1
1	1	-1	-1	1	1	-1	-1
1	-1	1	-1	1	-1	1	-1
1	1	1	1	1	1	1	1

(b)



The 2^3 design can also be represented geometrically using a cube with the eight treatment combinations lying at the eight corners as shown in the figure above.

Analysis of 2^*k* Designs

The 2^k designs are a special category of the factorial experiments where all the factors are at two levels. The fact that these designs contain factors at only two levels and are orthogonal greatly simplifies their analysis even when the number of factors is large. The use of 2^k designs in investigating a large number of factors calls for a revision of the notation used previously for the ANOVA models. The case for revised notation is made stronger by the fact that the ANOVA and multiple linear regression models are identical for 2^k designs because all factors are only at two levels. Therefore, the notation of the regression models is applied to the ANOVA models for these designs, as explained next.

Notation

Based on the notation used in General Full Factorial Designs, the ANOVA model for a two level factorial experiment with three factors would be as follows:

$$Y = \mu + \tau_1 \cdot x_1 + \delta_1 \cdot x_2 + (\tau\delta)_{11} \cdot x_1x_2 + \gamma_1 \cdot x_3 + (\tau\gamma)_{11} \cdot x_1x_3 + (\delta\gamma)_{11} \cdot x_2x_3 + (\tau\delta\gamma)_{111} \cdot x_1x_2x_3 + \epsilon$$

where:

- μ represents the overall mean
- τ_1 represents the independent effect of the first factor (factor A) out of the two effects τ_1 and τ_2
- δ_1 represents the independent effect of the second factor (factor B) out of the two effects δ_1 and δ_2
- $(\tau\delta)_{11}$ represents the independent effect of the interaction AB out of the other interaction effects
- γ_1 represents the effect of the third factor (factor C) out of the two effects γ_1 and γ_2
- $(\tau\gamma)_{11}$ represents the effect of the interaction AC out of the other interaction effects
- $(\delta\gamma)_{11}$ represents the effect of the interaction BC out of the other interaction effects
- $(\tau\delta\gamma)_{111}$ represents the effect of the interaction ABC out of the other interaction effects

and ϵ is the random error term.

The notation for a linear regression model having three predictor variables with interactions is:

$$\begin{aligned} Y = \quad & \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_{12} \cdot x_1 x_2 + \beta_3 \cdot x_3 \\ & + \beta_{13} \cdot x_1 x_3 + \beta_{23} \cdot x_2 x_3 + \beta_{123} \cdot x_1 x_2 x_3 + \epsilon \end{aligned}$$

The notation for the regression model is much more convenient, especially for the case when a large number of higher order interactions are present. In two level experiments, the ANOVA model requires only one indicator variable to represent each factor for both qualitative and quantitative factors. Therefore, the notation for the multiple linear regression model can be applied to the ANOVA model of the experiment that has all the factors at two levels. For example, for the experiment of the ANOVA model given above, β_0 can represent the overall mean instead of μ, and β_1 can represent the independent effect, τ_1, of factor A. Other main effects can be represented in a similar manner. The notation for the interaction effects is much more simplified (e.g., β_{123} can be used to represent the three factor interaction effect, $(\tau\beta\gamma)_{111}$).

As mentioned earlier, it is important to note that the coding for the indicator variables for the ANOVA models of two level factorial experiments is reversed from the coding followed in General Full Factorial Designs. Here -1 represents the first level of the factor while 1 represents the second level. This is because for a two level factor a single variable is needed to represent the factor for both qualitative and quantitative factors. For quantitative factors, using -1 for the first level (which is the low level) and 1 for the second level (which is the high level) keeps the coding consistent with the numerical value of the factors. The change in coding between the two

coding schemes does not affect the analysis except that signs of the estimated effect coefficients will be reversed (i.e., numerical values of $\hat{\tau}_1$, obtained based on the coding of General Full Factorial Designs, and $\hat{\beta}_1$, obtained based on the new coding, will be the same but their signs would be opposite).

Factor A Coding (two level factor)

Previous Coding	Coding for 2^k Designs
$Effect\ \tau_1 \ : \ x_1 = 1$	$Effect\ \tau_1\ (\text{or}\ -\beta_1) \ : \ x_1 = -1$
$Effect\ \tau_2 \ : \ x_1 = -1$	$Effect\ \tau_2\ (\text{or}\ \beta_1) \ : \ x_1 = 1$

In summary, the ANOVA model for the experiments with all factors at two levels is different from the ANOVA models for other experiments in terms of the notation in the following two ways:

- The notation of the regression models is used for the effect coefficients.
- The coding of the indicator variables is reversed.

Special Features

Consider the design matrix, X, for the 2^3 design discussed above. The $(X'X)^{-1}$ matrix is:

$$(X'X)^{-1} = \begin{bmatrix} 0.125 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.125 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.125 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.125 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.125 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.125 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.125 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.125 \end{bmatrix}$$

Notice that, due to the orthogonal design of the X matrix, the $(X'X)^{-1}$ has been simplified to a diagonal matrix which can be written as:

$$(X'X)^{-1} = 0.125 \cdot I = \frac{1}{8} \cdot I = \frac{1}{2^3} \cdot I$$

where I represents the identity matrix of the same order as the design matrix, X. Since there are eight observations per replicate of the 2^3 design, the $(X'X)^{-1}$ matrix for m replicates of this design can be written as:

$$(X'X)^{-1} = \frac{1}{(2^3 \cdot m)} \cdot I$$

The $(X'X)^{-1}$ matrix for any 2^k design can now be written as:

$$(X'X)^{-1} = \frac{1}{(2^k \cdot m)} \cdot I$$

Then the variance-covariance matrix for the 2^k design is:

$$C = \hat{\sigma}^2 \cdot (X'X)^{-1} = MS_E \cdot (X'X)^{-1} = \frac{MS_E}{(2^k \cdot m)} \cdot I$$

Note that the variance-covariance matrix for the 2^k design is also a diagonal matrix. Therefore, the estimated effect coefficients (β_1 , β_2 , β_{12}, etc.) for these designs are uncorrelated. This implies that the terms in the 2^k design (main effects, interactions) are independent of each other. Consequently, the extra sum of squares for each of the terms in these designs is independent of the sequence of terms in the model, and also independent of the presence of other terms in the model. As a result the sequential and partial sum of squares for the terms are identical for these designs and will always add up to the model sum of squares. Multicollinearity is also not an issue for these designs.

It can also be noted from the equation given above, that in addition to the C matrix being diagonal, all diagonal elements of the C matrix are identical. This means that the variance (or its square root, the standard error) of all estimated effect coefficients are the same. The standard error, $se(\hat{\beta}_j)$, for all the coefficients is:

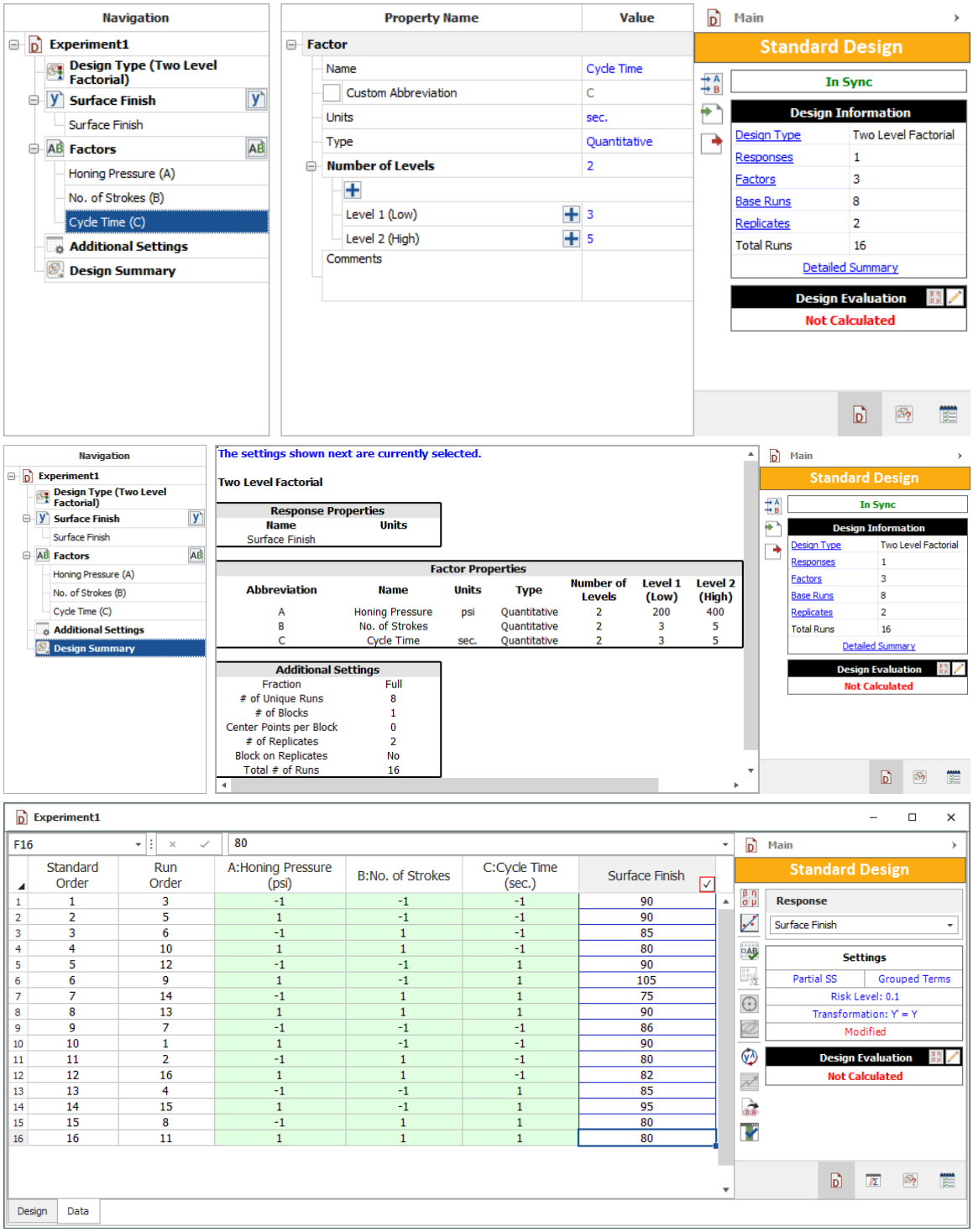
$$se(\hat{\beta}_j) = \sqrt{C_{jj}} = \sqrt{\frac{MS_E}{(2^k \cdot m)}} \quad for\ all\ j$$

This property is used to construct the normal probability plot of effects in 2^k designs and identify significant effects using graphical techniques. For details on the normal probability plot of effects in a Weibull++ DOE folio, refer to Normal Probability Plot of Effects.

Example

To illustrate the analysis of a full factorial 2^k design, consider a three factor experiment to investigate the effect of honing pressure, number of strokes and cycle time on the surface finish of automobile brake drums. Each of these factors is investigated at two levels. The honing pressure is investigated at levels of 200 psi and 400 psi, the number of strokes used is 3 and 5 and the two levels of the cycle time are 3 and 5 seconds. The design for this experiment is set up in a Weibull++ DOE folio as shown in the first two following figures. It is decided to run two

replicates for this experiment. The surface finish data collected from each run (using randomization) and the complete design is shown in the third following figure. The analysis of the experiment data is explained next.



The applicable model using the notation for 2^k designs is:

$$Y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_{12} \cdot x_1 x_2 + \beta_3 \cdot x_3 + \beta_{13} \cdot x_1 x_3 + \beta_{23} \cdot x_2 x_3 + \beta_{123} \cdot x_1 x_2 x_3 + \epsilon$$

where the indicator variable, x_1, represents factor A (honing pressure), $x_1 = -1$ represents the low level of 200 psi and $x_1 = 1$ represents the high level of 400 psi. Similarly, x_2 and x_3 represent factors B (number of strokes) and C (cycle time), respectively. β_0 is the overall mean, while β_1, β_2 and β_3 are the effect coefficients for the main effects of factors A, B and C, respectively. β_{12}, β_{13} and β_{23} are the effect coefficients for the AB, AC and BC interactions, while β_{123} represents the ABC interaction.

If the subscripts for the run (i ; $i =$ 1 to 8) and replicates (j ; $j =$ 1,2) are included, then the model can be written as:

$$Y_{ij} = \beta_0 + \beta_1 \cdot x_{ij1} + \beta_2 \cdot x_{ij2} + \beta_{12} \cdot x_{ij1} x_{ij2} + \beta_3 \cdot x_{ij3} + \beta_{13} \cdot x_{ij1} x_{ij3} + \beta_{23} \cdot x_{ij2} x_{ij3} + \beta_{123} \cdot x_{ij1} x_{ij2} x_{ij3} + \epsilon_{ij}$$

To investigate how the given factors affect the response, the following hypothesis tests need to be carried:

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

This test investigates the main effect of factor A (honing pressure). The statistic for this test is:

$$(F_0)_A = \frac{MS_A}{MS_E}$$

where MS_A is the mean square for factor A and MS_E is the error mean square. Hypotheses for the other main effects, B and C, can be written in a similar manner.

$$H_0 : \beta_{12} = 0$$
$$H_1 : \beta_{12} \neq 0$$

This test investigates the two factor interaction AB. The statistic for this test is:

$$(F_0)_{AB} = \frac{MS_{AB}}{MS_E}$$

where MS_{AB} is the mean square for the interaction AB and MS_E is the error mean square. Hypotheses for the other two factor interactions, AC and BC, can be written in a similar manner.

$$H_0 : \beta_{123} = 0$$
$$H_1 : \beta_{123} \neq 0$$

This test investigates the three factor interaction ABC. The statistic for this test is:

$$(F_0)_{ABC} = \frac{MS_{ABC}}{MS_E}$$

where MS_{ABC} is the mean square for the interaction ABC and MS_E is the error mean square. To calculate the test statistics, it is convenient to express the ANOVA model in the form $y = X\beta + \epsilon$.

Expression of the ANOVA Model as $y = X\beta + \epsilon$

In matrix notation, the ANOVA model can be expressed as:

$$y = X\beta + \epsilon$$

where:

$$y = \begin{bmatrix} Y_{11} \\ Y_{21} \\ . \\ Y_{81} \\ Y_{12} \\ . \\ Y_{82} \end{bmatrix} = \begin{bmatrix} 90 \\ 90 \\ . \\ 90 \\ 86 \\ . \\ 80 \end{bmatrix} \quad X = \begin{bmatrix} 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ . & . & . & . & . & . & . & . \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\ . & . & . & . & . & . & . & . \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_{12} \\ \beta_3 \\ \beta_{13} \\ \beta_{23} \\ \beta_{123} \end{bmatrix} \quad \epsilon = \begin{bmatrix} \epsilon_{11} \\ \epsilon_{21} \\ . \\ \epsilon_{81} \\ \epsilon_{12} \\ . \\ . \\ \epsilon_{82} \end{bmatrix}$$

Calculation of the Extra Sum of Squares for the Factors

Knowing the matrices y, X and β, the extra sum of squares for the factors can be calculated. These are used to calculate the mean squares that are used to obtain the test statistics. Since the experiment design is orthogonal, the partial and sequential extra sum of squares are identical. The extra sum of squares for each effect can be calculated as shown next. As an example, the extra sum of squares for the main effect of factor A is:

$$\begin{aligned} SS_A &= Model\ Sum\ of\ Squares - Sum\ of\ Squares\ of\ model\ excluding\ the\ main\ effect\ of\ A \\ &= y'[H - (1/16)J]y - y'[H_{\sim A} - (1/16)J]y \end{aligned}$$

where H is the hat matrix and J is the matrix of ones. The matrix $H_{\sim A}$ can be calculated using $H_{\sim A} = X_{\sim A}(X'_{\sim A}X_{\sim A})^{-1}X'_{\sim A}$ where $X_{\sim A}$ is the design matrix, X, excluding the second column that represents the main effect of factor A. Thus, the sum of squares for the main effect of factor A is:

$$\begin{aligned} SS_A &= y'[H - (1/16)J]y - y'[H_{\sim A} - (1/16)J]y \\ &= 654.4375 - 549.375 \\ &= 105.0625 \end{aligned}$$

Similarly, the extra sum of squares for the interaction effect AB is:

$$\begin{aligned} SS_{AB} &= y'[H - (1/16)J]y - y'[H_{\sim AB} - (1/16)J]y \\ &= 654.4375 - 636.375 \\ &= 18.0625 \end{aligned}$$

The extra sum of squares for other effects can be obtained in a similar manner.

Calculation of the Test Statistics

Knowing the extra sum of squares, the test statistic for the effects can be calculated. For example, the test statistic for the interaction AB is:

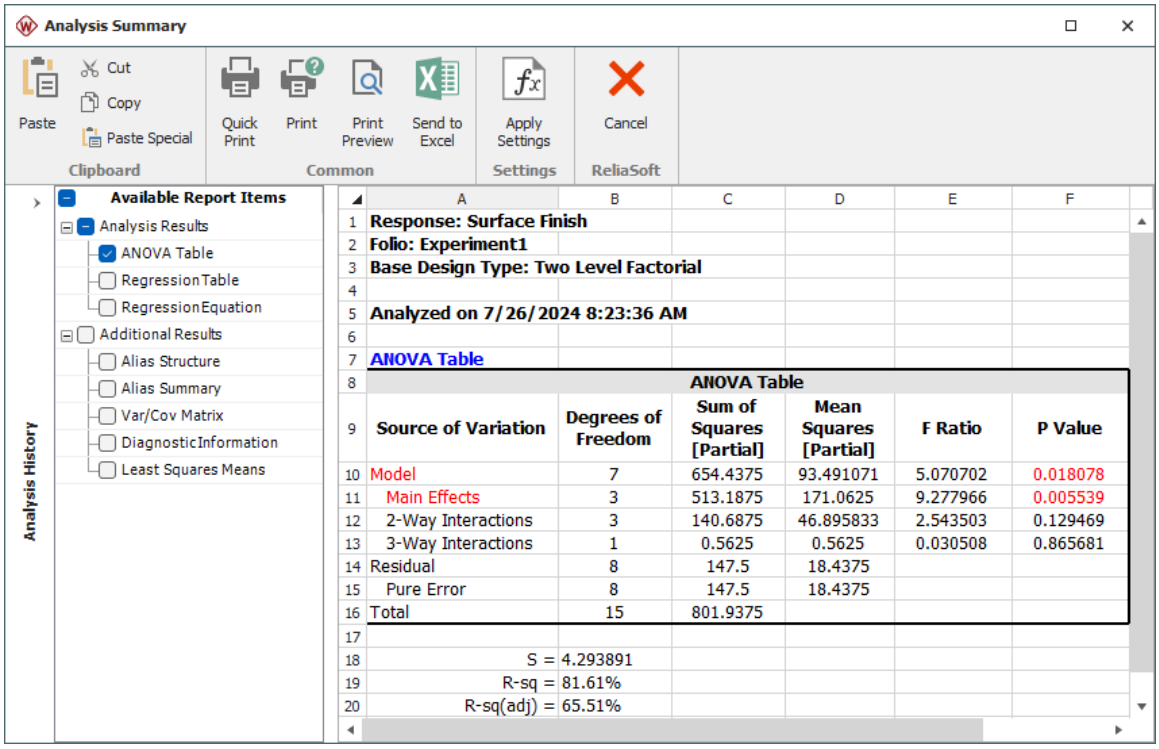
$$\begin{aligned} (f_0)_{AB} &= \frac{MS_{AB}}{MS_E} \\ &= \frac{SS_{AB}/dof(SS_{AB})}{SS_E/dof(SS_E)} \\ &= \frac{18.0625/1}{147.5/8} \\ &= 0.9797 \end{aligned}$$

where MS_{AB} is the mean square for the AB interaction and MS_E is the error mean square. The p value corresponding to the statistic, $(f_0)_{AB} = 0.9797$, based on the F distribution with one degree of freedom in the numerator and eight degrees of freedom in the denominator is:

$$\begin{aligned} p\ value &= 1 - P(F \leq (f_0)_{AB}) \\ &= 1 - 0.6487 \\ &= 0.3513 \end{aligned}$$

Assuming that the desired significance is 0.1, since p value > 0.1, it can be concluded that the interaction between honing pressure and number of strokes does not affect the surface finish of the brake drums. Tests for other effects can be carried out in a similar manner. The results are

shown in the ANOVA Table in the following figure. The values S, R-sq and R-sq(adj) in the figure indicate how well the model fits the data. The value of S represents the standard error of the model, R-sq represents the coefficient of multiple determination and R-sq(adj) represents the adjusted coefficient of multiple determination. For details on these values refer to Multiple Linear Regression Analysis.



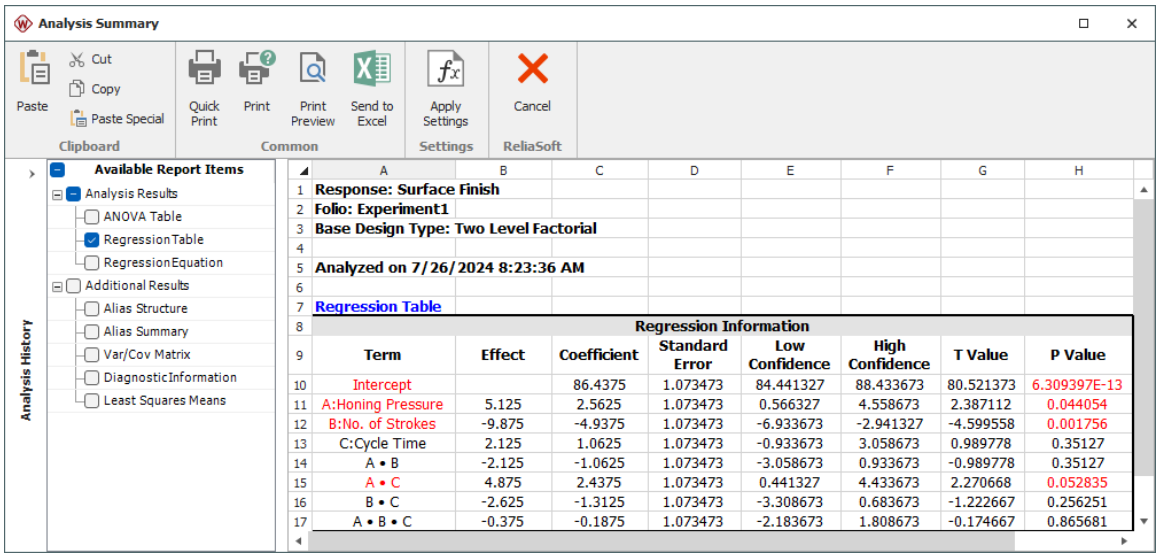
	A	B	C	D	E	F
1	**Response: Surface Finish**					
2	**Folio: Experiment1**					
3	**Base Design Type: Two Level Factorial**					
4						
5	**Analyzed on 7/26/2024 8:23:36 AM**					
6						
7	**ANOVA Table**					
8	**ANOVA Table**					
9	**Source of Variation**	**Degrees of Freedom**	**Sum of Squares [Partial]**	**Mean Squares [Partial]**	**F Ratio**	**P Value**
10	Model	7	654.4375	93.491071	5.070702	0.018078
11	Main Effects	3	513.1875	171.0625	9.277966	0.005539
12	2-Way Interactions	3	140.6875	46.895833	2.543503	0.129469
13	3-Way Interactions	1	0.5625	0.5625	0.030508	0.865681
14	Residual	8	147.5	18.4375		
15	Pure Error	8	147.5	18.4375		
16	Total	15	801.9375			
17						
18	S =	4.293891				
19	R-sq =	81.61%				
20	R-sq(adj) =	65.51%				

Calculation of Effect Coefficients

The estimate of effect coefficients can also be obtained:

$$\hat{\beta} = (X'X)^{-1}X'y$$

$$= \begin{bmatrix} 86.4375 \\ 2.5625 \\ -4.9375 \\ 1.0625 \\ -1.0625 \\ 2.4375 \\ -1.3125 \\ -0.1875 \end{bmatrix}$$



Regression Information

Term	Effect	Coefficient	Standard Error	Low Confidence	High Confidence	T Value	P Value
Intercept		86.4375	1.073473	84.441327	88.433673	80.521373	6.309397E-13
A:Honing Pressure	5.125	2.5625	1.073473	0.566327	4.558673	2.387112	0.044054
B:No. of Strokes	-9.875	-4.9375	1.073473	-6.933673	-2.941327	-4.599558	0.001756
C:Cycle Time	2.125	1.0625	1.073473	-0.933673	3.058673	0.989778	0.35127
A • B	-2.125	-1.0625	1.073473	-3.058673	0.933673	-0.989778	0.35127
A • C	4.875	2.4375	1.073473	0.441327	4.433673	2.270668	0.052835
B • C	-2.625	-1.3125	1.073473	-3.308673	0.683673	-1.222667	0.256251
A • B • C	-0.375	-0.1875	1.073473	-2.183673	1.808673	-0.174667	0.865681

The coefficients and related results are shown in the Regression Information table above. In the table, the Effect column displays the effects, which are simply twice the coefficients. The Standard Error column displays the standard error, $se(\hat{\beta}_j)$. The Low CI and High CI columns display the confidence interval on the coefficients. The interval shown is the 90% interval as the significance is chosen as 0.1. The T Value column displays the t statistic, t_0, corresponding to the coefficients. The P Value column displays the p value corresponding to the t statistic. (For details on how these results are calculated, refer to General Full Factorial Designs). Plots of residuals can also be obtained from the DOE folio to ensure that the assumptions related to the ANOVA model are not violated.

Model Equation

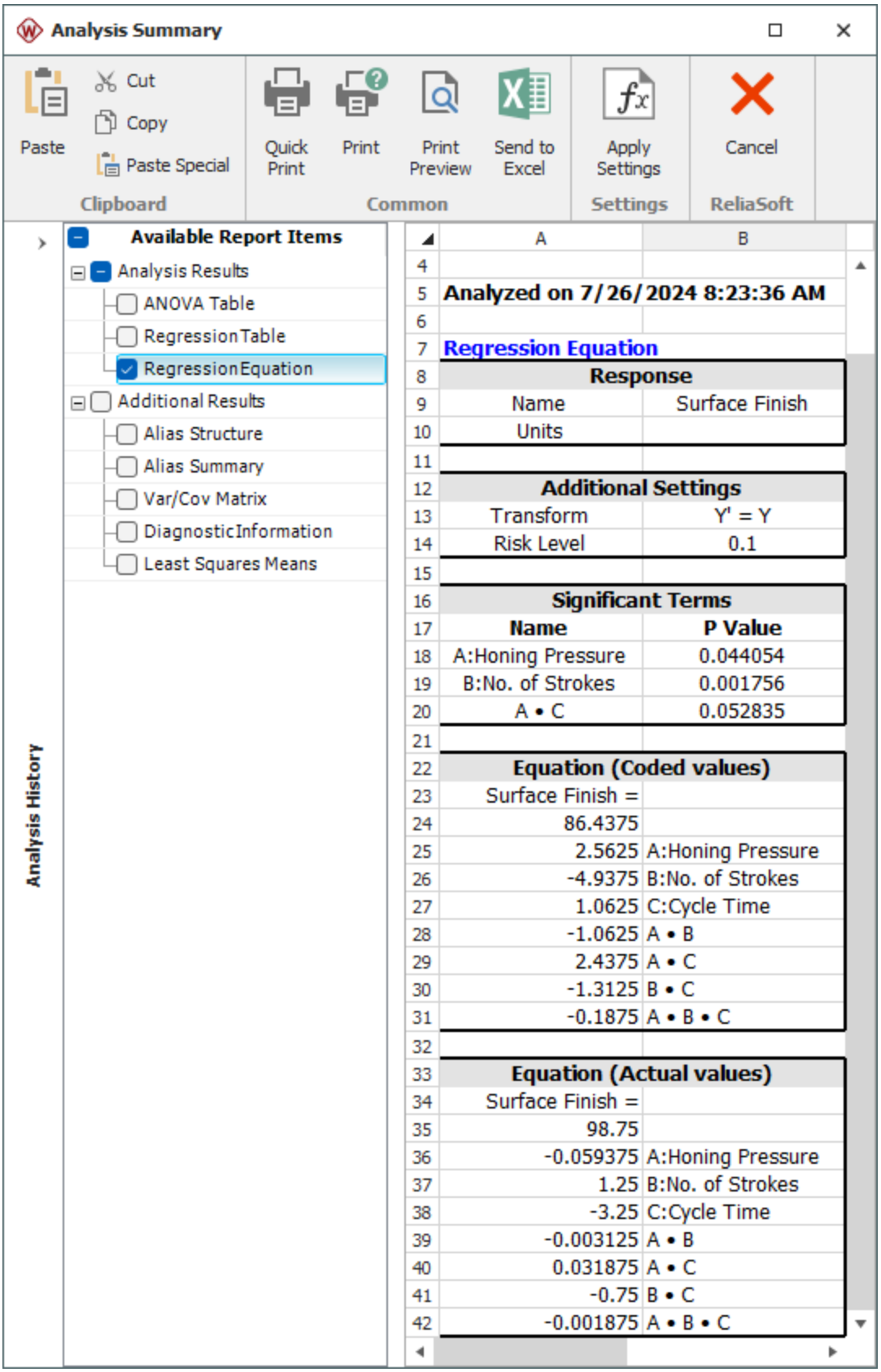
From the analysis results in the above figure within calculation of effect coefficients section, it is seen that effects A, B and AC are significant. In a DOE folio, the p values for the significant effects are displayed in red in the ANOVA Table for easy identification. Using the values of the estimated effect coefficients, the model for the present 2^3 design in terms of the coded values can be written as:

$$\begin{aligned}\hat{y} &= \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_{13} \cdot x_1 x_3 \\ &= 86.4375 + 2.5625x_1 - 4.9375x_2 + 2.4375x_1 x_3\end{aligned}$$

To make the model hierarchical, the main effect, C, needs to be included in the model (because the interaction AC is included in the model). The resulting model is:

$$\hat{y} = 86.4375 + 2.5625x_1 - 4.9375x_2 + 1.0625x_3 + 2.4375x_1 x_3$$

This equation can be viewed in a DOE folio, as shown in the following figure, using the Show Analysis Summary icon in the Control Panel. The equation shown in the figure will match the hierarchical model once the required terms are selected using the Select Effects icon.



	A	B
4		
5	**Analyzed on 7/26/2024 8:23:36 AM**	
6		
7	**Regression Equation**	
8	**Response**	
9	Name	Surface Finish
10	Units	
11		
12	**Additional Settings**	
13	Transform	Y' = Y
14	Risk Level	0.1
15		
16	**Significant Terms**	
17	**Name**	**P Value**
18	A:Honing Pressure	0.044054
19	B:No. of Strokes	0.001756
20	A • C	0.052835
21		
22	**Equation (Coded values)**	
23	Surface Finish =	
24	86.4375	
25	2.5625	A:Honing Pressure
26	-4.9375	B:No. of Strokes
27	1.0625	C:Cycle Time
28	-1.0625	A • B
29	2.4375	A • C
30	-1.3125	B • C
31	-0.1875	A • B • C
32		
33	**Equation (Actual values)**	
34	Surface Finish =	
35	98.75	
36	-0.059375	A:Honing Pressure
37	1.25	B:No. of Strokes
38	-3.25	C:Cycle Time
39	-0.003125	A • B
40	0.031875	A • C
41	-0.75	B • C
42	-0.001875	A • B • C

Replicated and Repeated Runs

In the case of replicated experiments, it is important to note the difference between replicated runs and repeated runs. Both repeated and replicated runs are multiple response readings taken at the same factor levels. However, repeated runs are response observations taken at the same time or in succession. Replicated runs are response observations recorded in a random order. Therefore, replicated runs include more variation than repeated runs. For example, a baker, who wants to investigate the effect of two factors on the quality of cakes, will have to bake four cakes to complete one replicate of a 2^2 design. Assume that the baker bakes eight cakes in all. If, for each of the four treatments of the 2^2 design, the baker selects one treatment at random and then bakes two cakes for this treatment at the same time then this is a case of two repeated runs. If, however, the baker bakes all the eight cakes randomly, then the eight cakes represent two sets of replicated runs. For repeated measurements, the average values of the response for each treatment should be entered into a DOE folio as shown in the following figure (a) when the two cakes for a particular treatment are baked together. For replicated measurements, when all the cakes are baked randomly, the data is entered as shown in the following figure (b).

Treatment Name	Factors		Cake Quality
	A	*B*	
(1)	-1	-1	7.5
a	1	-1	5.5
b	-1	1	6.5
ab	1	1	8.5

(a)

Treatment Name	Factors		Cake Quality
	A	*B*	
(1)	-1	-1	8
a	1	-1	5
b	-1	1	7
ab	1	1	9
(1)	-1	-1	7
a	1	-1	6
b	-1	1	6
ab	1	1	8

(b)

Unreplicated 2^*k* Designs

If a factorial experiment is run only for a single replicate then it is not possible to test hypotheses about the main effects and interactions as the error sum of squares cannot be obtained. This is because the number of observations in a single replicate equals the number of terms in the ANOVA model. Hence the model fits the data perfectly and no degrees of freedom are available to obtain the error sum of squares.

However, sometimes it is only possible to run a single replicate of the 2^k design because of constraints on resources and time. In the absence of the error sum of squares, hypothesis tests to identify significant factors cannot be conducted. A number of methods of analyzing information obtained from unreplicated 2^k designs are available. These include pooling higher order interactions, using the normal probability plot of effects or including center point replicates in the design.

Pooling Higher Order Interactions

One of the ways to deal with unreplicated 2^k designs is to use the sum of squares of some of the higher order interactions as the error sum of squares provided these higher order interactions can be assumed to be insignificant. By dropping some of the higher order interactions from the model, the degrees of freedom corresponding to these interactions can be used to estimate the error mean square. Once the error mean square is known, the test statistics to conduct hypothesis tests on the factors can be calculated.

Normal Probability Plot of Effects

Another way to use unreplicated 2^k designs to identify significant effects is to construct the normal probability plot of the effects. As mentioned in Special Features, the standard error for all effect coefficients in the 2^k designs is the same. Therefore, on a normal probability plot of effect coefficients, all non-significant effect coefficients (with $\beta = 0$) will fall along the straight line representative of the normal distribution, N($0, \sigma^2/(2^k \cdot m)$). Effect coefficients that show large deviations from this line will be significant since they do not come from this normal distribution. Similarly, since effects $= 2 \times$ effect coefficients, all non-significant effects will also follow a straight line on the normal probability plot of effects. For replicated designs, the Effects Probability plot of a DOE folio plots the normalized effect values (or the T Values) on the standard normal probability line, N(0,1). However, in the case of unreplicated 2^k designs, σ^2 remains unknown since MS_E cannot be obtained. Lenth's method is used in this case to estimate the variance of the effects. For details on Lenth's method, please refer to Montgomery (2001). The DOE folio then uses this variance value to plot effects along the N(0, Lenth's effect variance) line. The method is illustrated in the following example.

Example

Vinyl panels, used as instrument panels in a certain automobile, are seen to develop defects after a certain amount of time. To investigate the issue, it is decided to carry out a two level factorial experiment. Potential factors to be investigated in the experiment are vacuum rate (factor A), material temperature (factor B), element intensity (factor C) and pre-stretch (factor D). The two levels of the factors used in the experiment are as shown in below.

Factors	Levels	
	-1	1
A: Vacuum Rate	75%	100%
B: Material Temperature	340 F	360 F
C: Element Intensity	80%	90%
D: Pre-Stretch	no	yes

With a 2^4 design requiring 16 runs per replicate it is only feasible for the manufacturer to run a single replicate.

The experiment design and data, collected as percent defects, are shown in the following figure. Since the present experiment design contains only a single replicate, it is not possible to obtain an estimate of the error sum of squares, SS_E. It is decided to use the normal probability plot of effects to identify the significant effects. The effect values for each term are obtained as shown in the following figure.

A1 | -1

	A:Vacuum Rate	B:Material Temperature	C:Element Intensity	D:Pre-Stretch	Defects
1	-1	-1	-1	-1	5
2	1	-1	-1	-1	16
3	-1	1	-1	-1	10
4	1	1	-1	-1	15
5	-1	-1	1	-1	13
6	1	-1	1	-1	14
7	-1	1	1	-1	10
8	1	1	1	-1	13
9	-1	-1	-1	1	50
10	1	-1	-1	1	24
11	-1	1	-1	1	55
12	1	1	-1	1	36
13	-1	-1	1	1	51
14	1	-1	1	1	36
15	-1	1	1	1	56
16	1	1	1	1	22

Lenth's method uses these values to estimate the variance. As described in [Lenth, 1989], if all effects are arranged in ascending order, using their absolute values, then s_0 is defined as 1.5 times the median value:

$$\begin{aligned} s_0 &= 1.5 \cdot median(|effect|) \\ &= 1.5 \cdot 2 \\ &= 3 \end{aligned}$$

Using s_0, the "pseudo standard error" (PSE) is calculated as 1.5 times the median value of all effects that are less than 2.5 s_0 :

$$\begin{aligned} PSE &= 1.5 \cdot median(|effect| \quad : \quad |effect| < 2.5 s_0) \\ &= 1.5 \cdot 1.5 \\ &= 2.25 \end{aligned}$$

Using PSE as an estimate of the effect variance, the effect variance is 2.25. Knowing the effect variance, the normal probability plot of effects for the present unreplicated experiment can be constructed as shown in the following figure. The line on this plot is the line N(0, 2.25). The plot shows that the effects A, D and the interaction AD do not follow the distribution represented by this line. Therefore, these effects are significant.

The significant effects can also be identified by comparing individual effect values to the margin of error or the threshold value using the pareto chart (see the third following figure). If the required significance is 0.1, then:

$$margin\ of\ error = t_{\alpha/2,d} \cdot PSE$$

The t statistic, $t_{\alpha/2,d}$, is calculated at a significance of $\alpha/2$ (for the two-sided hypothesis) and degrees of freedom $d = (\text{ number of effects })/3$. Thus:

$$\begin{aligned} margin\ of\ error &= t_{0.05,5} \cdot PSE \\ &= 2.015 \cdot 2.25 \\ &= 4.534 \end{aligned}$$

The value of 4.534 is shown as the critical value line in the third following figure. All effects with absolute values greater than the margin of error can be considered to be significant. These effects are A, D and the interaction AD. Therefore, the vacuum rate, the pre-stretch and their interaction have a significant effect on the defects of the vinyl panels.

Analysis Summary

Paste | Cut | Copy | Paste Special | Quick Print | Print | Print Preview | Send to Excel | Apply Settings | Cancel

Clipboard | Common | Settings | ReliaSoft

Available Report Items

- Analysis Results
  - ANOVA Table
  - Regression Table
  - Regression Equation
- Additional Results
  - Alias Structure
  - Alias Summary
  - Var/Cov Matrix
  - Diagnostic Information
  - Least Squares Means

Analysis History

Response: Defects
Folio: Vinyl Panel Defect
Base Design Type: Two Level Factorial

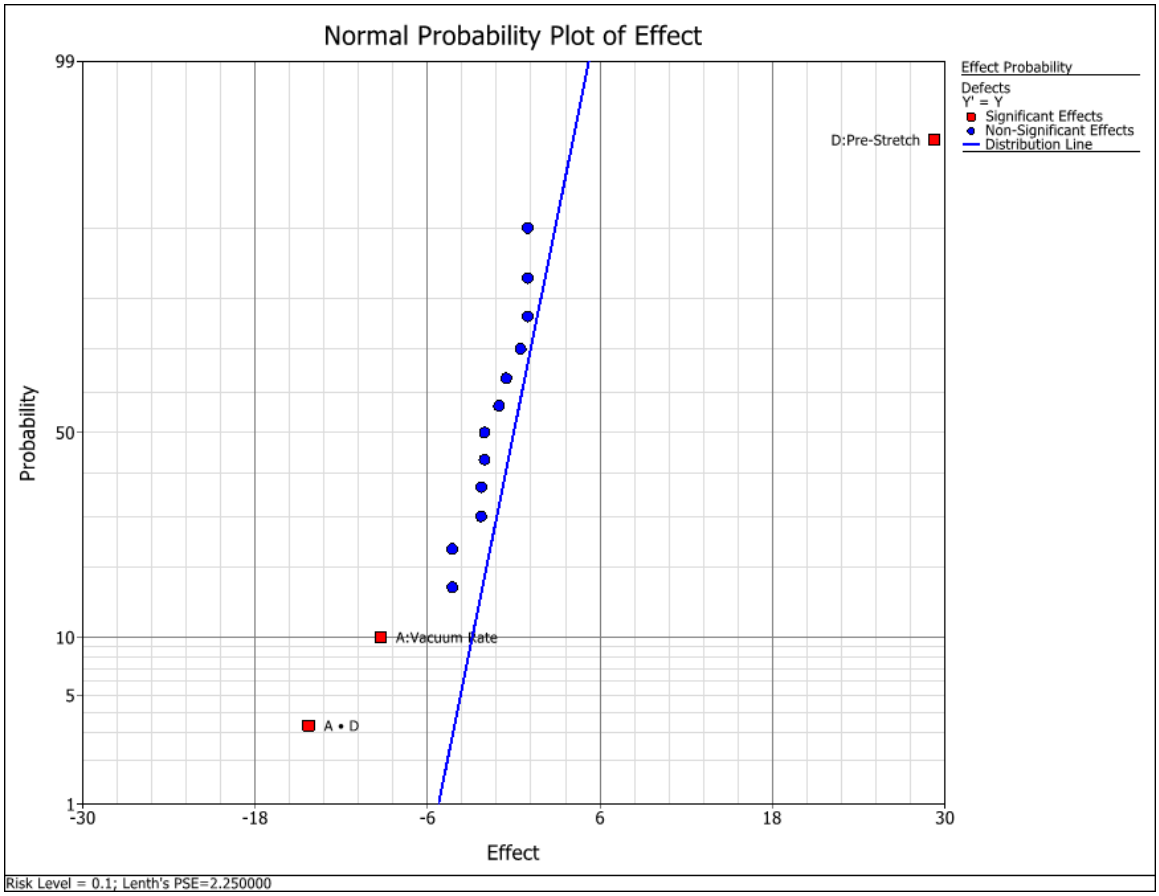
Analyzed on 6/18/2024 1:29:47 PM

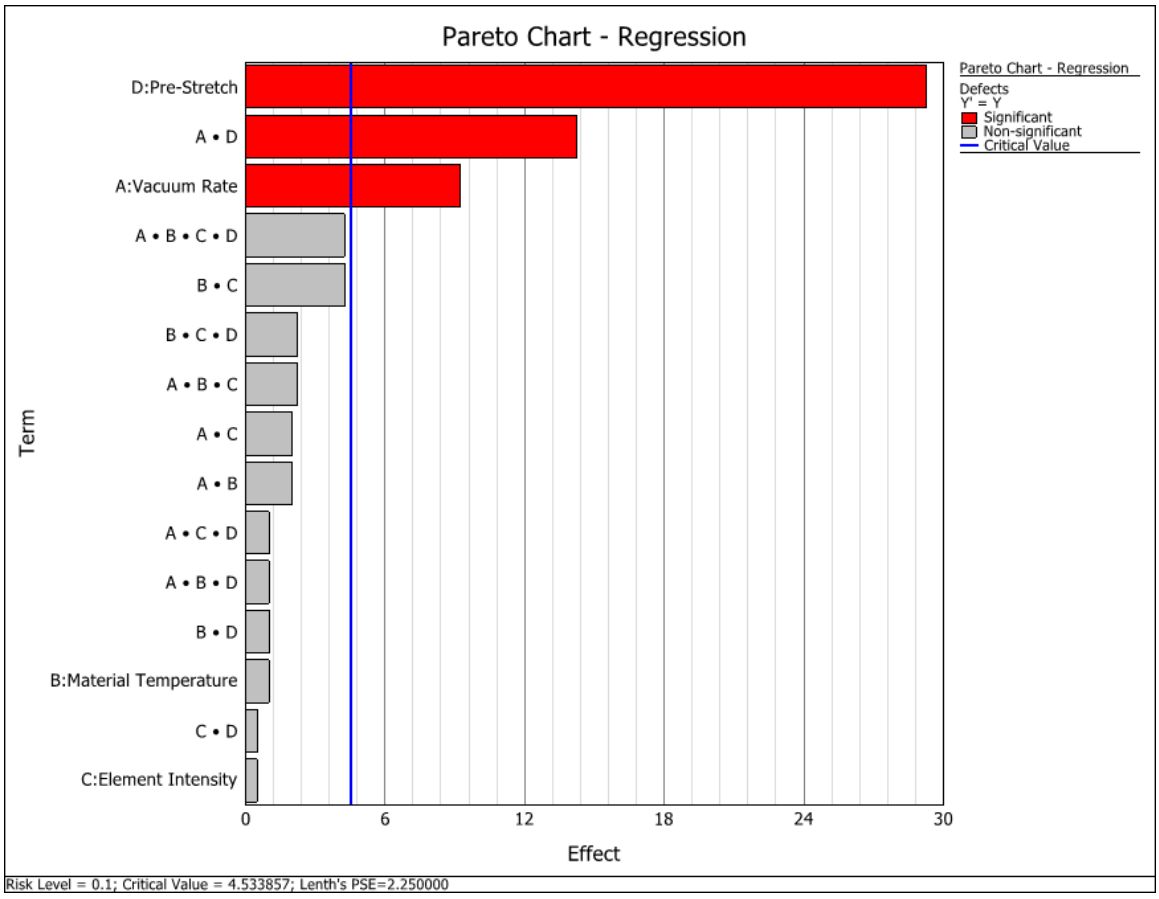
Regression Table

Regression Information

Term	Effect	Coefficient	Standard Error	Low Confidence	High Confidence	T Value	P Value
Intercept		26.625	-	-	-	-	-
A:Vacuum Rate*	-9.25	-4.625	-	-	-	-	-
B:Material Temperature	1	0.5	-	-	-	-	-
C:Element Intensity	0.5	0.25	-	-	-	-	-
D:Pre-Stretch*	29.25	14.625	-	-	-	-	-
A • B	-2	-1	-	-	-	-	-
A • C	-2	-1	-	-	-	-	-
A • D*	-14.25	-7.125	-	-	-	-	-
B • C	-4.25	-2.125	-	-	-	-	-
B • D	1	0.5	-	-	-	-	-
C • D	-0.5	-0.25	-	-	-	-	-
A • B • C	-2.25	-1.125	-	-	-	-	-
A • B • D	-1	-0.5	-	-	-	-	-
A • C • D	1	0.5	-	-	-	-	-
B • C • D	-2.25	-1.125	-	-	-	-	-
A • B • C • D	-4.25	-2.125	-	-	-	-	-

*: Significant terms according to Lenth's method





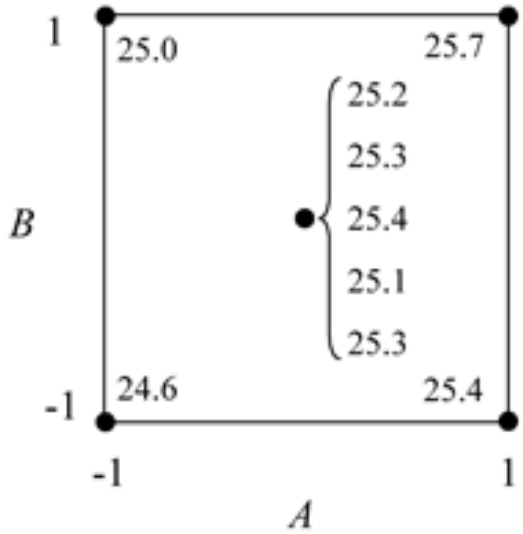
Center Point Replicates

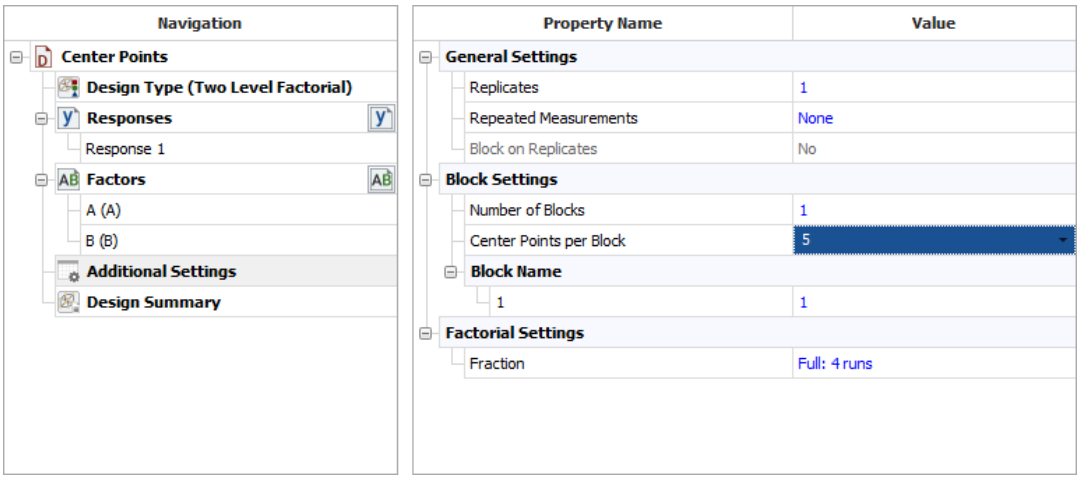
Another method of dealing with unreplicated 2^k designs that only have quantitative factors is to use replicated runs at the center point. The center point is the response corresponding to the treatment exactly midway between the two levels of all factors. Running multiple replicates at this point provides an estimate of pure error. Although running multiple replicates at any treatment level can provide an estimate of pure error, the other advantage of running center point replicates in the 2^k design is in checking for the presence of curvature. The test for curvature investigates whether the model between the response and the factors is linear and is discussed in Center Pt. Replicates to Test Curvature.

Example: Use Center Point to Get Pure Error

Consider a 2^2 experiment design to investigate the effect of two factors, A and B, on a certain response. The energy consumed when the treatments of the 2^2 design are run is considerably larger than the energy consumed for the center point run (because at the center point the factors are at their middle levels). Therefore, the analyst decides to run only a single replicate of the design and augment the design by five replicated runs at the center point as shown in the

following figure. The design properties for this experiment are shown in the second following figure. The complete experiment design is shown in the third following figure. The center points can be used in the identification of significant effects as shown next.





A1 | 1

	Standard Order	Run Order	A:A	B:B	Response 1
1	1	2	-1	-1	24.6
2	2	8	1	-1	25.4
3	3	5	-1	1	25
4	4	4	1	1	25.7
5	5	6	0	0	25.2
6	6	3	0	0	25.3
7	7	1	0	0	25.4
8	8	9	0	0	25.1
9	9	7	0	0	25.3

Since the present 2^2 design is unreplicated, there are no degrees of freedom available to calculate the error sum of squares. By augmenting this design with five center points, the response values at the center points, y_i^c, can be used to obtain an estimate of pure error, SS_{PE}. Let $\bar{y}^c$ represent the average response for the five replicates at the center. Then:

$$SS_{PE} = Sum\ of\ Squares\ for\ center\ points$$

$$\begin{aligned} SS_{PE} &= \sum_{i=1}^{5} (y_i^c - \bar{y}^c)^2 \\ &= (25.2 - 25.26)^2 + \ldots + (25.3 - 25.26)^2 \\ &= 0.052 \end{aligned}$$

Then the corresponding mean square is:

$$\begin{aligned} MS_{PE} &= \frac{SS_{PE}}{degrees\ of\ freedom} \\ &= \frac{0.052}{5-1} \\ &= 0.013 \end{aligned}$$

Alternatively, MS_{PE} can be directly obtained by calculating the variance of the response values at the center points:

$$\begin{aligned} MS_{PE} &= s^2 \\ &= \frac{\sum_{i=1}^{5} (y_i^c - \bar{y}^c)^2}{5-1} \end{aligned}$$

Once MS_{PE} is known, it can be used as the error mean square, MS_E, to carry out the test of significance for each effect. For example, to test the significance of the main effect of factor A, the sum of squares corresponding to this effect is obtained in the usual manner by considering only the four runs of the original 2^2 design.

$$\begin{aligned} SS_A &= y'[H - (1/4)J]y - y'[H_{-A} - (1/4)J]y \\ &= 0.5625 \end{aligned}$$

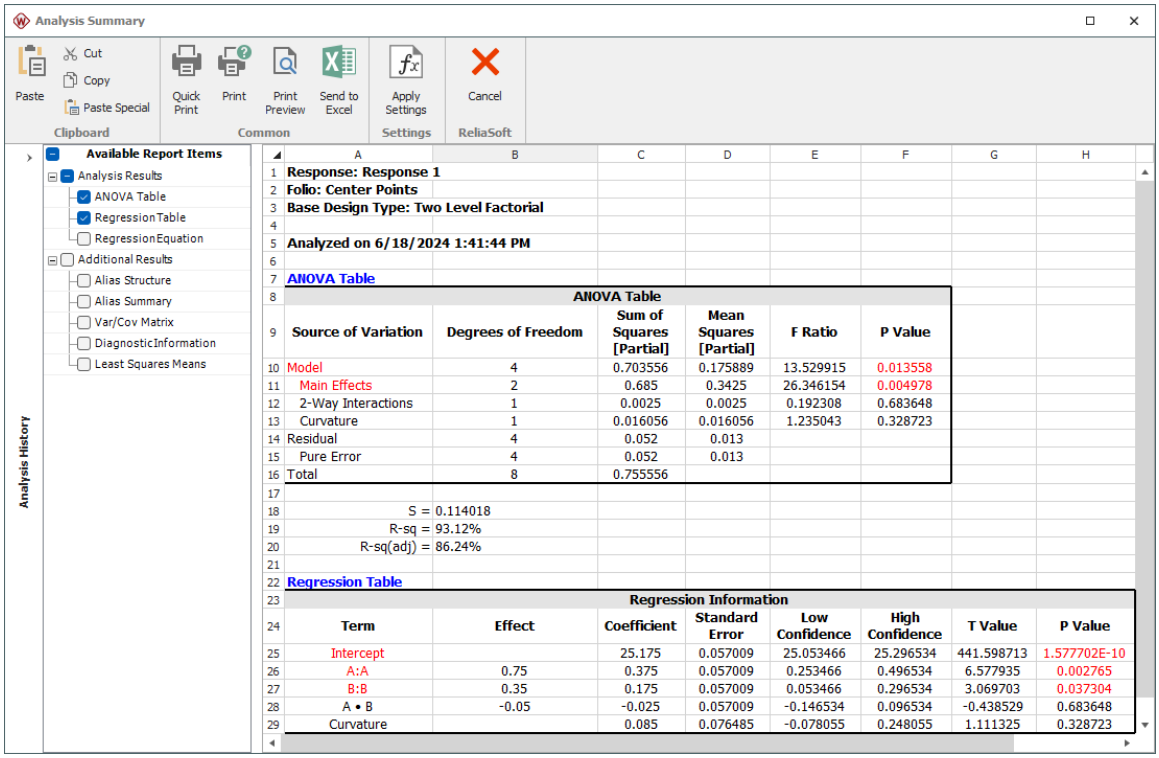
Then, the test statistic to test the significance of the main effect of factor A is:

$$\begin{aligned} (f_0)_A &= \frac{MS_A}{MS_E} \\ &= \frac{0.5625/1}{0.052/4} \\ &= 43.2692 \end{aligned}$$

The p value corresponding to the statistic, $(f_0)_A = 43.2692$, based on the F distribution with one degree of freedom in the numerator and eight degrees of freedom in the denominator is:

$$
\begin{aligned}
p\ value &= 1 - P(F \leq (f_0)_A) \\
&= 1 - 0.9972 \\
&= 0.0028
\end{aligned}
$$

Assuming that the desired significance is 0.1, since p value < 0.1, it can be concluded that the main effect of factor A significantly affects the response. This result is displayed in the ANOVA table as shown in the following figure. Test for the significance of other factors can be carried out in a similar manner.



Response: Response 1
Folio: Center Points
Base Design Type: Two Level Factorial

Analyzed on 6/18/2024 1:41:44 PM

ANOVA Table

ANOVA Table					
Source of Variation	Degrees of Freedom	Sum of Squares [Partial]	Mean Squares [Partial]	F Ratio	P Value
Model	4	0.703556	0.175889	13.529915	0.013558
Main Effects	2	0.685	0.3425	26.346154	0.004978
2-Way Interactions	1	0.0025	0.0025	0.192308	0.683648
Curvature	1	0.016056	0.016056	1.235043	0.328723
Residual	4	0.052	0.013		
Pure Error	4	0.052	0.013		
Total	8	0.755556			

S = 0.114018
R-sq = 93.12%
R-sq(adj) = 86.24%

Regression Table

Regression Information							
Term	Effect	Coefficient	Standard Error	Low Confidence	High Confidence	T Value	P Value
Intercept		25.175	0.057009	25.053466	25.296534	441.598713	1.577702E-10
A:A	0.75	0.375	0.057009	0.253466	0.496534	6.577935	0.002765
B:B	0.35	0.175	0.057009	0.053466	0.296534	3.069703	0.037304
A • B	-0.05	-0.025	0.057009	-0.146534	0.096534	-0.438529	0.683648
Curvature		0.085	0.076485	-0.078055	0.248055	1.111325	0.328723

Using Center Point Replicates to Test Curvature

Center point replicates can also be used to check for curvature in replicated or unreplicated 2^k designs. The test for curvature investigates whether the model between the response and the factors is linear. The way the DOE folio handles center point replicates is similar to its handling of blocks. The center point replicates are treated as an additional factor in the model. The factor is labeled as Curvature in the results of the DOE folio. If Curvature turns out to be a significant factor in the results, then this indicates the presence of curvature in the model.

Example: Use Center Point to Test Curvature

To illustrate the use of center point replicates in testing for curvature, consider again the data of the single replicate 2^2 experiment from a preceding figure(labeled "2^2 design augmented by five

center point runs"). Let x_1 be the indicator variable to indicate if the run is a center point:

$$\begin{aligned} x_1 &= 0 \quad \text{Center point run} \\ x_1 &= 1 \quad \text{Other run} \end{aligned}$$

If x_2 and x_3 are the indicator variables representing factors A and B, respectively, then the model for this experiment is:

$$Y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 + \beta_{23} \cdot x_2 x_3$$

To investigate the presence of curvature, the following hypotheses need to be tested:

$$\begin{aligned} H_0 : &\quad \beta_1 = 0 \text{ (Curvature is absent)} \\ H_1 : &\quad \beta_1 \neq 0 \end{aligned}$$

The test statistic to be used for this test is:

$$(F_0)_{curvature} = \frac{MS_{curvature}}{MS_E}$$

where $MS_{curvature}$ is the mean square for Curvature and MS_E is the error mean square.

Calculation of the Sum of Squares

The X matrix and y vector for this experiment are:

$$X = \begin{bmatrix} 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix} \quad y = \begin{bmatrix} 24.6 \\ 25.4 \\ 25.0 \\ 25.7 \\ 25.2 \\ 25.3 \\ 25.4 \\ 25.1 \\ 25.3 \end{bmatrix}$$

The sum of squares can now be calculated. For example, the error sum of squares is:

$$\begin{aligned} SS_E = &\quad y'[I - H]y \\ = &\quad 0.052 \end{aligned}$$

where I is the identity matrix and H is the hat matrix. It can be seen that this is equal to SS_{PE} (the sum of squares due to pure error) because of the replicates at the center point, as obtained in the example. The number of degrees of freedom associated with SS_E , $dof(SS_E)$ is four. The extra sum of squares corresponding to the center point replicates (or Curvature) is:

$$\begin{aligned} SS_{Curvature} = & \quad Model\ Sum\ of\ Squares - \\ & Sum\ of\ Squares\ of\ model\ excluding\ the\ center\ point \\ = & \quad y'[H - (1/9)J]y - y'[H_{\text{-}Curvature} - (1/9)J]y \end{aligned}$$

where H is the hat matrix and J is the matrix of ones. The matrix $H_{\text{-}Curvature}$ can be calculated using $H_{\text{-}Curvature} = X_{\text{-}Curv}(X'_{\text{-}Curv}X_{\text{-}Curv})^{-1}X'_{\text{-}Curv}$ where $X_{\text{-}Curv}$ is the design matrix, X, excluding the second column that represents the center point. Thus, the extra sum of squares corresponding to Curvature is:

$$\begin{aligned} SS_{Curvature} = & \quad y'[H - (1/9)J]y - y'[H_{\text{-}Center} - (1/9)J]y \\ = & \quad 0.7036 - 0.6875 \\ = & \quad 0.0161 \end{aligned}$$

This extra sum of squares can be used to test for the significance of curvature. The corresponding mean square is:

$$\begin{aligned} MS_{Curvature} = & \quad \frac{Sum\ of\ squares\ corresponding\ to\ Curvature}{degrees\ of\ freedom} \\ = & \quad \frac{0.0161}{1} \\ = & \quad 0.0161 \end{aligned}$$

Calculation of the Test Statistic

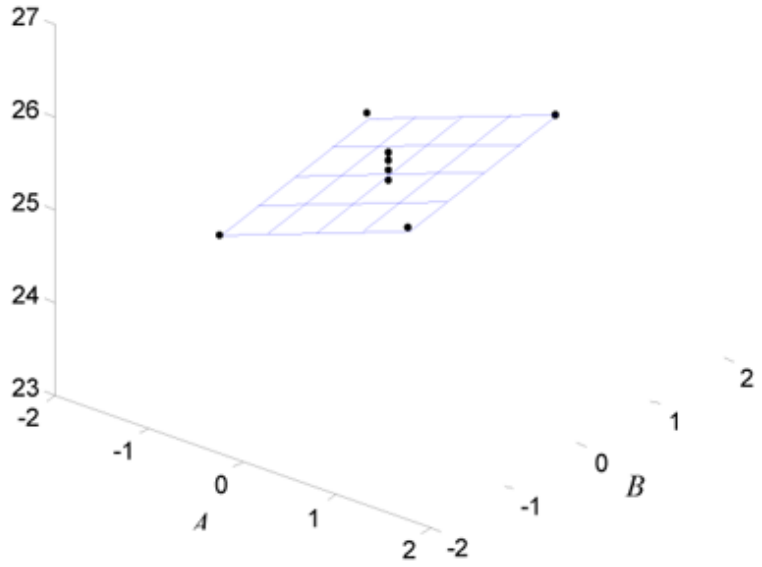
Knowing the mean squares, the statistic to check the significance of curvature can be calculated.

$$\begin{aligned} (f_0)_{Curvature} = & \quad \frac{MS_{Curvature}}{MS_E} \\ = & \quad \frac{0.0161/1}{0.052/4} \\ = & \quad 1.24 \end{aligned}$$

The p value corresponding to the statistic, $(f_0)_{Curvature} = 1.24$, based on the F distribution with one degree of freedom in the numerator and four degrees of freedom in the denominator is:

$$\begin{aligned} p\ value = & \quad 1 - P(F \leq (f_0)_{Curvature}) \\ = & \quad 1 - 0.6713 \\ = & \quad 0.3287 \end{aligned}$$

Assuming that the desired significance is 0.1, since p value > 0.1, it can be concluded that curvature does not exist for this design. This results is shown in the ANOVA table in the figure above. The surface of the fitted model based on these results, along with the observed response values, is shown in the figure below.



Blocking in 2^k Designs

Blocking can be used in the 2^k designs to deal with cases when replicates cannot be run under identical conditions. Randomized complete block designs that were discussed in Randomization and Blocking in DOE for factorial experiments are also applicable here. At times, even with just two levels per factor, it is not possible to run all treatment combinations for one replicate of the experiment under homogeneous conditions. For example, each replicate of the 2^2 design requires four runs. If each run requires two hours and testing facilities are available for only four hours per day, two days of testing would be required to run one complete replicate. Blocking can be used to separate the treatment runs on the two different days. Blocks that do not contain all treatments of a replicate are called incomplete blocks. In incomplete block designs, the block effect is confounded with certain effect(s) under investigation. For the 2^2 design assume that treatments (1) and ab were run on the first day and treatments a and b were run on the second day. Then, the incomplete block design for this experiment is:

Block 1	Block 2
(1)	a
ab	b

For this design the block effect may be calculated as:

$$\begin{aligned} Block\ Effect = & \quad Average\ response\ for\ Block\ 1 - \\ & Average\ response\ for\ Block\ 2 \\ = & \quad \frac{(1)+ab}{2} - \frac{a+b}{2} \\ = & \quad \frac{1}{2}[(1)+ab-a-b] \end{aligned}$$

The AB interaction effect is:

$$\begin{aligned} AB = & \quad Average\ response\ at\ A_{\text{high}}\text{-}B_{\text{high}}\ and\ A_{\text{low}}\text{-}B_{\text{low}} - \\ & Average\ response\ at\ A_{\text{low}}\text{-}B_{\text{high}}\ and\ A_{\text{high}}\text{-}B_{\text{low}} \\ = & \quad \frac{ab+(1)}{2} - \frac{b+a}{2} \\ = & \quad \frac{1}{2}[(1)+ab-a-b] \end{aligned}$$

The two equations given above show that, in this design, the AB interaction effect cannot be distinguished from the block effect because the formulas to calculate these effects are the same. In other words, the AB interaction is said to be confounded with the block effect and it is not possible to say if the effect calculated based on these equations is due to the AB interaction effect, the block effect or both. In incomplete block designs some effects are always confounded with the blocks. Therefore, it is important to design these experiments in such a way that the important effects are not confounded with the blocks. In most cases, the experimenter can assume that higher order interactions are unimportant. In this case, it would better to use incomplete block designs that confound these effects with the blocks. One way to design incomplete block designs is to use defining contrasts as shown next:

$$L = \alpha_1 q_1 + \alpha_2 q_2 + \ldots + \alpha_k q_k$$

where the α_i s are the exponents for the factors in the effect that is to be confounded with the block effect and the q_i s are values based on the level of the i the factor (in a treatment that is to be allocated to a block). For 2^k designs the α_i s are either 0 or 1 and the q_i s have a value of 0 for the low level of the i th factor and a value of 1 for the high level of the factor in the treatment under consideration. As an example, consider the 2^2 design where the interaction effect AB is confounded with the block. Since there are two factors, $k = 2$, with $i = 1$ representing factor A and $i = 2$ representing factor B. Therefore:

$$L = \alpha_1 q_1 + \alpha_2 q_2$$

The value of α_1 is one because the exponent of factor A in the confounded interaction AB is one. Similarly, the value of α_2 is one because the exponent of factor B in the confounded interaction AB is also one. Therefore, the defining contrast for this design can be written as:

$$\begin{aligned} L = & \quad \alpha_1 q_1 + \alpha_2 q_2 \\ = & \quad 1 \cdot q_1 + 1 \cdot q_2 \\ = & \quad q_1 + q_2 \end{aligned}$$

Once the defining contrast is known, it can be used to allocate treatments to the blocks. For the 2^2 design, there are four treatments (1), a, b and ab. Assume that $L = 0$ represents block 2 and $L = 1$ represents block 1. In order to decide which block the treatment (1) belongs to, the levels of factors A and B for this run are used. Since factor A is at the low level in this treatment, $q_1 = 0$. Similarly, since factor B is also at the low level in this treatment, $q_2 = 0$. Therefore:

$$\begin{aligned} L = & \quad q_1 + q_2 \\ = & \quad 0 + 0 = 0 \pmod 2 \end{aligned}$$

Note that the value of L used to decide the block allocation is "mod 2" of the original value. This value is obtained by taking the value of 1 for odd numbers and 0 otherwise. Based on the value of L, treatment (1) is assigned to block 1. Other treatments can be assigned using the following calculations:

$$\begin{aligned} (1): & \quad L = 0 + 0 = 0 = 0 \pmod 2 \\ a: & \quad L = 1 + 0 = 1 = 1 \pmod 2 \\ b: & \quad L = 0 + 1 = 1 = 1 \pmod 2 \\ ab: & \quad L = 1 + 1 = 2 = 0 \pmod 2 \end{aligned}$$

Therefore, to confound the interaction AB with the block effect in the 2^2 incomplete block design, treatments (1) and ab (with $L = 0$) should be assigned to block 2 and treatment combinations a and b (with $L = 1$) should be assigned to block 1.

Example: Two Level Factorial Design with Two Blocks

This example illustrates how treatments can be allocated to two blocks for an unreplicated 2^k design. Consider the unreplicated 2^4 design to investigate the four factors affecting the defects in automobile vinyl panels discussed in Normal Probability Plot of Effects. Assume that the 16 treatments required for this experiment were run by two different operators with each operator conducting 8 runs. This experiment is an example of an incomplete block design. The analyst in charge of this experiment assumed that the interaction $ABCD$ was not significant and decided to allocate treatments to the two operators so that the $ABCD$ interaction was confounded with the block effect (the two operators are the blocks). The allocation scheme to assign treatments to the two operators can be obtained as follows.

The defining contrast for the 2^4 design where the $ABCD$ interaction is confounded with the blocks is:

$$L = q_1 + q_2 + q_3 + q_4$$

The treatments can be allocated to the two operators using the values of the defining contrast. Assume that $L=0$ represents block 2 and $L=1$ represents block 1. Then the value of the defining contrast for treatment a is:

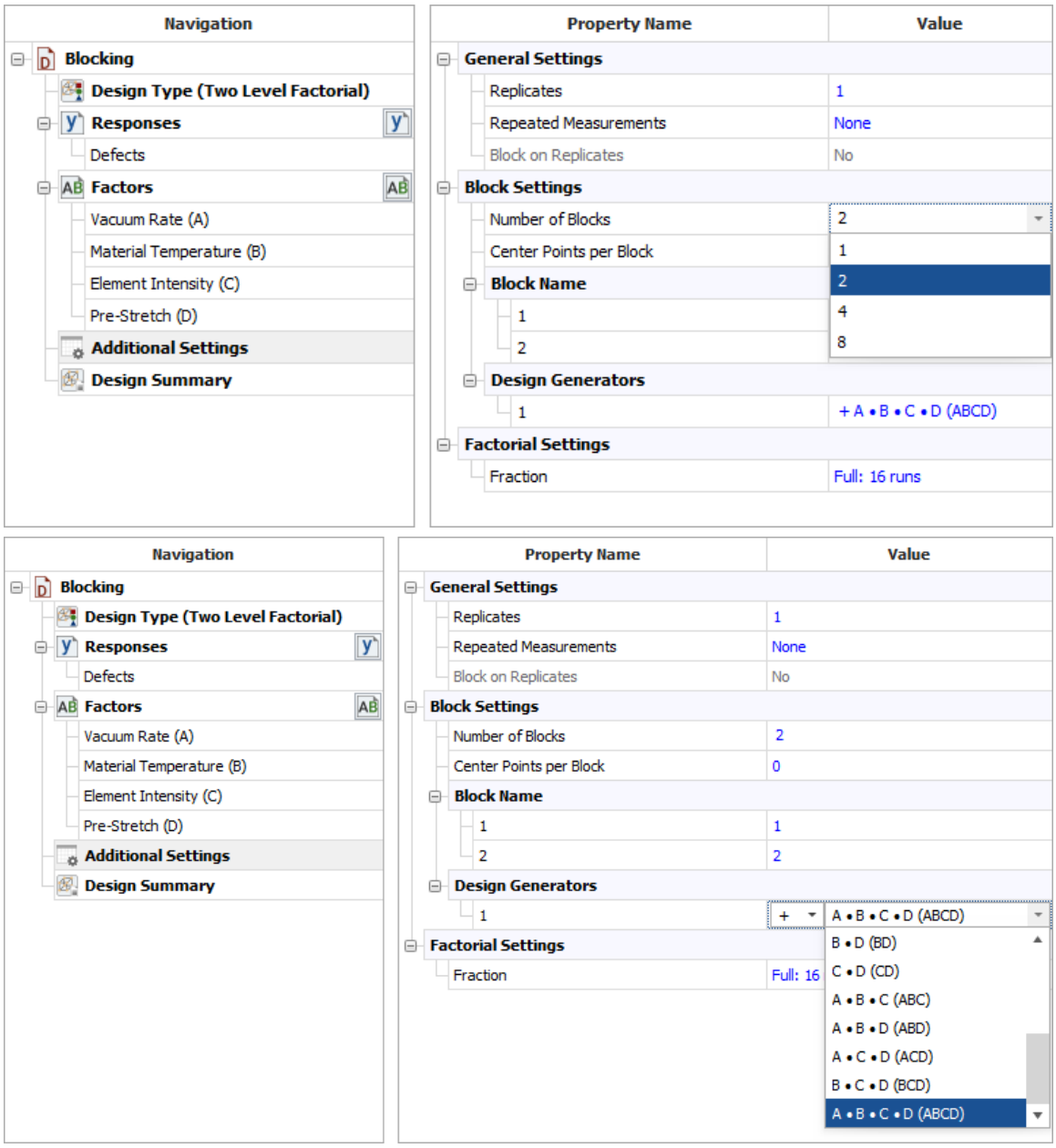
$$a \ : \ L=1+0+0+0=1=1 \ (\text{mod } 2)$$

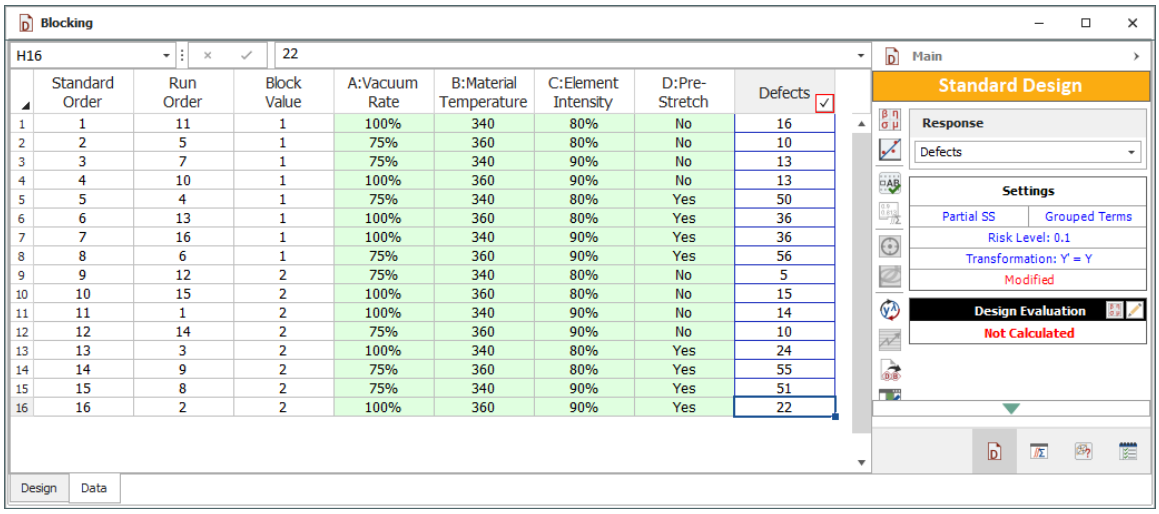
Therefore, treatment a should be assigned to Block 1 or the first operator. Similarly, for treatment ab we have:

$$ab \ : \ L=1+1+0+0=2=0 \ (\text{mod } 2)$$

Block 2 (L=0)	Block 1 (L=1)
(1)	a
ab	b
ac	c
bc	d
ad	abc
bd	bcd
cd	acd
$abcd$	abd

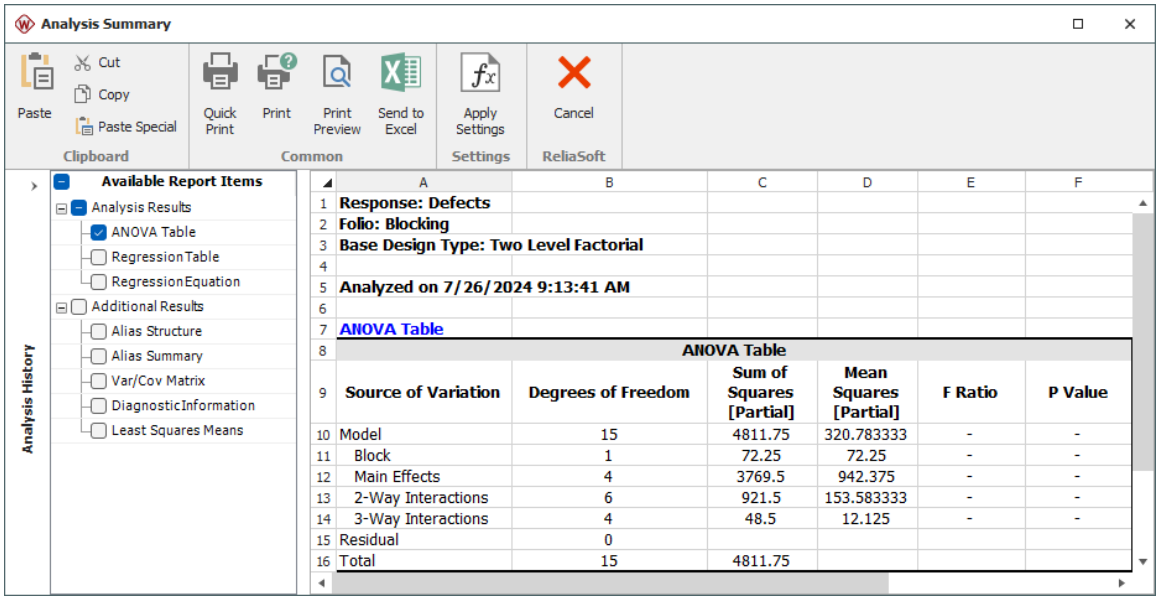
Therefore, ab should be assigned to Block 2 or the second operator. Other treatments can be allocated to the two operators in a similar manner to arrive at the allocation scheme shown in the figure below. In a DOE folio, to confound the $ABCD$ interaction for the 2^4 design into two blocks, the number of blocks are specified as shown in the figure below. Then the interaction $ABCD$ is entered in the Block Generator window (second following figure) which is available using the Block Generator button in the following figure. The design generated by the Weibull++ DOE folio is shown in the third of the following figures. This design matches the allocation scheme of the preceding figure.


Navigation
Blocking
Design Type (Two Level Factorial)
Responses
Defects
Factors
Vacuum Rate (A)
Material Temperature (B)
Element Intensity (C)
Pre-Stretch (D)
Additional Settings
Design Summary
Property Name
Value
General Settings
Replicates
1
Repeated Measurements
None
Block on Replicates
No
Block Settings
Number of Blocks
2
Center Points per Block
1
Block Name
2
1
2
4
8
Design Generators
1
+ A • B • C • D (ABCD)
Factorial Settings
Fraction
Full: 16 runs
Navigation
Blocking
Design Type (Two Level Factorial)
Responses
Defects
Factors
Vacuum Rate (A)
Material Temperature (B)
Element Intensity (C)
Pre-Stretch (D)
Additional Settings
Design Summary
Property Name
Value
General Settings
Replicates
1
Repeated Measurements
None
Block on Replicates
No
Block Settings
Number of Blocks
2
Center Points per Block
0
Block Name
1
1
2
2
Design Generators
1
+
A • B • C • D (ABCD)
B • D (BD)
C • D (CD)
A • B • C (ABC)
A • B • D (ABD)
A • C • D (ACD)
B • C • D (BCD)
A • B • C • D (ABCD)
Factorial Settings
Fraction
Full: 16



	Standard Order	Run Order	Block Value	A:Vacuum Rate	B:Material Temperature	C:Element Intensity	D:Pre-Stretch	Defects
1	1	11	1	100%	340	80%	No	16
2	2	5	1	75%	360	80%	No	10
3	3	7	1	75%	340	90%	No	13
4	4	10	1	100%	360	90%	No	13
5	5	4	1	75%	340	80%	Yes	50
6	6	13	1	100%	360	80%	Yes	36
7	7	16	1	100%	340	90%	Yes	36
8	8	6	1	75%	360	90%	Yes	56
9	9	12	2	75%	340	80%	No	5
10	10	15	2	100%	360	80%	No	15
11	11	1	2	100%	340	90%	No	14
12	12	14	2	75%	360	90%	No	10
13	13	3	2	100%	340	80%	Yes	24
14	14	9	2	75%	360	80%	Yes	55
15	15	8	2	75%	340	90%	Yes	51
16	16	2	2	100%	360	90%	Yes	22

For the analysis of this design, the sum of squares for all effects are calculated assuming no blocking. Then, to account for blocking, the sum of squares corresponding to the $ABCD$ interaction is considered as the sum of squares due to blocks and $ABCD$. In the DOE folio, this is done by displaying this sum of squares as the sum of squares due to the blocks. This is shown in the following figure where the sum of squares in question is obtained as 72.25 and is displayed against Block. The interaction ABCD, which is confounded with the blocks, is not displayed. Since the design is unreplicated, any of the methods to analyze unreplicated designs mentioned in Unreplicated 2^k designs have to be used to identify significant effects.



Response: Defects
Folio: Blocking
Base Design Type: Two Level Factorial

Analyzed on 7/26/2024 9:13:41 AM

ANOVA Table

ANOVA Table

Source of Variation	Degrees of Freedom	Sum of Squares [Partial]	Mean Squares [Partial]	F Ratio	P Value
Model	15	4811.75	320.783333	-	-
Block	1	72.25	72.25	-	-
Main Effects	4	3769.5	942.375	-	-
2-Way Interactions	6	921.5	153.583333	-	-
3-Way Interactions	4	48.5	12.125	-	-
Residual	0				
Total	15	4811.75			

Unreplicated 2^*k* Designs in 2^*p* Blocks

A single replicate of the 2^k design can be run in up to 2^p blocks where $p < k$. The number of effects confounded with the blocks equals the degrees of freedom associated with the block

effect.

If two blocks are used (the block effect has two levels), then one $(2-1=1)$ effect is confounded with the blocks. If four blocks are used, then three $(4-1=3)$ effects are confounded with the blocks and so on. For example an unreplicated 2^4 design may be confounded in 2^2 (four) blocks using two contrasts, L_1 and L_2. Let AC and BD be the effects to be confounded with the blocks. Corresponding to these two effects, the contrasts are respectively:

$$\begin{aligned} L_1 &= q_1 + q_3 \\ L_2 &= q_2 + q_4 \end{aligned}$$

Based on the values of L_1 and L_2, the treatments can be assigned to the four blocks as follows:

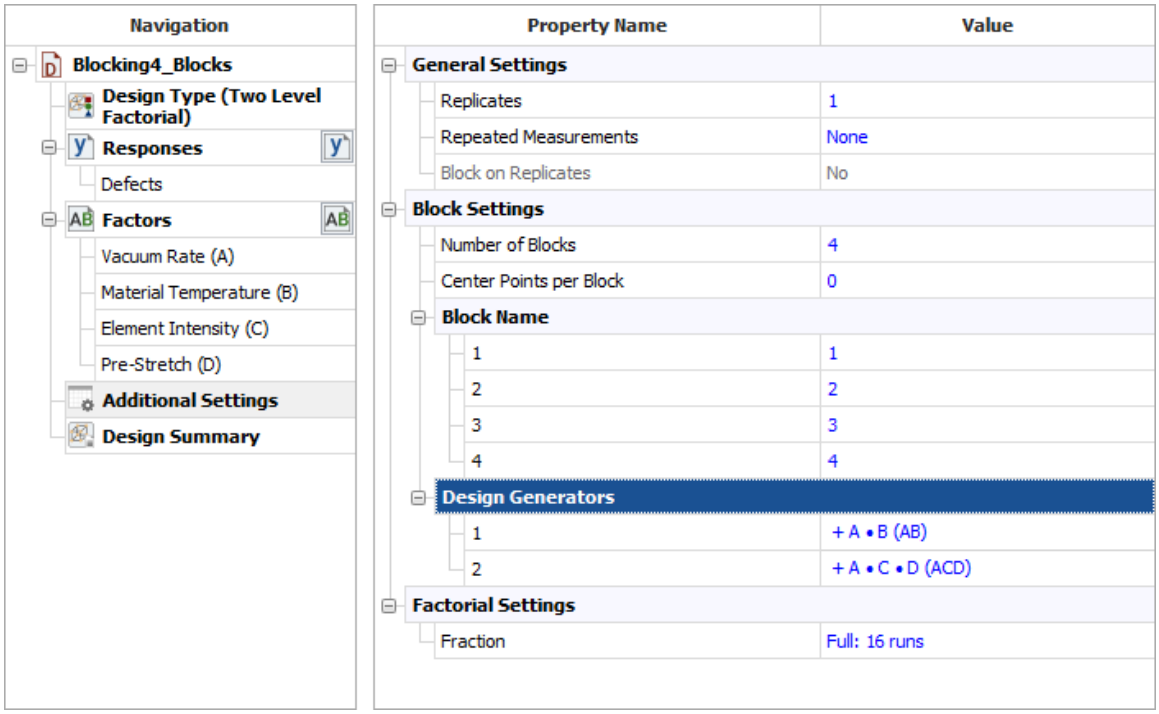
Block 4	Block 3	Block 2	Block 1
$L_1=0, L_2=0$	$L_1=1, L_2=0$	$L_1=0, L_2=1$	$L_1=1, L_2=1$
$\begin{bmatrix}(1)\\ac\\bd\\abcd\end{bmatrix}$	$\begin{bmatrix}a\\c\\abd\\bcd\end{bmatrix}$	$\begin{bmatrix}b\\abc\\d\\acd\end{bmatrix}$	$\begin{bmatrix}ab\\bc\\ad\\cd\end{bmatrix}$

Since the block effect has three degrees of freedom, three effects are confounded with the block effect. In addition to AC and BD, the third effect confounded with the block effect is their generalized interaction, $(AC)(BD)=ABCD$. In general, when an unreplicated 2^k design is confounded in 2^p blocks, p contrasts are needed $(L_1, L_2 \ldots L_p)$. p effects are selected to define these contrasts such that none of these effects are the generalized interaction of the others. The 2^p blocks can then be assigned the treatments using the p contrasts. $2^p-(p+1)$ effects, that are also confounded with the blocks, are then obtained as the generalized interaction of the p effects. In the statistical analysis of these designs, the sum of squares are computed as if no blocking were used. Then the block sum of squares is obtained by adding the sum of squares for all the effects confounded with the blocks.

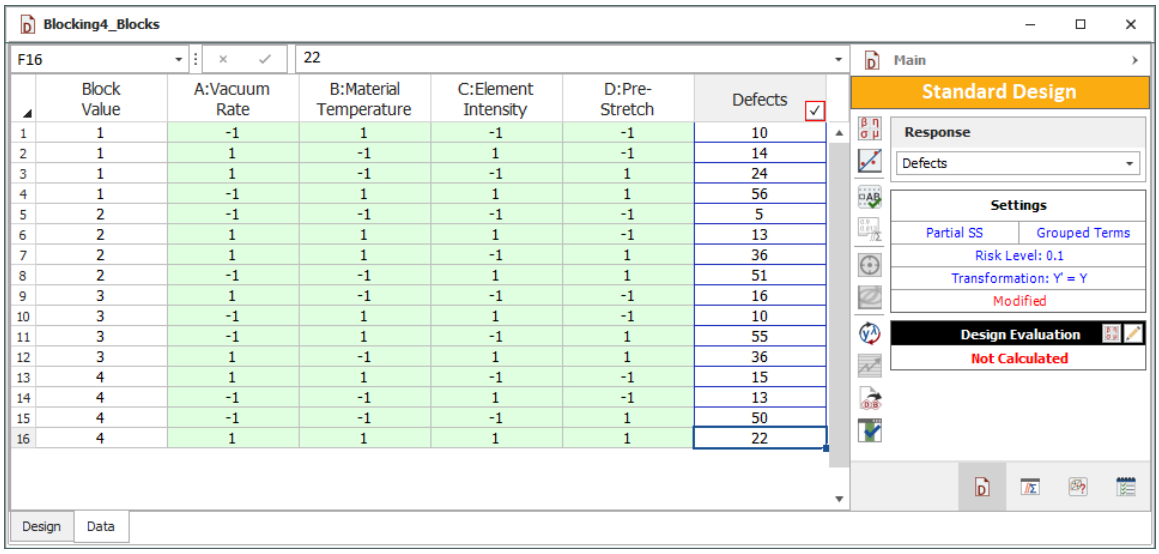
Example: 2 Level Factorial Design with Four Blocks

This example illustrates how a DOE folio obtains the sum of squares when treatments for an unreplicated 2^k design are allocated among four blocks. Consider again the unreplicated 2^4 design used to investigate the defects in automobile vinyl panels presented in Normal Probability Plot of Effects. Assume that the 16 treatments needed to complete the experiment were run by four operators. Therefore, there are four blocks. Assume that the treatments were allocated to the blocks using the generators mentioned in the previous section, i.e., treatments were allocated among the four operators by confounding the effects, AC and BD, with the blocks. These effects can be specified as Block Generators as shown in the following figure. (The generalized interaction of these two effects, interaction $ABCD$, will also get confounded with the

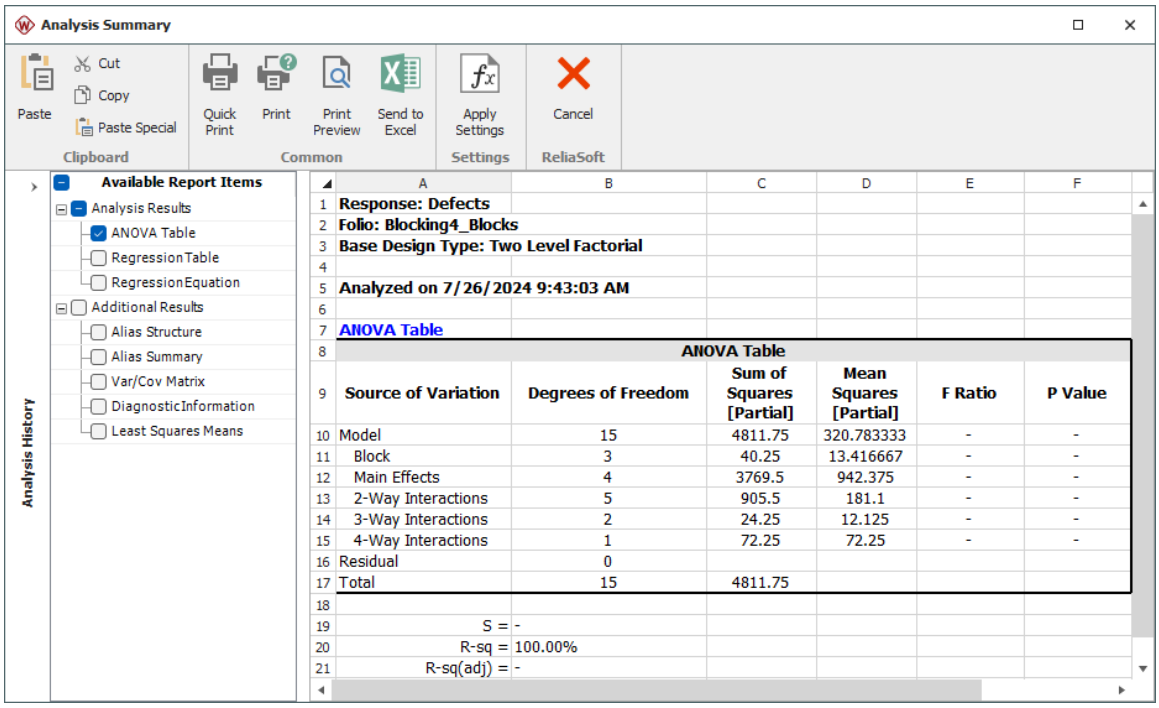
blocks.) The resulting design is shown in the second following figure and matches the allocation scheme obtained in the previous section.



The sum of squares in this case can be obtained by calculating the sum of squares for each of the effects assuming there is no blocking. Once the individual sum of squares have been obtained, the block sum of squares can be calculated. The block sum of squares is the sum of the sum of squares of effects, AC, BD and $ABCD$, since these effects are confounded with the block effect. As shown in the second following figure, this sum of squares is 92.25 and is displayed against Block. The interactions AC, BD and $ABCD$, which are confounded with the blocks, are not displayed. Since the present design is unreplicated any of the methods to analyze unreplicated designs mentioned in Unreplicated 2^k designs have to be used to identify significant effects.



	Block Value	A:Vacuum Rate	B:Material Temperature	C:Element Intensity	D:Pre-Stretch	Defects
1	1	-1	1	-1	-1	10
2	1	1	-1	1	-1	14
3	1	1	-1	-1	1	24
4	1	-1	1	1	1	56
5	2	-1	-1	-1	-1	5
6	2	1	1	1	-1	13
7	2	1	1	-1	1	36
8	2	-1	-1	1	1	51
9	3	1	-1	-1	-1	16
10	3	-1	1	1	-1	10
11	3	-1	1	-1	1	55
12	3	1	-1	1	1	36
13	4	1	1	-1	-1	15
14	4	-1	-1	1	-1	13
15	4	-1	-1	-1	1	50
16	4	1	1	1	1	22



Response: Defects
Folio: Blocking4_Blocks
Base Design Type: Two Level Factorial

Analyzed on 7/26/2024 9:43:03 AM

ANOVA Table

ANOVA Table

Source of Variation	Degrees of Freedom	Sum of Squares [Partial]	Mean Squares [Partial]	F Ratio	P Value
Model	15	4811.75	320.783333	-	-
Block	3	40.25	13.416667	-	-
Main Effects	4	3769.5	942.375	-	-
2-Way Interactions	5	905.5	181.1	-	-
3-Way Interactions	2	24.25	12.125	-	-
4-Way Interactions	1	72.25	72.25	-	-
Residual	0				
Total	15	4811.75			

S = -
R-sq = 100.00%
R-sq(adj) = -

Variability Analysis

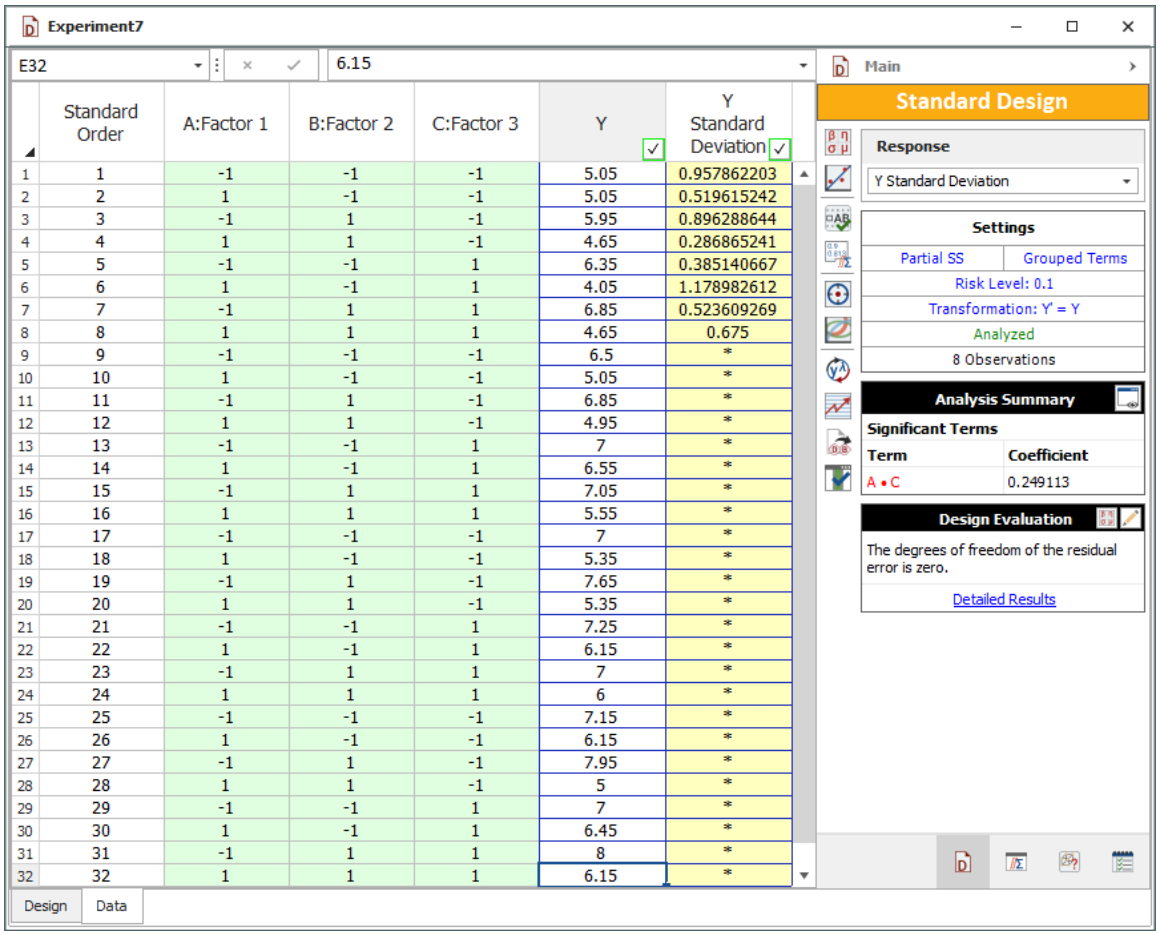
For replicated two level factorial experiments, the DOE folio provides the option of conducting variability analysis (using the Variability Analysis icon under the Data menu). The analysis is used to identify the treatment that results in the least amount of variation in the product or process being investigated. Variability analysis is conducted by treating the standard deviation of the response for each treatment of the experiment as an additional response. The standard deviation for a treatment is obtained by using the replicated response values at that treatment run. As

an example, consider the 2^3 design shown in the following figure where each run is replicated four times. A variability analysis can be conducted for this design. The DOE folio calculates eight standard deviation values corresponding to each treatment of the design (see second following figure). Then, the design is analyzed as an unreplicated 2^3 design with the standard deviations (displayed as Y Standard Deviation. in second following figure) as the response. The normal probability plot of effects identifies AC as the effect that influences variability (see third figure following). Based on the effect coefficients obtained in the fourth figure following, the model for Y Std. is:

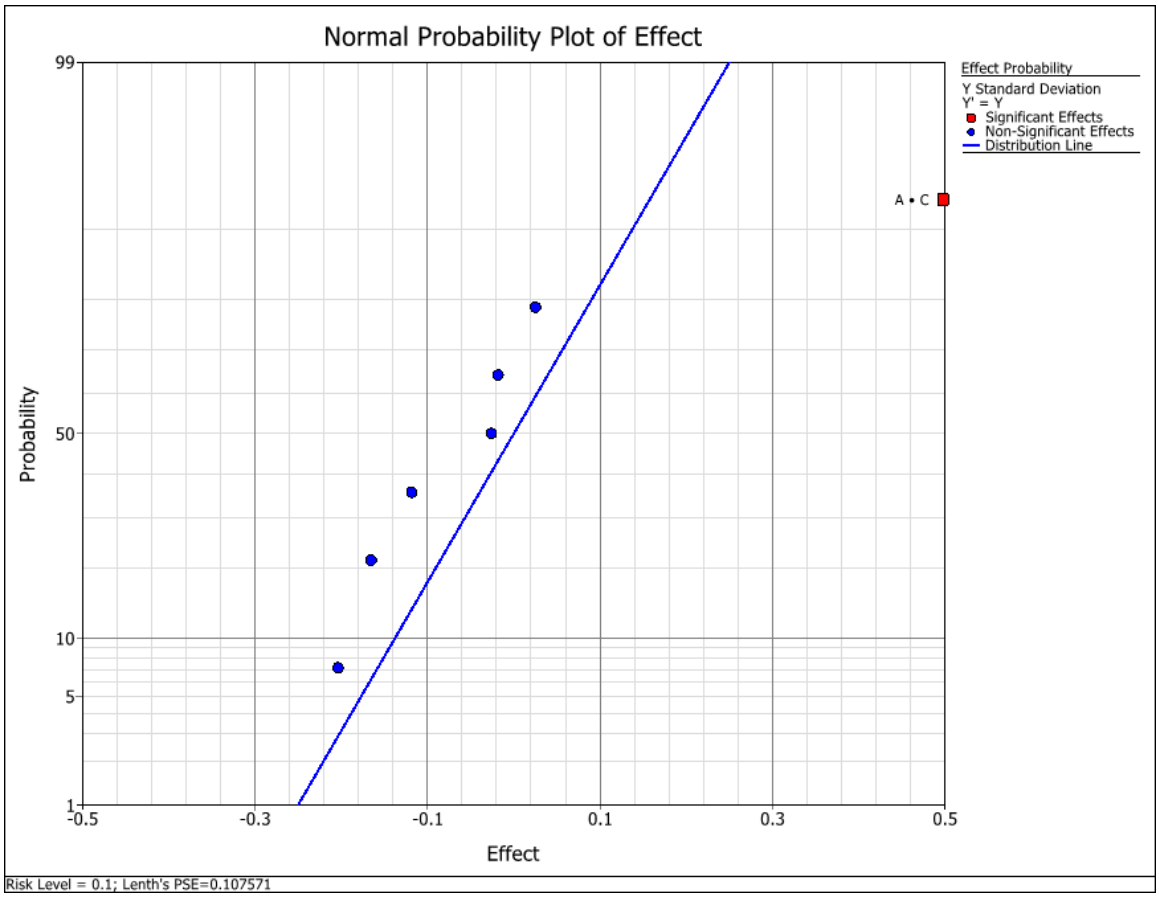
$$\begin{aligned} \text{Y Std.} = & \quad 0.6779 + 0.2491 \cdot AC \\ = & \quad 0.6779 + 0.2491 x_1 x_3 \end{aligned}$$

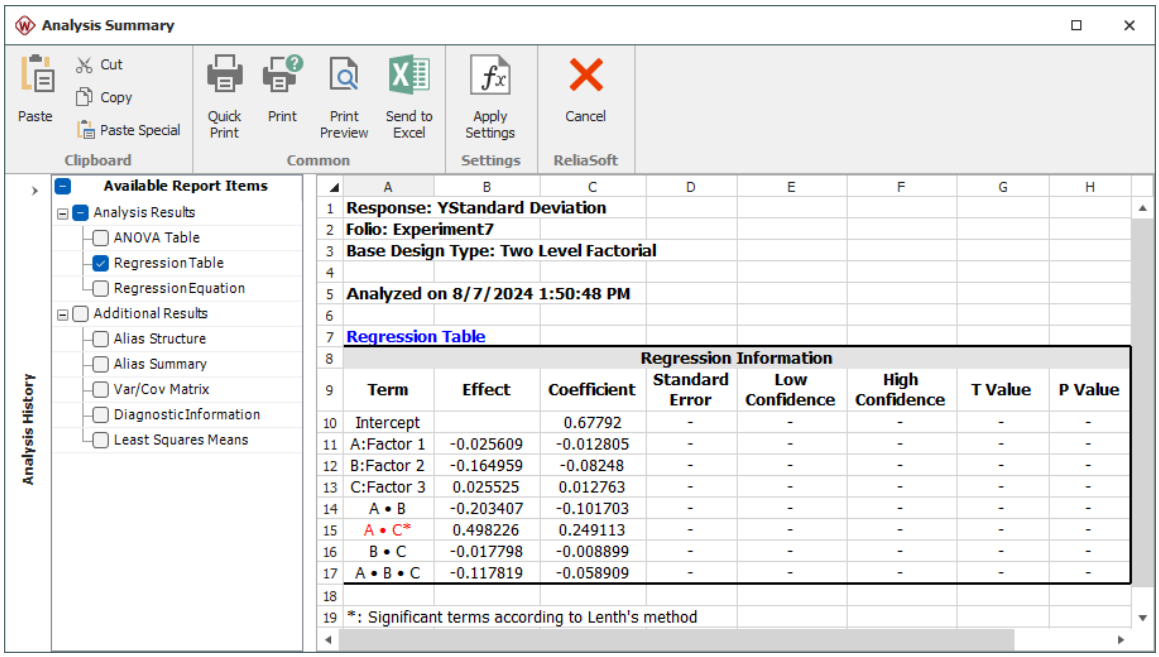
Based on the model, the experimenter has two choices to minimize variability (by minimizing Y Std.). The first choice is that x_1 should be 1 (i.e., A should be set at the high level) and x_3 should be -1 (i.e., C should be set at the low level). The second choice is that x_1 should be -1 (i.e., A should be set at the low level) and x_3 should be -1 (i.e., C should be set at the high level). The experimenter can select the most feasible choice.

Factors			Replicates			
A	*B*	*C*	1	2	3	4
-1	-1	-1	5.05	6.50	7.00	7.15
1	-1	-1	5.05	5.05	5.35	6.15
-1	1	-1	5.95	6.85	7.65	7.95
1	1	-1	4.65	4.95	5.35	5.00
-1	-1	1	6.35	7.00	7.25	7.00
1	-1	1	4.05	6.55	6.15	6.45
-1	1	1	6.85	7.05	7.00	8.00
1	1	1	4.65	5.55	6.00	6.15



	Standard Order	A:Factor 1	B:Factor 2	C:Factor 3	Y	Y Standard Deviation
1	1	-1	-1	-1	5.05	0.957862203
2	2	1	-1	-1	5.05	0.519615242
3	3	-1	1	-1	5.95	0.896288644
4	4	1	1	-1	4.65	0.286865241
5	5	-1	-1	1	6.35	0.385140667
6	6	1	-1	1	4.05	1.178982612
7	7	-1	1	1	6.85	0.523609269
8	8	1	1	1	4.65	0.675
9	9	-1	-1	-1	6.5	*
10	10	1	-1	-1	5.05	*
11	11	-1	1	-1	6.85	*
12	12	1	1	-1	4.95	*
13	13	-1	-1	1	7	*
14	14	1	-1	1	6.55	*
15	15	-1	1	1	7.05	*
16	16	1	1	1	5.55	*
17	17	-1	-1	-1	7	*
18	18	1	-1	-1	5.35	*
19	19	-1	1	-1	7.65	*
20	20	1	1	-1	5.35	*
21	21	-1	-1	1	7.25	*
22	22	1	-1	1	6.15	*
23	23	-1	1	1	7	*
24	24	1	1	1	6	*
25	25	-1	-1	-1	7.15	*
26	26	1	-1	-1	6.15	*
27	27	-1	1	-1	7.95	*
28	28	1	1	-1	5	*
29	29	-1	-1	1	7	*
30	30	1	-1	1	6.45	*
31	31	-1	1	1	8	*
32	32	1	1	1	6.15	*





	A	B	C	D	E	F	G	H
1	**Response: YStandard Deviation**							
2	**Folio: Experiment7**							
3	**Base Design Type: Two Level Factorial**							
4								
5	**Analyzed on 8/7/2024 1:50:48 PM**							
6								
7	**Regression Table**							
8	**Regression Information**							
9	**Term**	**Effect**	**Coefficient**	**Standard Error**	**Low Confidence**	**High Confidence**	**T Value**	**P Value**
10	Intercept		0.67792	-	-	-	-	-
11	A:Factor 1	-0.025609	-0.012805	-	-	-	-	-
12	B:Factor 2	-0.164959	-0.08248	-	-	-	-	-
13	C:Factor 3	0.025525	0.012763	-	-	-	-	-
14	A • B	-0.203407	-0.101703	-	-	-	-	-
15	A • C*	0.498226	0.249113	-	-	-	-	-
16	B • C	-0.017798	-0.008899	-	-	-	-	-
17	A • B • C	-0.117819	-0.058909	-	-	-	-	-
18								
19	*: Significant terms according to Lenth's method							

Two Level Fractional Factorial Designs

Two Level Fractional Factorial Designs

As the number of factors in a two level factorial design increases, the number of runs for even a single replicate of the 2^k design becomes very large. For example, a single replicate of an eight factor two level experiment would require 256 runs. Fractional factorial designs can be used in these cases to draw out valuable conclusions from fewer runs. The basis of fractional factorial designs is the *sparsity of effects* principle.[Wu, 2000] The principle states that, most of the time, responses are affected by a small number of main effects and lower order interactions, while higher order interactions are relatively unimportant. Fractional factorial designs are used as screening experiments during the initial stages of experimentation. At these stages, a large number of factors have to be investigated and the focus is on the main effects and two factor interactions. These designs obtain information about main effects and lower order interactions with fewer experiment runs by confounding these effects with unimportant higher order interactions. As an example, consider a 2^8 design that requires 256 runs. This design allows for the investigation of 8 main effects and 28 two factor interactions. However, 219 degrees of freedom are devoted to three factor or higher order interactions. This full factorial design can prove to be very inefficient when these higher order interactions can be assumed to be unimportant. Instead, a fractional design can be used here to identify the important factors that can then be investigated more thoroughly in subsequent experiments. In unreplicated fractional factorial designs, no degrees of freedom are available to calculate the error sum of squares and the techniques mentioned in Unreplicated 2^k designs should be employed for the analysis of these designs.

Half-fraction Designs

A half-fraction of the 2^k design involves running only half of the treatments of the full factorial design. For example, consider a 2^3 design that requires eight runs in all. The design matrix for this design is shown in the figure (a) below. A half-fraction of this design is the design in which only four of the eight treatments are run. The fraction is denoted as 2^{3-1} with the "-1" in the index denoting a half-fraction. Assume that the treatments chosen for the half-fraction design are the ones where the interaction ABC is at the high level (i.e., only those rows are chosen from the following figure (a) where the column for ABC has entries of 1). The resulting 2^{3-1} design has a design matrix as shown in figure (b) below.

I	A	B	AB	C	AC	BC	ABC
1	-1	-1	1	-1	1	1	-1
1	1	-1	-1	-1	-1	1	1
1	-1	1	-1	-1	1	-1	1
1	1	1	1	-1	-1	-1	-1
1	-1	-1	1	1	-1	-1	1
1	1	-1	-1	1	1	-1	-1
1	-1	1	-1	1	-1	1	-1
1	1	1	1	1	1	1	1

(a)

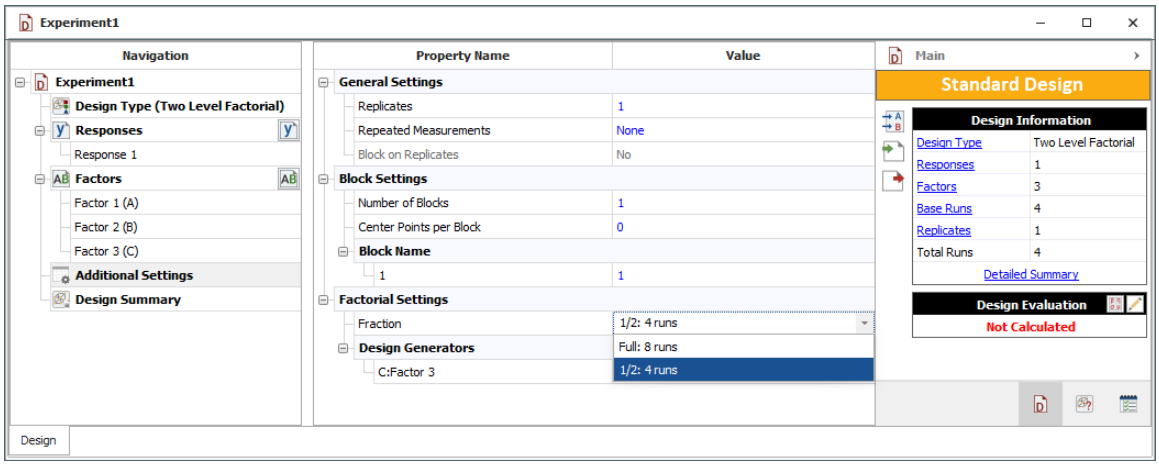
I	A	B	AB	C	AC	BC	ABC
1	1	-1	-1	-1	-1	1	1
1	-1	1	-1	-1	1	-1	1
1	-1	-1	1	1	-1	-1	1
1	1	1	1	1	1	1	1

(b)

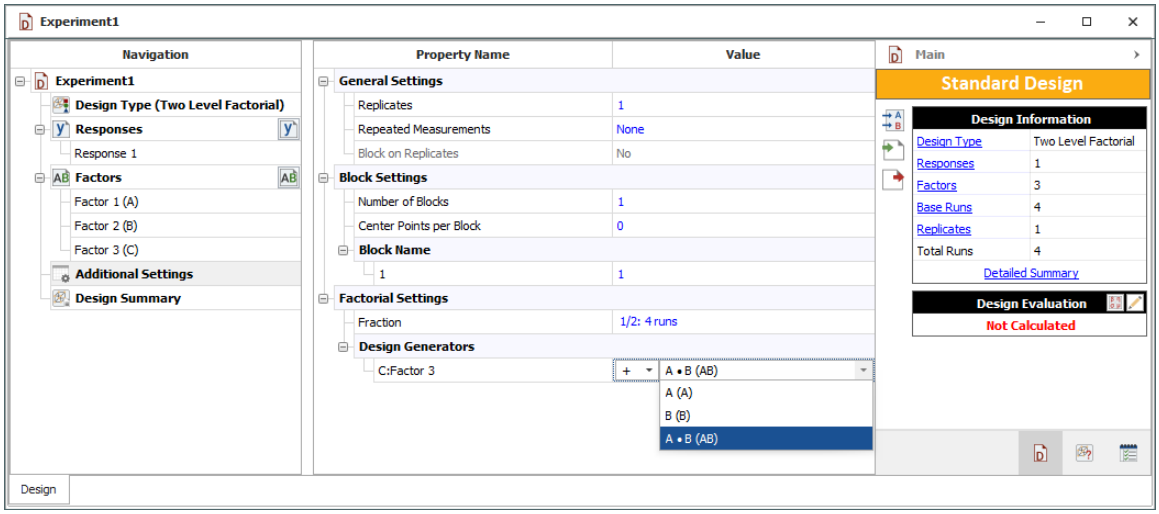
I	A	B	AB	C	AC	BC	ABC
1	-1	-1	1	-1	1	1	-1
1	1	1	1	-1	-1	-1	-1
1	1	-1	-1	1	1	-1	-1
1	-1	1	-1	1	-1	1	-1

(c)

In the 2^{3-1} design of figure (b), since the interaction ABC is always included at the same level (the high level represented by 1), it is not possible to measure this interaction effect. The effect, ABC, is called the *generator* or *word* for this design. It can be noted that, in the design matrix of the following figure (b), the column corresponding to the intercept, I, and column corresponding to the interaction ABC, are identical. The identical columns are written as $I = ABC$ and this equation is called the *defining relation* for the design. In a DOE folio, the present 2^{3-1} design can be obtained by specifying the design properties as shown in the following figure.



The defining relation, $I = ABC$, is entered in the Fraction Generator window as shown next.



Note that in the figure following that, the defining relation is specified as $C = AB$. This relation is obtained by multiplying the defining relation, $I = ABC$, by the last factor, C, of the design.

Calculation of Effects

Using the four runs of the 2^{3-1} design in figure (b) discussed above, the main effects can be calculated as follows:

$$A = \frac{(a+abc)}{2} - \frac{(b+c)}{2} = \frac{1}{2}(a-b-c+abc)$$

$$B = \frac{(b+abc)}{2} - \frac{(a+c)}{2} = \frac{1}{2}(-a+b-c+abc)$$

$$C = \frac{(c+abc)}{2} - \frac{(a+b)}{2} = \frac{1}{2}(-a-b+c+abc)$$

where a, b, c and abc are the treatments included in the 2^{3-1} design.

Similarly, the two factor interactions can also be obtained as:

$$
\begin{aligned}
BC = &\quad \frac{(a+abc)}{2} - \frac{(b+c)}{2} = \frac{1}{2}(a-b-c+abc) \\
AC = &\quad \frac{(b+abc)}{2} - \frac{(a+c)}{2} = \frac{1}{2}(-a+b-c+abc) \\
AB = &\quad \frac{(c+abc)}{2} - \frac{(a+b)}{2} = \frac{1}{2}(-a-b+c+abc)
\end{aligned}
$$

The equations for A and BC above result in the same effect values showing that effects A and BC are confounded in the present 2^{3-1} design. Thus, the quantity, $\frac{1}{2}(a-b-c+abc)$, estimates $A+BC$ (i.e., both the main effect A and the two-factor interaction BC). The effects, A and BC, are called *aliases*. From the remaining equations given above, it can be seen that the other aliases for this design are B and AC, and C and AB. Therefore, the equations to calculate the effects in the present 2^{3-1} design can be written as follows:

$$
\begin{aligned}
A+BC = &\quad \frac{1}{2}(a-b-c+abc) \\
B+AC = &\quad \frac{1}{2}(-a+b-c+abc) \\
C+AB = &\quad \frac{1}{2}(-a-b+c+abc)
\end{aligned}
$$

Calculation of Aliases

Aliases for a fractional factorial design can be obtained using the defining relation for the design. The defining relation for the present 2^{3-1} design is:

$$I = ABC$$

Multiplying both sides of the previous equation by the main effect, A, gives the alias effect of A:

$$
\begin{aligned}
A \cdot I = &\quad A \cdot ABC \\
A = &\quad A^2BC \\
A = &\quad BC
\end{aligned}
$$

Note that in calculating the alias effects, any effect multiplied by I remains the same ($A \cdot I = A$), while an effect multiplied by itself results in I ($A^2 = I$). Other aliases can also be obtained:

$$
\begin{aligned}
B \cdot I = &\quad B \cdot ABC \\
B = &\quad AB^2C \\
B = &\quad AC
\end{aligned}
$$

and:

$$\begin{aligned} C \cdot I &= C \cdot ABC \\ C &= ABC^2 \\ C &= AB \end{aligned}$$

Fold-over Design

If it can be assumed for this design that the two-factor interactions are unimportant, then in the absence of BC, AC and AB, the equations for (A+BC), (B+AC) and (C+AB) can be used to estimate the main effects, A, B and C, respectively. However, if such an assumption is not applicable, then to uncouple the main effects from their two factor aliases, the alternate fraction that contains runs having ABC at the lower level should be run. The design matrix for this design is shown in the preceding figure (c). The defining relation for this design is $I = -ABC$ because the four runs for this design are obtained by selecting the rows of the preceding figure (a) for which the value of the ABC column is -1. The aliases for this fraction can be obtained as explained in Half-fraction Designs as $A = -BC$, $B = -AC$ and $C = -AB$. The effects for this design can be calculated as:

$$\begin{aligned} A - BC &= \frac{1}{2}(ab + ac - (1) - bc) \\ B - AC &= \frac{1}{2}(ab - ac + (1) - bc) \\ C - AB &= \frac{1}{2}(-ab + ac - (1) + bc) \end{aligned}$$

These equations can be combined with the equations for (A+BC), (B+AC) and (C+AB) to obtain the de-aliased main effects and two factor interactions. For example, adding equations (A+BC) and (A-BC) returns the main effect A.

$$\begin{aligned} 2A = \quad & \frac{1}{2}(a - b - c + abc) + \\ & \frac{1}{2}(ab + ac - (1) - bc) \end{aligned}$$

The process of augmenting a fractional factorial design by a second fraction of the same size by simply reversing the signs (of all effect columns except I) is called *folding over*. The combined design is referred to as a *fold-over design*.

Quarter and Smaller Fraction Designs

At times, the number of runs even for a half-fraction design are very large. In these cases, smaller fractions are used. A quarter-fraction design, denoted as 2^{k-2}, consists of a fourth of the runs of the full factorial design. Quarter-fraction designs require two defining relations. The first defining relation returns the half-fraction or the 2^{k-1} design. The second defining relation selects half of the runs of the 2^{k-1} design to give the quarter-fraction. For example, consider the 2^4 design. To obtain a

2^{4-2} design from this design, first a half-fraction of this design is obtained by using a defining relation. Assume that the defining relation used is $I = ABCD$. The design matrix for the resulting 2^{4-1} design is shown in figure (a) below. Now, a quarter-fraction can be obtained from the 2^{4-1} design shown in figure (a) below using a second defining relation $I = AD$. The resulting 2^{4-2} design obtained is shown in figure (b) below.

I	A	B	AB	C	AC	BC	D	AD	BD	ABC	ABD	CD	ACD	BCD	ABCD
1	-1	-1	1	-1	1	1	-1	1	1	-1	-1	1	-1	-1	1
1	1	-1	-1	-1	-1	1	1	1	-1	1	-1	-1	-1	1	1
1	-1	1	-1	-1	1	-1	1	-1	1	1	-1	-1	1	-1	1
1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1
1	-1	-1	1	1	-1	-1	1	-1	-1	1	1	1	-1	-1	1
1	1	-1	-1	1	1	-1	-1	-1	1	-1	1	-1	-1	1	1
1	-1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

(a)

I	A	B	AB	C	AC	BC	D	AD	BD	ABC	ABD	CD	ACD	BCD	ABCD
1	-1	-1	1	-1	1	1	-1	1	1	-1	-1	1	-1	-1	1
1	1	-1	-1	-1	-1	1	1	1	-1	1	-1	-1	-1	1	1
1	-1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

(b)

The complete defining relation for this 2^{4-2} design is:

$$I = ABCD = AD = BC$$

Note that the effect, BC, in the defining relation is the generalized interaction of $ABCD$ and AD and is obtained using $(ABCD)(AD) = A^2BCD^2 = BC$. In general, a 2^{k-p} fractional factorial design requires p independent generators. The defining relation for the design consists of the p independent generators and their 2^p - (p +1) generalized interactions.

Calculation of Aliases

The alias structure for the present 2^{4-2} design can be obtained using the defining relation of equation (I=ABCD=AD=BC) following the procedure explained in Half-fraction Designs. For example, multiplying the defining relation by A returns the effects aliased with the main effect, A, as follows:

$$A \cdot I = A \cdot ABCD = A \cdot AD = A \cdot BC$$
$$A = A^2BCD = A^2D = ABC$$
$$A = BCD = D = ABC$$

Therefore, in the present 2^{4-2} design, it is not possible to distinguish between effects A, D, BCD and ABC. Similarly, multiplying the defining relation by B and AB returns the effects that are aliased with these effects:

$$B = ACD = ABD = C$$
$$AB = CD = AD = AC$$

Other aliases can be obtained in a similar way. It can be seen that each effect in this design has three aliases. In general, each effect in a 2^{k-p} design has 2^{p-1} aliases. The aliases for the 2^{4-2} design show that in this design the main effects are aliased with each other (A is aliased with D and B is aliased with C). Therefore, this design is not a useful design and is not available in a Weibull++ DOE folio. It is important to ensure that main effects and lower order interactions of interest are not aliased in a fractional factorial design. This is known by looking at the resolution of the fractional factorial design.

Design Resolution

The resolution of a fractional factorial design is defined as the number of factors in the lowest order effect in the defining relation. For example, in the defining relation $I = ABCD = AD = BC$ of the previous 2^{4-2} design, the lowest-order effect is either AD or BC, containing two factors. Therefore, the resolution of this design is equal to two. The resolution of a fractional factorial design is represented using Roman numerals. For example, the previously mentioned 2^{4-2} design with a resolution of two can be represented as $2^{\frac{4-2}{II}}$. The resolution provides information about the confounding in the design as explained next:

1. **Resolution III Designs**
   In these designs, the lowest order effect in the defining relation has three factors (e.g., a 2^{5-2} design with the defining relation $I = ABDE = ABC = CDE$). In resolution III designs, no main effects are aliased with any other main effects, but main effects are aliased with two factor interactions. In addition, some two factor interactions are aliased with each other.

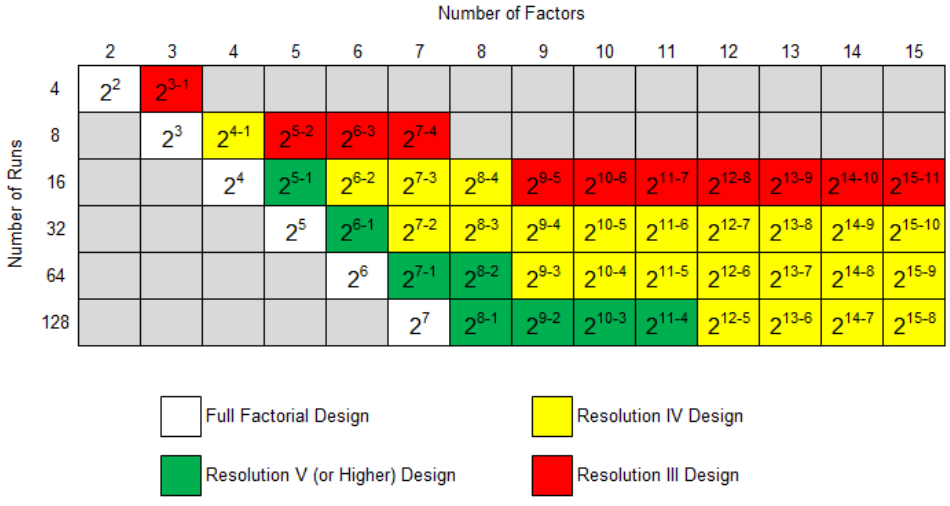
2. **Resolution IV Designs**
   In these designs, the lowest order effect in the defining relation has four factors (e.g., a 2^{5-1} design with the defining relation $I = ABDE$). In resolution IV designs, no main effects are aliased with any other main effects or two factor interactions. However, some main effects are aliased with three factor interactions and the two factor interactions are aliased with each other.

3. **Resolution V Designs**
   In these designs the lowest order effect in the defining relation has five factors (e.g., a 2^{5-1} design with the defining relation $I = ABCDE$). In resolution V designs, no main effects or two factor interactions are aliased with any other main effects or two factor interactions. However, some main effects are aliased with four factor interactions and the two factor interactions are aliased with three factor interactions.

Fractional factorial designs with the highest resolution possible should be selected because the higher the resolution of the design, the less severe the degree of confounding. In general, designs with a resolution less than III are never used because in these designs some of the main effects are aliased with each other. The table below shows fractional factorial designs with the highest available resolution for three to ten factor designs along with their defining relations.

Number of Factors	Fraction	Resolution	Runs	Defining Relation (excluding generalized interactions)
3	2^{3-1}	III	4	*I=ABC*
4	2^{4-1}	IV	8	*I=ABCD*
5	2^{5-1}	V	16	*I=ABCDE*
6	2^{6-1}	VI	32	*I=ABCDEF*
7	2^{7-1}	VII	64	*I=ABCDEFG*
8	2^{8-2}	V	64	*I=ABCDG=ABEFH*
9	2^{9-2}	VI	128	*I=ACDFGH=BCEFGJ*
10	2^{10-3}	V	128	*I=ABCGH=ACDEJ=ACDFK*

All of the two level fractional factorial designs available in a DOE folio are shown next.



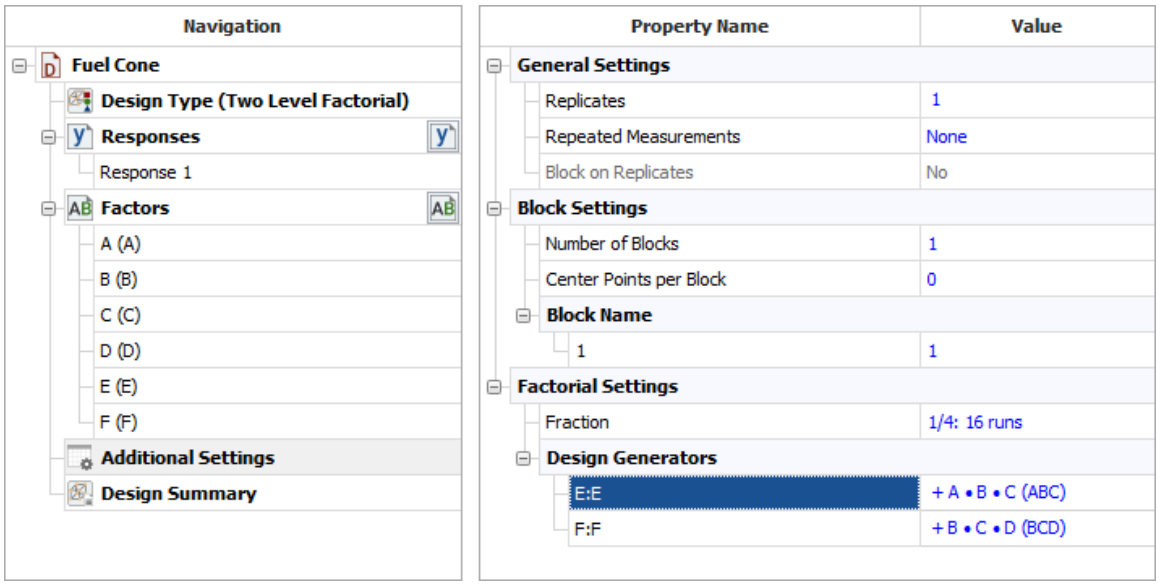
Minimum Aberration Designs

At times, different designs with the same resolution but different aliasing may be available. The best design to select in such a case is the minimum aberration design. For example, all 2^{7-2} designs in the fourth table have a resolution of four (since the generator with the minimum number of factors in each design has four factors). Design 1 has three generators of length four ($ABCF, BCDG, ADFG$). Design 2 has two generators of length four ($ABCF, ADEG$). Design 3 has one generator of length four ($CEFG$). Therefore, design 3 has the least number of generators with the minimum length of four. Design 3 is called the minimum aberration design. It can be seen that the alias structure for design 3 is less involved compared to the other designs. For details refer to [Wu, 2000].

2^{7-2} Design	Defining Relation	Alias relations (for two-factor interactions)
1	$I=ABCF=BCDG=ADFG$	$AB=CF, AC=BF, AD=FG,$ $AG=DF, BD=CG, BG=CD,$ $AF=BC=DG$
2	$I=ABCF=ADEG=BCDEFG$	$AB=CF, AC=BF, AD=EG,$ $AE=DG, AF=BC, AG=DE$
3	$I=ABCDF=ABDEG=CEFG$	$CE=FG, CF=EG, CG=EF$

EXAMPLE

The design of an automobile fuel cone is thought to be affected by six factors in the manufacturing process: cavity temperature (factor A), core temperature (factor B), melt temperature (factor C), hold pressure (factor D), injection speed (factor E) and cool time (factor F). The manufacturer of the fuel cone is unable to run the $2^6 = 64$ runs required to complete one replicate for a two level full factorial experiment with six factors. Instead, they decide to run a fractional factorial design. Considering that three factor and higher order interactions are likely to be inactive, the manufacturer selects a 2^{6-2} design that will require only 16 runs. The manufacturer chooses the resolution IV design which will ensure that all main effects are free from aliasing (assuming three factor and higher order interactions are absent). However, in this design the two factor interactions may be aliased with each other. It is decided that, if important two factor interactions are found to be present, additional experiment trials may be conducted to separate the aliased effects. The performance of the fuel cone is measured on a scale of 1 to 15. In a Weibull++ DOE folio, the design for this experiment is set up using the properties shown in the following figure. The Fraction Generators for the design, $E = ABC$ and $F = BCD$, are the same as the defaults used in a DOE folio. The resulting 2^{6-2} design and the corresponding response values are shown in the following two figures.



H16 | 13.5

	Standard Order	A:A	B:B	C:C	D:D	E:E	F:F	Response 1
1	1	-1	-1	-1	-1	-1	-1	7
2	2	1	-1	-1	-1	1	-1	7
3	3	-1	1	-1	-1	1	1	6.5
4	4	1	1	-1	-1	-1	1	7
5	5	-1	-1	1	-1	1	1	5
6	6	1	-1	1	-1	-1	1	5
7	7	-1	1	1	-1	-1	-1	3.5
8	8	1	1	1	-1	1	-1	4.5
9	9	-1	-1	-1	1	-1	1	7
10	10	1	-1	-1	1	1	1	7.5
11	11	-1	1	-1	1	1	-1	7
12	12	1	1	-1	1	-1	-1	4.5
13	13	-1	-1	1	1	1	-1	14
14	14	1	-1	1	1	-1	-1	11
15	15	-1	1	1	1	-1	1	13
16	16	1	1	1	1	1	1	13.5

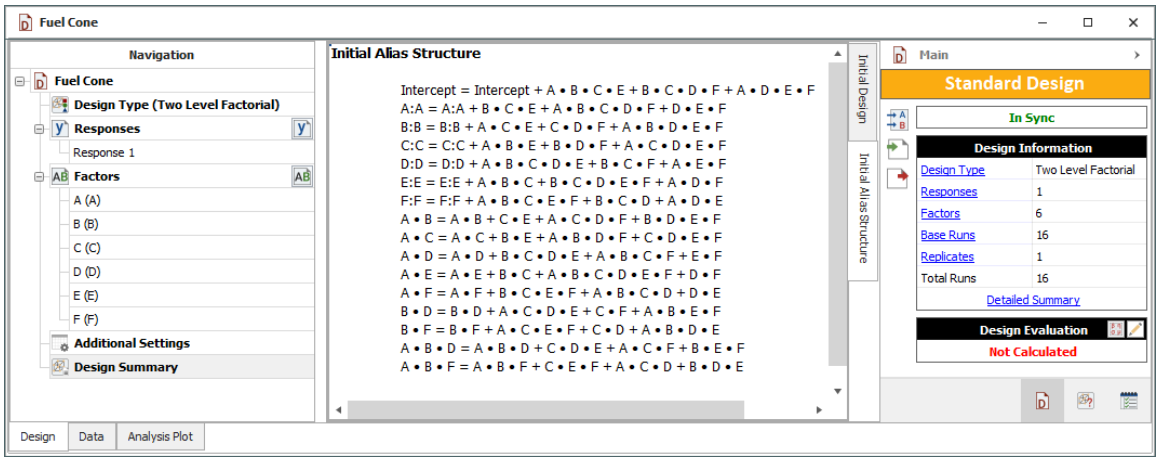
The complete alias structure for the $2^{\mathbf{6-2}}_{\mathbf{IV}}$ design is shown next.

$$I = ABCE = ADEF = BCDF$$

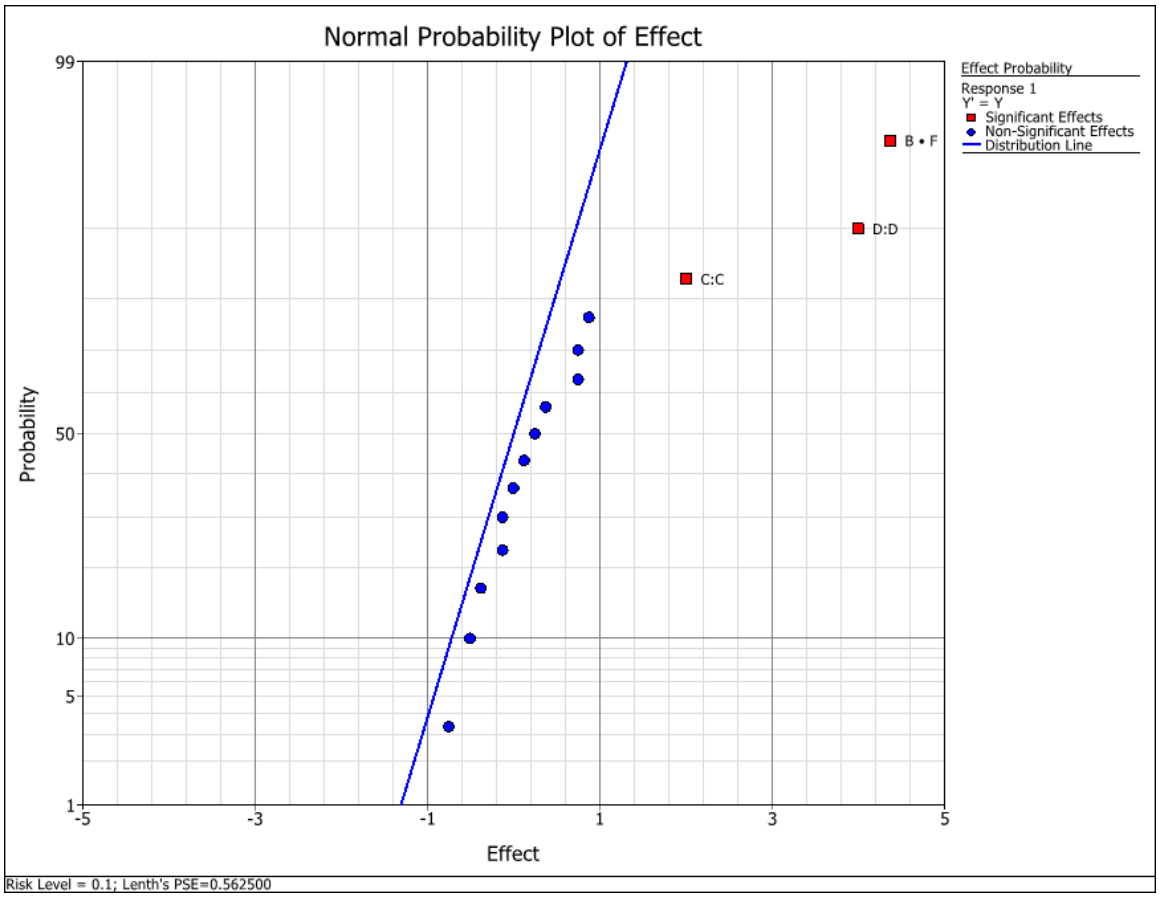
$$\begin{aligned} A &= BCE = DEF = ABCDF \\ B &= ACE = CDF = ABDEF \\ C &= ABE = BDF = ACDEF \\ D &= AEF = BCF = ABCDE \\ E &= ABC = ADF = BCDEF \\ F &= ADE = BCD = ABCEF \end{aligned}$$

$$
\begin{aligned}
AB &= CE = ACDF = BDEF \\
AC &= BE = ABDF = CDEF \\
AD &= EF = ABCF = BCDE \\
AE &= BC = DF = ABCDEF \\
AF &= DE = ABCD = BCEF \\
BD &= CF = ABEF = ACDE \\
BF &= CD = ABDE = ACEF \\
ABD &= ACF = BEF = CDE \\
ABF &= ACD = BDE = CEF
\end{aligned}
$$

In a DOE folio, the alias structure is displayed in the Design Summary and as part of the Design Evaluation result, as shown next:

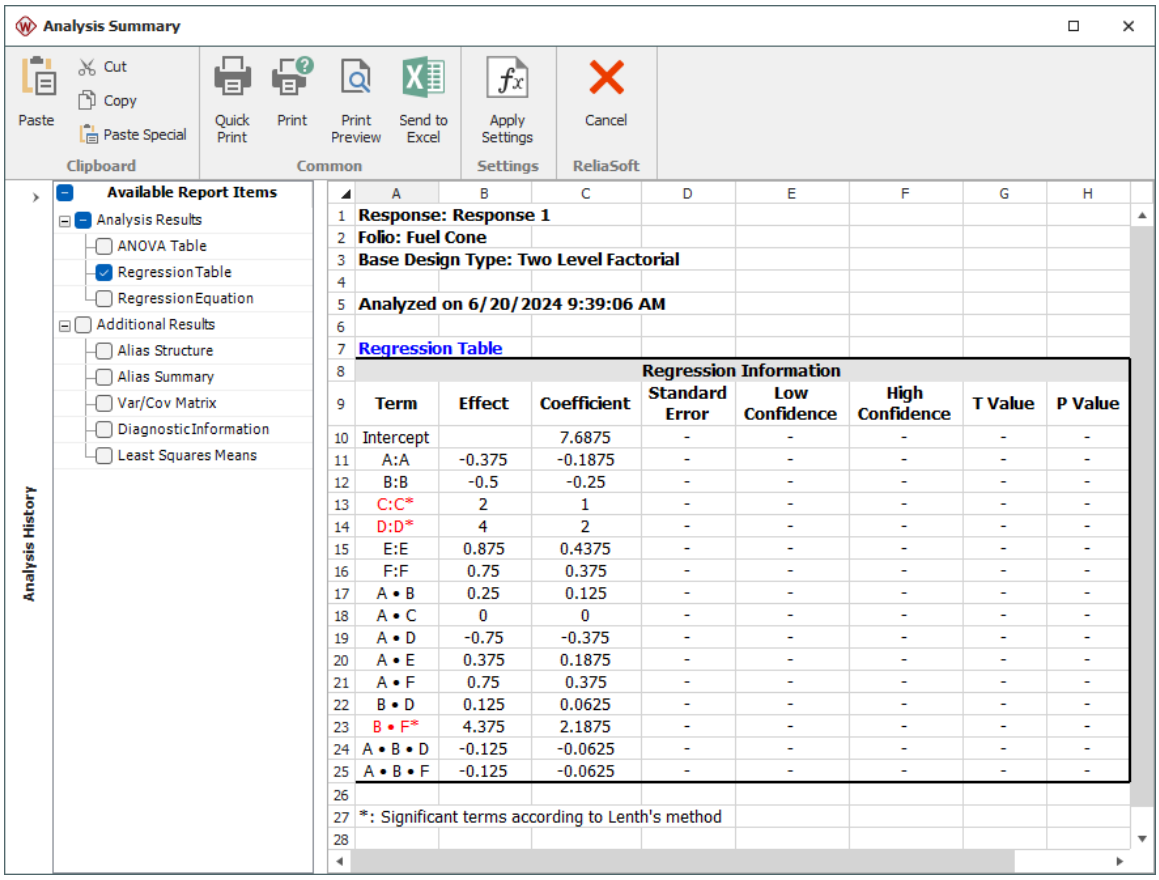


The normal probability plot of effects for this unreplicated design shows the main effects of factors C and D and the interaction effect, BF, to be significant (see the following figure).



From the alias structure, it can be seen that for the present design interaction effect, BF, is confounded with CD. Therefore, the actual source of this effect cannot be known on the basis of the present experiment. However because neither factor B nor F is found to be significant there is an indication the observed effect is likely due to interaction, CD. To confirm this, a follow-up 2^2 experiment is run involving only factors B and F. The interaction, BF, is found to be inactive, leading to the conclusion that the interaction effect in the original experiment is effect, CD. Given these results, the fitted regression model for the fuel cone design as per the coefficients obtained from the DOE folio is shown next.

$$\hat{y} = 7.6875 + C + 2 \cdot D + 2.1875 \cdot CD$$



Response: Response 1
Folio: Fuel Cone
Base Design Type: Two Level Factorial

Analyzed on 6/20/2024 9:39:06 AM

Regression Table

Regression Information

Term	Effect	Coefficient	Standard Error	Low Confidence	High Confidence	T Value	P Value
Intercept		7.6875	-	-	-	-	-
A:A	-0.375	-0.1875	-	-	-	-	-
B:B	-0.5	-0.25	-	-	-	-	-
C:C*	2	1	-	-	-	-	-
D:D*	4	2	-	-	-	-	-
E:E	0.875	0.4375	-	-	-	-	-
F:F	0.75	0.375	-	-	-	-	-
A • B	0.25	0.125	-	-	-	-	-
A • C	0	0	-	-	-	-	-
A • D	-0.75	-0.375	-	-	-	-	-
A • E	0.375	0.1875	-	-	-	-	-
A • F	0.75	0.375	-	-	-	-	-
B • D	0.125	0.0625	-	-	-	-	-
B • F*	4.375	2.1875	-	-	-	-	-
A • B • D	-0.125	-0.0625	-	-	-	-	-
A • B • F	-0.125	-0.0625	-	-	-	-	-

*: Significant terms according to Lenth's method

Projection

Projection refers to the reduction of a fractional factorial design to a full factorial design by dropping out some of the factors of the design. Any fractional factorial design of resolution, R, can be reduced to complete factorial designs in any subset of $R-1$ factors. For example, consider the 2^{7-3}_{IV} design. The resolution of this design is four. Therefore, this design can be reduced to full factorial designs in any three $(4-1=3)$ of the original seven factors (by dropping the remaining four of factors). Further, a fractional factorial design can also be reduced to a full factorial design in any R of the original factors, as long as these R factors are not part of the generator in the defining relation. Again consider the 2^{7-3}_{IV} design. This design can be reduced to a full factorial design in four factors provided these four factors do not appear together as a generator in the defining relation. The complete defining relation for this design is:

$$I = \quad ABCE = BCDF = ACDG = ADEF = ABFG = BDEG = CEFG$$

Therefore, there are seven four factor combinations out of the 35 $(\binom{4}{7} = 35)$ possible four-factor combinations that are used as generators in the defining relation. The designs with the

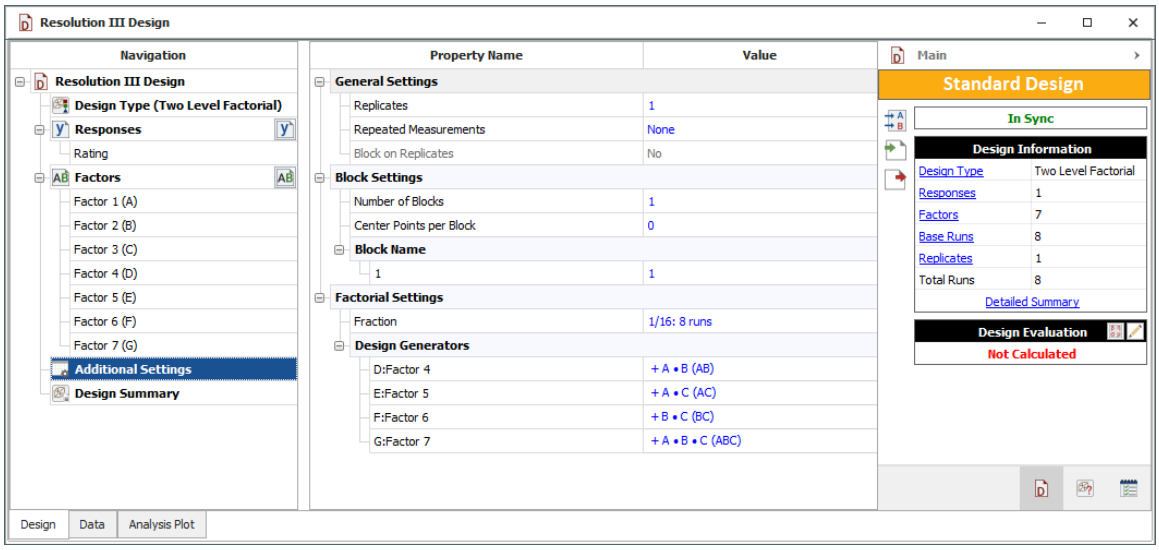
remaining 28 four factor combinations would be full factorial 16-run designs. For example, factors A, B, C and D do not occur as a generator in the defining relation of the 2^{7-3}_{IV} design. If the remaining factors, E, F and G, are dropped, the 2^{7-3}_{IV} design will reduce to a full factorial design in A, B, C and D.

Resolution III Designs

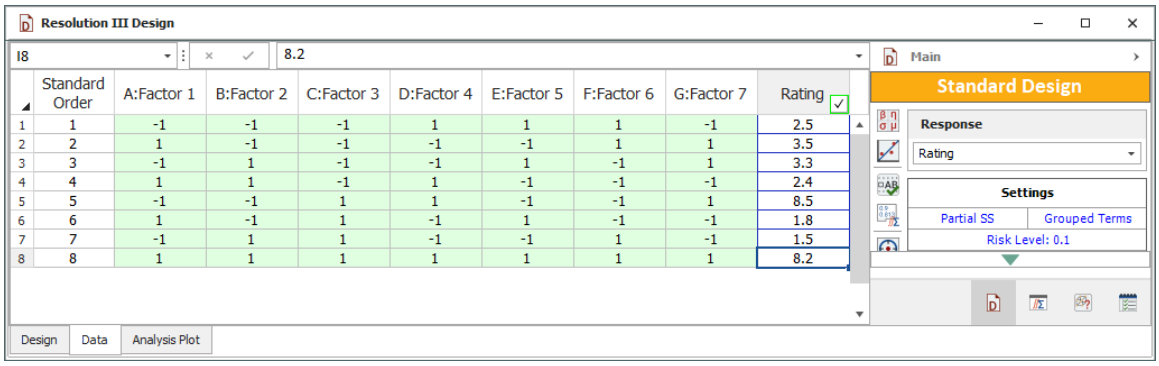
At times, the factors to be investigated in screening experiments are so large that even running a fractional factorial design is impractical. This can be partially solved by using resolution III fractional factorial designs in the cases where three factor and higher order interactions can be assumed to be unimportant. Resolution III designs, such as the 2^{3-1}_{III} design, can be used to estimate k main effects using just $k+1$ runs. In these designs, the main effects are aliased with two factor interactions. Once the results from these designs are obtained, and knowing that three factor and higher order interactions are unimportant, the experimenter can decide if there is a need to run a fold-over design to de-alias the main effects from the two factor interactions. Thus, the 2^{3-1}_{III} design can be used to investigate three factors in four runs, the 2^{7-4}_{III} design can be used to investigate seven factors in eight runs, the 2^{15-11}_{III} design can be used to investigate fifteen factors in sixteen runs and so on.

Example

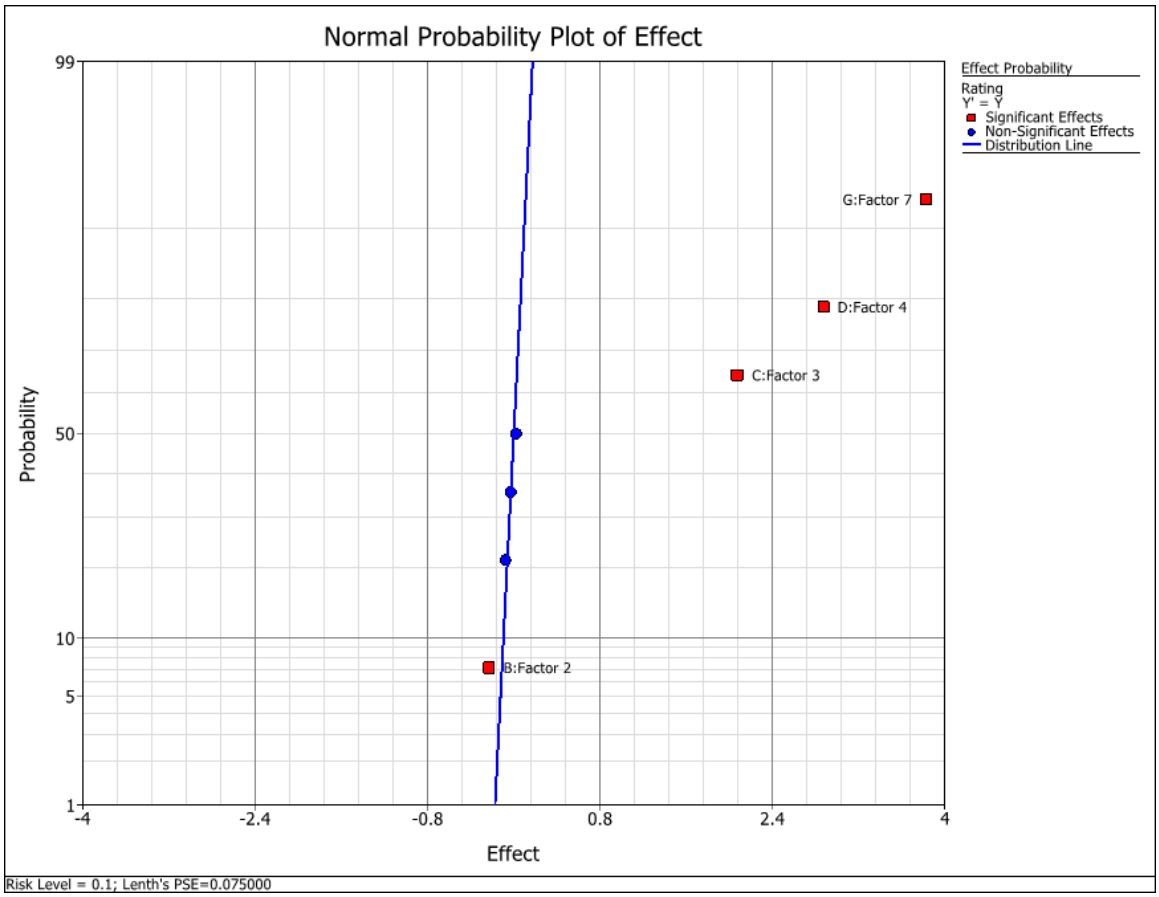
A baker wants to investigate the factors that most affect the taste of the cakes made in his bakery. He chooses to investigate seven factors, each at two levels: flour type (factor A), conditioner type (factor B), sugar quantity (factor C), egg quantity (factor D), preservative type (factor E), bake time (factor F) and bake temperature (factor G). The baker expects most of these factors and all higher order interactions to be inactive. On the basis of this, he decides to run a screening experiment using a 2^{7-4}_{III} design that requires just 8 runs. The cakes are rated on a scale of 1 to 10. The design properties for the 2^{7-4}_{III} design (with generators $D = AB$, $E = AC$, $F = BC$ and $G = ABC$) are shown in the following figure.



The resulting design along with the rating of the cakes corresponding to each run is shown in the following figure.



The normal probability plot of effects for the unreplicated design shows main effects B, C, D and G to be significant, as shown in the next figure.



However, for this design, the following alias relations exist for the main effects:

$$
\begin{aligned}
A &= A + BD + CE + FG \\
B &= B + AD + CF + EG \\
C &= C + AE + BF + DG \\
D &= D + AB + CG + EF \\
E &= E + AC + BG + DF \\
F &= F + BC + AG + DE \\
G &= G + CD + BE + AF
\end{aligned}
$$

Based on the alias structure, three separate possible conclusions can be drawn. It can be concluded that effect CD is active instead of G so that effects C, D and their interaction, CD, are the significant effects. Another conclusion can be that effect DG is active instead of C so that effects D, G and their interaction, DG, are significant. Yet another conclusion can be that effects C, G and their interaction, CG, are significant. To accurately discover the active effects, the baker decides to a run a fold-over of the present design and base his conclusions on the effect values calculated once results from both the designs are available.

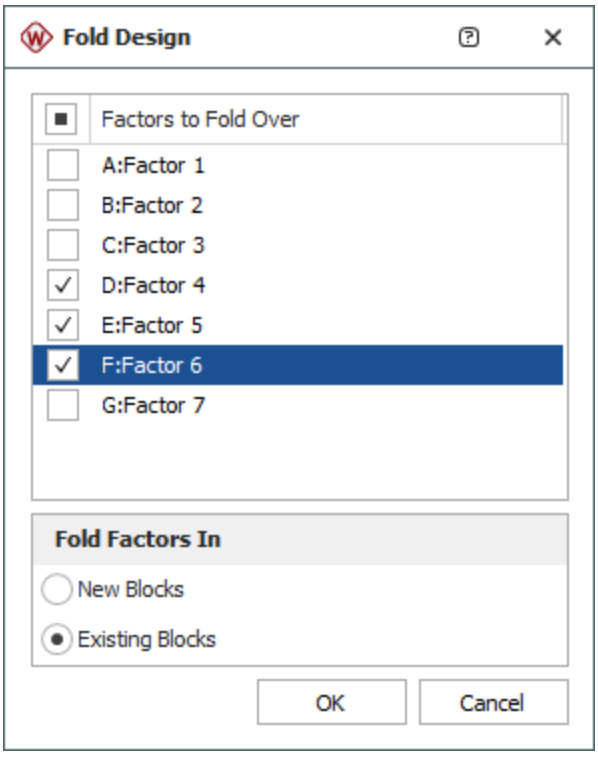
The present design is shown next.

Regression Information							
Term	Effect	Coefficient	Standard Error	Low Confidence	High Confidence	T Value	P Value
Intercept		3.9625	-	-	-	-	-
A:Factor 1	0.025	0.0125	-	-	-	-	-
B:Factor 2*	-0.225	-0.1125	-	-	-	-	-
C:Factor 3*	2.075	1.0375	-	-	-	-	-
D:Factor 4*	2.875	1.4375	-	-	-	-	-
E:Factor 5	-0.025	-0.0125	-	-	-	-	-
F:Factor 6	-0.075	-0.0375	-	-	-	-	-
G:Factor 7*	3.825	1.9125	-	-	-	-	-

Using the alias relations, the effects obtained from the DOE folio for the present design can be expressed as:

$$
\begin{aligned}
Effect_A &= 0.025 = A + BD + CE + FG \\
Effect_B &= -0.225 = B + AD + CF + EG \\
Effect_C &= 2.075 = C + AE + BF + DG \\
Effect_D &= 2.875 = D + AB + CG + EF \\
Effect_E &= -0.025 = E + AC + BG + DF \\
Effect_F &= -0.075 = F + BC + AG + DE \\
Effect_G &= 3.825 = G + CD + BE + AF
\end{aligned}
$$

The fold-over design for the experiment is obtained by reversing the signs of the columns D, E, and F. In a DOE folio, you can fold over a design using the following window.



The resulting design and the corresponding response values obtained are shown in the following figures.

A9 | 9

	Standard Order	A:Factor 1	B:Factor 2	C:Factor 3	D:Factor 4	E:Factor 5	F:Factor 6	G:Factor 7	Rating
1	1	-1	-1	-1	1	1	1	-1	2.5
2	2	1	-1	-1	-1	-1	1	1	3.5
3	3	-1	1	-1	-1	1	-1	1	3.3
4	4	1	1	-1	1	-1	-1	-1	2.4
5	5	-1	-1	1	1	-1	-1	1	8.5
6	6	1	-1	1	-1	1	-1	-1	1.8
7	7	-1	1	1	-1	-1	1	-1	1.5
8	8	1	1	1	1	1	1	1	8.2
9	9	-1	-1	-1	-1	-1	-1	-1	3.1
10	10	1	-1	-1	1	1	-1	1	2.7
11	11	-1	1	-1	1	-1	1	1	2.2
12	12	1	1	-1	-1	1	1	-1	3.4
13	13	-1	-1	1	-1	1	1	1	1.3
14	14	1	-1	1	1	-1	1	-1	8.4
15	15	-1	1	1	1	1	-1	-1	7.7
16	16	1	1	1	-1	-1	-1	1	1.2

Analysis Summary

Cut, Copy, Paste Special, Paste (Clipboard); Quick Print, Print, Print Preview, Send to Excel (Common); Apply Settings (Settings); Cancel (ReliaSoft)

Available Report Items
- Analysis Results
  - ANOVA Table
  - Regression Table
  - Regression Equation
- Additional Results
  - Alias Structure
  - Alias Summary
  - Var/Cov Matrix
  - Diagnostic Information
  - Least Squares Means

Analysis History

Response: Rating
Folio: Resolution III Design
Base Design Type: Two Level Factorial

Analyzed on 7/26/2024 1:39:39 PM

Regression Table

Regression Information

Term	Effect	Coefficient	Standard Error	Low Confidence	High Confidence	T Value	P Value
Intercept		3.85625	0.062276	3.738262	3.974238	61.921545	7.523471E-11
A:Factor 1	0.1875	0.09375	0.062276	-0.024238	0.211738	1.505386	0.175943
B:Factor 2	-0.2375	-0.11875	0.062276	-0.236738	-0.000762	-1.906822	0.098216
C:Factor 3	1.9375	0.96875	0.062276	0.850762	1.086738	15.555655	0.000001
D:Factor 4	2.9375	1.46875	0.062276	1.350762	1.586738	23.584381	6.258985E-8
E:Factor 5	0.0125	0.00625	0.062276	-0.111738	0.124238	0.100359	0.922873
F:Factor 6	0.0375	0.01875	0.062276	-0.099238	0.136738	0.301077	0.772102
G:Factor 7	0.0125	0.00625	0.062276	-0.111738	0.124238	0.100359	0.922873
C • D	3.8125	1.90625	0.062276	1.788262	2.024238	30.609516	1.024996E-8

Comparing the absolute values of the effects, the active effects are B, C, D and the interaction CD. Therefore, the most important factors affecting the taste of the cakes in the present case are sugar quantity, egg quantity and their interaction.

Alias Matrix

In Half-fraction designs and Quarter and Smaller Fraction Designs, the alias structure for fractional factorial designs was obtained using the defining relation. However, this method of obtaining the alias structure is not very efficient when the alias structure is very complex or when

partial aliasing is involved. One of the ways to obtain the alias structure for any design, regardless of its complexity, is to use the alias matrix. The alias matrix for a design is calculated using $(X_1'X_1)^{-1}X_1'X_2$ where X_1 is the portion of the design matrix, X, that contains the effects for which the aliases need to be calculated, and X_2 contains the remaining columns of the design matrix, other than those included in X_1.

To illustrate the use of the alias matrix, consider the design matrix for the 2^{4-1}_{IV} design (using the defining relation $I = ABCD$) shown next:

I	A	B	AB	C	AC	BC	D	AD	BD	ABC	ABD	CD	ACD	BCD	$ABCD$
1	-1	-1	1	-1	1	1	-1	1	1	-1	-1	1	-1	-1	1
1	1	-1	-1	-1	-1	1	1	1	-1	1	-1	-1	-1	1	1
1	-1	1	-1	-1	1	-1	1	-1	1	1	-1	-1	1	-1	1
1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1
1	-1	-1	1	1	-1	-1	1	-1	-1	1	1	1	-1	-1	1
1	1	-1	-1	1	1	-1	-1	-1	1	-1	1	-1	-1	1	1
1	-1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

The alias structure for this design can be obtained by defining X_1 using eight columns since the 2^{4-1}_{IV} design estimates eight effects. If the first eight columns of X are used then X_1 is:

I	A	B	AB	C	AC	BC	D
1	-1	-1	1	-1	1	1	-1
1	1	-1	-1	-1	-1	1	1
1	-1	1	-1	-1	1	-1	1
1	1	1	1	-1	-1	-1	-1
1	-1	-1	1	1	-1	-1	1
1	1	-1	-1	1	1	-1	-1
1	-1	1	-1	1	-1	1	-1
1	1	1	1	1	1	1	1

X_2 is obtained using the remaining columns as:

AD	BD	ABC	ABD	CD	ACD	BCD	$ABCD$
1	1	-1	1	-1	-1	-1	1
1	-1	1	-1	-1	-1	1	1
-1	1	1	-1	1	1	-1	1
-1	-1	-1	1	1	1	1	1
-1	-1	1	1	-1	-1	-1	1
-1	1	-1	-1	-1	-1	1	1
1	-1	-1	-1	1	1	-1	1
1	1	1	1	1	1	1	1

Then the alias matrix $(X_1'X_1)^{-1}X_1'X_2$ is:

	AD	BD	ABC	ABD	CD	ACD	BCD	$ABCD$
I	0	0	0	0	0	0	0	1
A	0	0	0	0	0	0	1	0
B	0	0	0	0	0	1	0	0
AB	0	0	0	0	1	0	0	0
C	0	0	0	1	0	0	0	0
AC	0	1	0	0	0	0	0	0
BC	1	0	0	0	0	0	0	0
D	0	0	1	0	0	0	0	0

The alias relations can be easily obtained by observing the alias matrix as:

$$
\begin{aligned}
I &= ABCD \\
A &= BCD \\
B &= ACD \\
AB &= CD \\
C &= ABD \\
AC &= BD \\
BC &= AD \\
D &= ABC
\end{aligned}
$$

Highly Fractional Factorial Designs

IN THIS CHAPTER

Plackett-Burman Designs **215**

Taguchi's Orthogonal Arrays **216**
Example 220
Preferred Columns in Taguchi OA 223

This chapter discusses factorial designs that are commonly used in designed experiments, but are not necessarily limited to two level factors. These designs are the Plackett-Burman designs and Taguchi's orthogonal arrays.

Plackett-Burman Designs

It was mentioned in Two Level Factorial Experiments that resolution III designs can be used as highly fractional designs to investigate k main effects using $k+1$ runs (provided that three factor and higher order interaction effects are not important to the experimenter). A limitation with these designs is that all runs in these designs have to be a power of 2. The valid runs for these designs are 4, 8, 16, 32, etc. Therefore, the next design after the 2^{3-1}_{III} design with 4 runs is the 2^{7-4}_{III} design with 8 runs, and the design after this is the 2^{15-11}_{III} design with 32 runs and so on, as shown in the next table.

Design	Runs	Type	Effects Investigated
2^{3-1}_{III}	4	Two level Fractional Factorial	Up to 3 main effects
2^{7-4}_{III}	8	Two level Fractional Factorial	Up to 7 main effects
12-run P-B	12	Plackett-Burman	Up to 11 main effects
2^{15-11}_{III}	16	Two level Fractional Factorial	Up to 15 main effects
20-run P-B	20	Plackett-Burman	Up to 19 main effects
24-run P-B	24	Plackett-Burman	Up to 23 main effects
28-run P-B	28	Plackett-Burman	Up to 27 main effects
2^{31-26}_{III}	32	Two level Fractional Factorial	Up to 31 main effects

Plackett-Burman designs solve this problem. These designs were proposed by R. L. Plackett and J.P. Burman (1946). They allow the estimation of k main effects using $k+1$ runs. In these designs, runs are a multiple of 4 (i.e., 4, 8, 12, 16, 20 and so on). When the runs are a power of 2, the designs correspond to the resolution III two factor fractional factorial designs. Although Plackett-Burman designs are all two level orthogonal designs, the alias structure for these designs is complicated when runs are not a power of 2.

As an example, consider the 12-run Plackett-Burman design shown in the figure below.

Run	Factors										
	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*J*	*K*	*L*
1	1	1	-1	1	1	1	-1	-1	-1	1	-1
2	-1	1	1	-1	1	1	1	-1	-1	-1	1
3	1	-1	1	1	-1	1	1	1	-1	-1	-1
4	-1	1	-1	1	1	-1	1	1	1	-1	-1
5	-1	-1	1	-1	1	1	-1	1	1	1	-1
6	-1	-1	-1	1	-1	1	1	-1	1	1	1
7	1	-1	-1	-1	1	-1	1	1	-1	1	1
8	1	1	-1	-1	-1	1	-1	1	1	-1	1
9	1	1	1	-1	-1	-1	1	-1	1	1	-1
10	-1	1	1	1	-1	-1	-1	1	-1	1	1
11	1	-1	1	1	1	-1	-1	-1	1	-1	1
12	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

If 11 main effects are to be estimated using this design, then each of these main effects is partially aliased with all other two factor interactions not containing that main effect. For example, the main effect A is partially aliased with all two factor interactions except AB, AC, AD, AE, AF, AG, AH, AJ, AK and AL. There are 45 such two factor interactions that are aliased with A.

$$A = A - \frac{1}{3}BC - \frac{1}{3}BD + \frac{1}{3}CD - \frac{1}{3}BE - \frac{1}{3}CE + \frac{1}{3}DE + \ldots$$
$$\ldots + \frac{1}{3}EL - \frac{1}{3}FL - \frac{1}{3}GL + \frac{1}{3}HL + \frac{1}{3}JL - \frac{1}{3}KL$$

Due to the complex aliasing, Plackett-Burman designs involving a large number of factors should be used with care. Some of the Plackett-Burman designs available in the DOE folio are included in Appendix C.

Taguchi's Orthogonal Arrays

Taguchi's orthogonal arrays are highly fractional orthogonal designs proposed by Dr. Genichi Taguchi, a Japanese industrialist. These designs can be used to estimate main effects using only a few experimental runs. These designs are not only applicable to two level factorial experiments; they can also investigate main effects when factors have more than two levels. Designs

are also available to investigate main effects for certain mixed level experiments where the factors included do not have the same number of levels. As in the case of Placket-Burman designs, these designs require the experimenter to assume that interaction effects are unimportant and can be ignored. A few of Taguchi's orthogonal arrays available in a DOE folio are included in Appendix D.

Some of Taguchi's arrays, with runs that are a power of 2, are similar to the corresponding 2^{k-f}_{III} designs. For example, consider the L4 array shown in figure (a) below. The L4 array is denoted as L4(2^3) in a DOE folio. L4 means the array requires 4 runs. 2^3 indicates that the design estimates up to three main effects at 2 levels each. The L4 array can be used to estimate three main effects using four runs provided that the two factor and three factor interactions can be ignored. Figure (b) below shows the 2^{3-1}_{III} design (defining relation $I = -ABC$) which also requires four runs and can be used to estimate up to three main effects, assuming that all two factor and three factor interactions are unimportant. A comparison between the two designs shows that the columns in the two designs are the same except for the arrangement of the columns. In figure (c) below, columns of the L4 array are marked with the name of the effect from the corresponding column of the 2^{3-1}_{III} design.

Run	Columns		
	1	2	3
1	1	1	1
2	1	2	2
3	2	1	2
4	2	2	1

(a)

Run	Factors		
	A	*B*	*C*
(1)	-1	-1	-1
ab	1	1	-1
ac	1	-1	1
bc	-1	1	1

(b)

Run	Columns		
	1	2	3
1	1	1	1
2	1	2	2
3	2	1	2
4	2	2	1
	C	*B*	*A*

(c)

Similarly, consider the L8(2^7) array shown in figure (a) below. This design can be used to estimate up to seven main effects using eight runs. This array is again similar to the 2^{7-4}_{III} design shown in figure (b) below, except that the aliasing between the columns of the two designs differs in sign for some of the columns (see figure (c)).

Run	Columns						
	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2

(a)

Run	Factors						
	A	*B*	*C*	*D=AB*	*E=AC*	*F=BC*	*G=ABC*
def	-1	-1	-1	1	1	1	-1
afg	1	-1	-1	-1	-1	1	1
beg	-1	1	-1	-1	1	-1	1
abd	1	1	-1	1	-1	-1	-1
cdg	-1	-1	1	1	-1	-1	1
ace	1	-1	1	-1	1	-1	-1
bcf	-1	1	1	-1	-1	1	-1
abcdefg	1	1	1	1	1	1	1

(b)

Run	Columns						
	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2
	C	*B*	*-BC*	*A*	*-AC*	*-AB*	*ABC*

(c)

The L8 array can also be used as a full factorial three factor experiment design in the same way as a 2^3 design. However, the orthogonal arrays should be used carefully in such cases, taking into consideration the alias relationships between the columns of the array. For the L8 array, figure (c) above shows that the third column of the array is the product of the first two columns. If

the L8 array is used as a two level full factorial design in the place of a 2^3 design, and if the main effects are assigned to the first three columns, the main effect assigned to the third column will be aliased with the two factor interaction of the first two main effects. The proper assignment of the main effects to the columns of the L8 array requires the experimenter to assign the three main effects to the first, second and fourth columns. These columns are sometimes referred to as the *preferred columns* for the L8 array. To know the preferred columns for any of the orthogonal arrays, the alias relationships between the array columns must be known. The alias relations between the main effects and two factor interactions of the columns for the L8 array are shown in the next table.

	1	2	3	4	5	6	7
1		3	2	5	4	7	6
2			1	6	7	4	5
3				7	6	5	4
4					1	2	3
5						3	2
6							1
7							

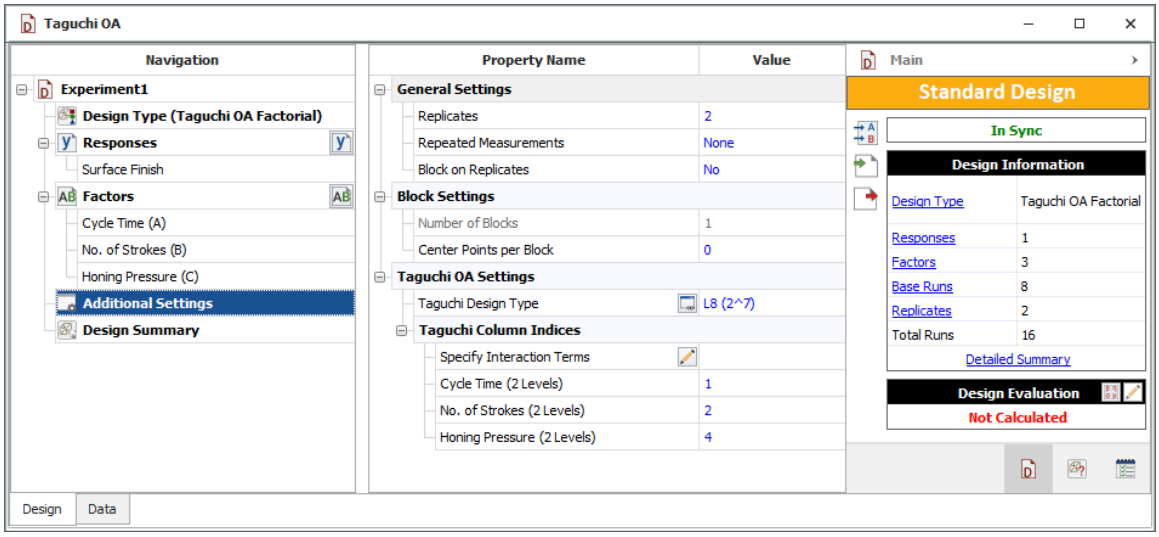
The cell value in any (i, j) cell of the table gives the column number of the two factor interaction for the ith row and jth column. For example, to know which column is confounded with the interaction of the first and second columns, look at the value in the $(1, 2)$ cell. The value of 3 indicates that the third column is the same as the product of the first and second columns. The alias relations for some of Taguchi's orthogonal arrays are available in Appendix E.

Example

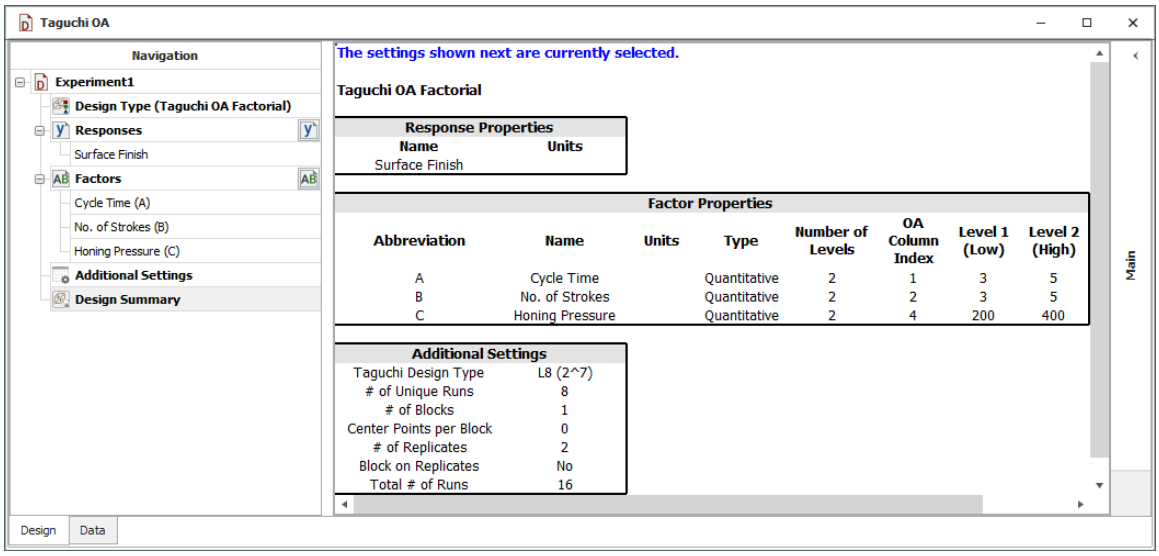
Recall the experiment to investigate factors affecting the surface finish of automobile brake drums discussed in Two Level Factorial Experiments. The three factors investigated in the experiment were honing pressure (factor A), number of strokes (factor B) and cycle time (factor C). Assume that you used Taguchi's L8 orthogonal array to investigate the three factors instead of the 2^3 design that was used in Two Level Factorial Experiments. Based on the discussion in the previous section, the preferred columns for the L8 array are the first, second and fourth columns. Therefore, the three factors should be assigned to these columns. The three factors are assigned to these columns based on the figure (c) above, so that you can easily compare results obtained from the L8 array to the ones included in Two Level Factorial Experiments. Based on this assignment, the L8 array for the two replicates, along with the respective response values, should be as shown in the third table. Note that to run the experiment using the L8 array, you would use only the first, the second and the fourth column to set the three factors.

A:Cycle Time	B:No. of Strokes	C:AB	D:Honing Press	E:AD	F:BD	G:ABD	Surface Finish
1	1	1	1	1	1	1	90
1	1	1	2	2	2	2	90
1	2	2	1	1	2	2	85
1	2	2	2	2	1	1	80
2	1	2	1	2	1	2	90
2	1	2	2	1	2	1	105
2	2	1	1	2	2	1	75
2	2	1	2	1	1	2	90
1	1	1	1	1	1	1	86
1	1	1	2	2	2	2	90
1	2	2	1	1	2	2	80
1	2	2	2	2	1	1	82
2	1	2	1	2	1	2	85
2	1	2	2	1	2	1	95
2	2	1	1	2	2	1	80
2	2	1	2	1	1	2	80

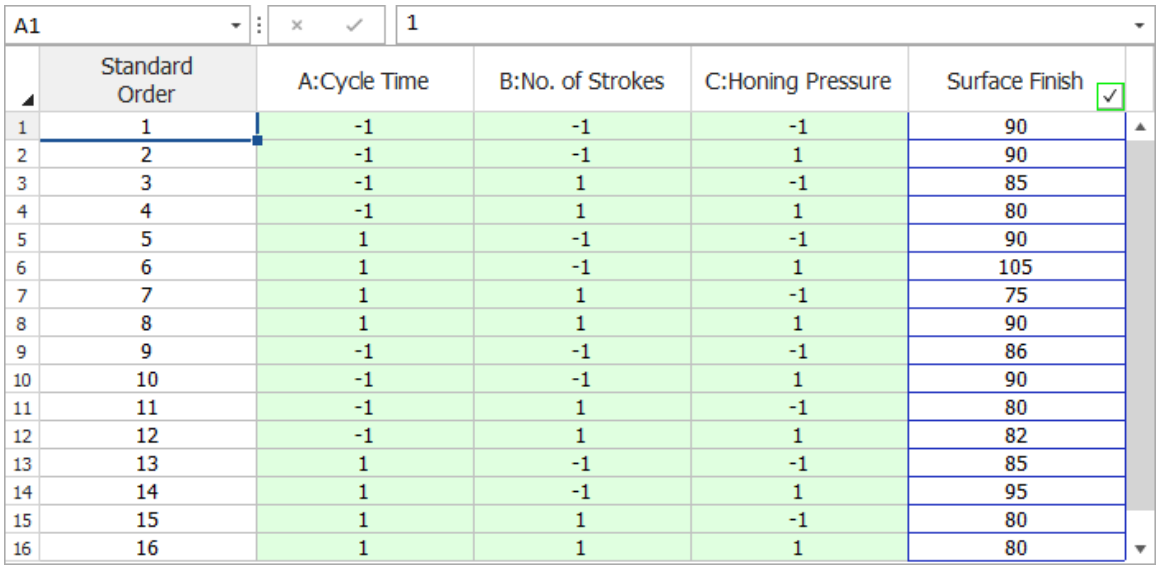
The experiment design for this example can be set using the properties shown in the figure below.



Note that for this design, the factor properties are set up as shown in the design summary.



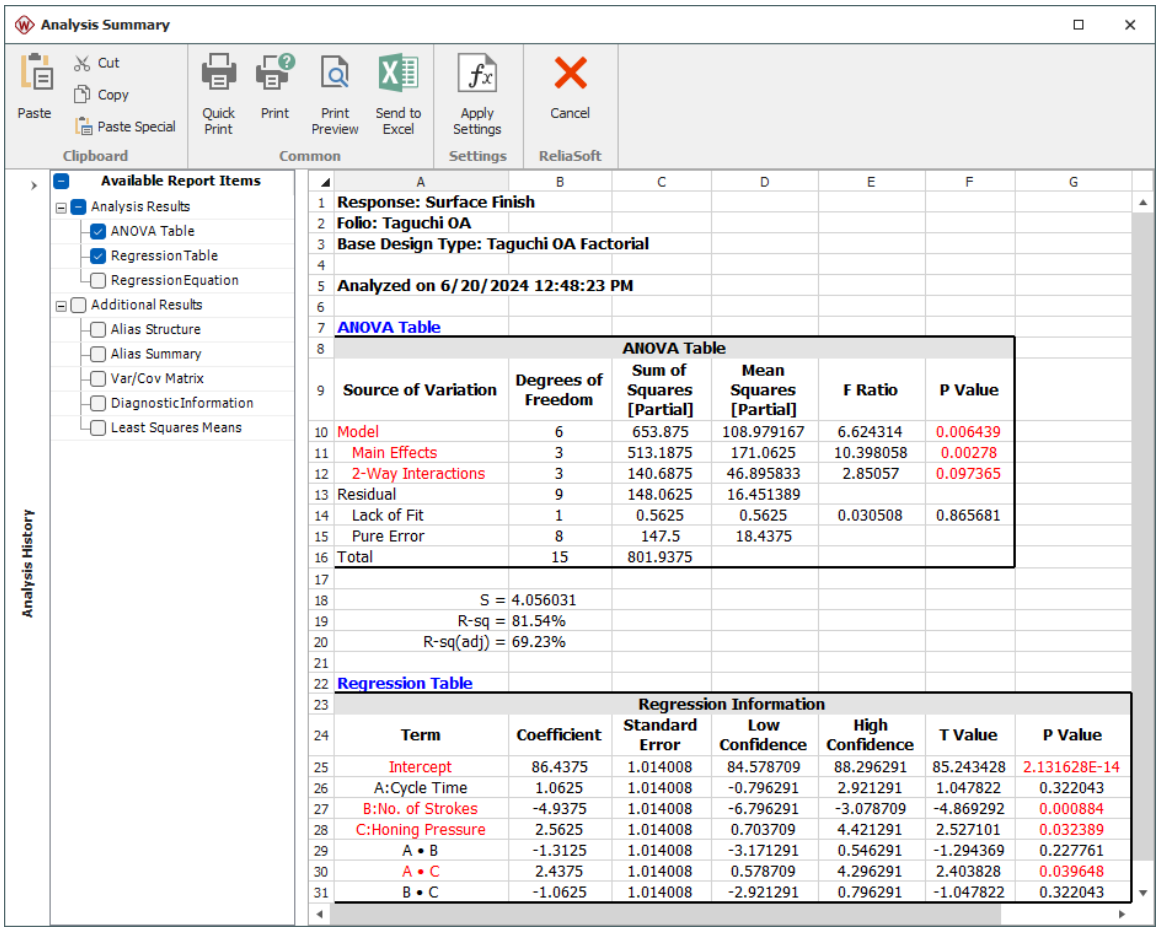
The resulting design along with the response values is shown in the figure below.



A1 | 1

	Standard Order	A:Cycle Time	B:No. of Strokes	C:Honing Pressure	Surface Finish
1	1	-1	-1	-1	90
2	2	-1	-1	1	90
3	3	-1	1	-1	85
4	4	-1	1	1	80
5	5	1	-1	-1	90
6	6	1	-1	1	105
7	7	1	1	-1	75
8	8	1	1	1	90
9	9	-1	-1	-1	86
10	10	-1	-1	1	90
11	11	-1	1	-1	80
12	12	-1	1	1	82
13	13	1	-1	-1	85
14	14	1	-1	1	95
15	15	1	1	-1	80
16	16	1	1	1	80

And the results from the DOE folio for the design are shown in the next figure.



The results identify honing pressure, number of strokes, and the interaction between honing pressure and cycle time to be significant effects. This is identical to the conclusion obtained from the 2^3 design used in Two Level Factorial Experiments.

Preferred Columns in Taguchi OA

One of the difficulties of using Taguchi OA is to assign factors to the appropriate columns of the array. For example, take a simple Taguchi OA L8(2^7), which can be used for experiments with up to 7 factors. If you have only 3 factors, which 3 columns in this array should be used? The DOE folio provides a simple utility to help users utilize Taguchi OA more effectively by assigning factors to the appropriate columns. Let's use Taguchi OA L8(2^7) as an example. The design table for this array is:

Run	Columns						
	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2

This is a fractional factorial design for 7 factors. For any fractional factorial design, the first thing we need to do is check its alias structure. In general, the alias structures for Taguchi OAs are very complicated. People usually use the following table to represent the alias relations between each factor. For the above orthogonal array, the alias table is:

1	2	3	4	5	6	7
2x3	1x3	1x2	1x5	1x4	1x7	1x6
4x5	4x6	4x7	2x6	2x7	2x4	2x5
6x7	5x7	5x6	3x7	3x6	3x5	3x4

In the above table, an Arabic number is used to represent a factor. For instance, "1" represents the factor assigned to the 1st column in the array. "2x3" represents the interaction effect of the two factors assigned to column 2 and 3. Each column in the above alias table lists all the 2-way interaction effects that are aliased with the main effect of the factor assigned to this column. For example, for the 1st column, the main effect of the factor assigned to it is aliased with interaction effects of 2x3, 4x5 and 6x7. If an experiment has only 3 factors and these 3 factors A, B and C are assigned to the first 3 columns of the above L8(2^7) array, then the design table will be:

Run	A (Column 1)	B (Column 2)	C (Column 3)
1	1	1	1
2	1	1	1
3	1	2	2
4	1	2	2
5	2	1	2
6	2	1	2

7	2	2	1
8	2	2	1

The alias structure for the above table is:

[I] = I - ABC
[A] = A - BC
[B] = B - AC
[C] = C - AB

This is a resolution 3 design. All the main effects are aliased with 2-way interactions. There are many ways to choose 3 columns from the 7 columns of L8(2^7). If the 3 factors are assigned to column 1, 2, and 4, then the design table is:

Run	A (Column 1)	B (Column 2)	C (Column 4)
1	1	1	1
2	1	1	1
3	1	2	2
4	1	2	2
5	2	1	2
6	2	1	2
7	2	2	1
8	2	2	1

For experiments using the above design table, all the effects will be alias free. Therefore, this design is much better than the previous one which used column 1, 2, and 3 of L8(2^7). Although both designs have the same number of runs, more information can be obtained from this design since it is alias free.

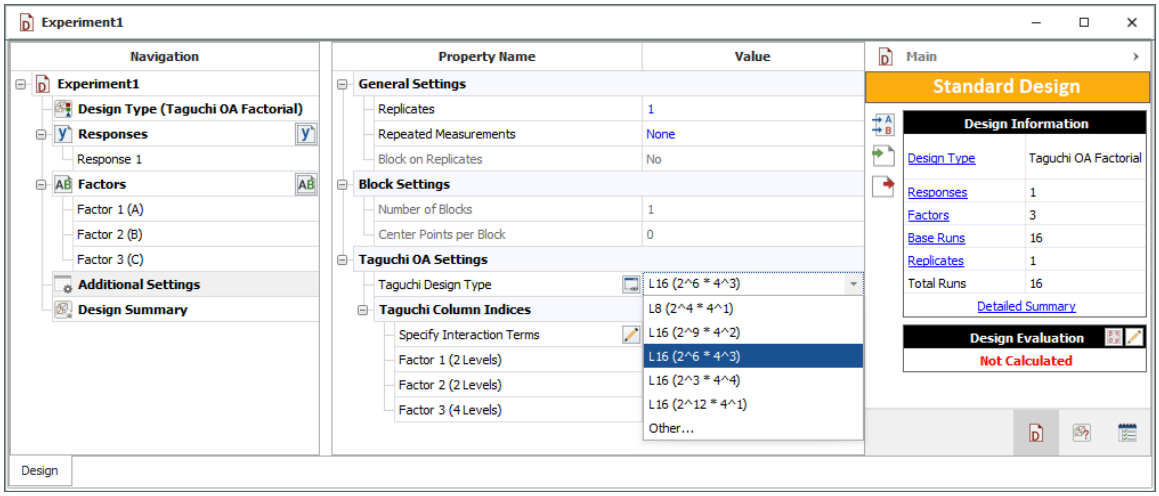
Clearly, it is very important to assign factors to the right columns when applying Taguchi OA. The DOE folio can help users automatically choose the right columns when the number of factors is less than the number of columns in a Taguchi OA. The selection is based on the specified model terms by users. Let's use an example to explain this.

Example

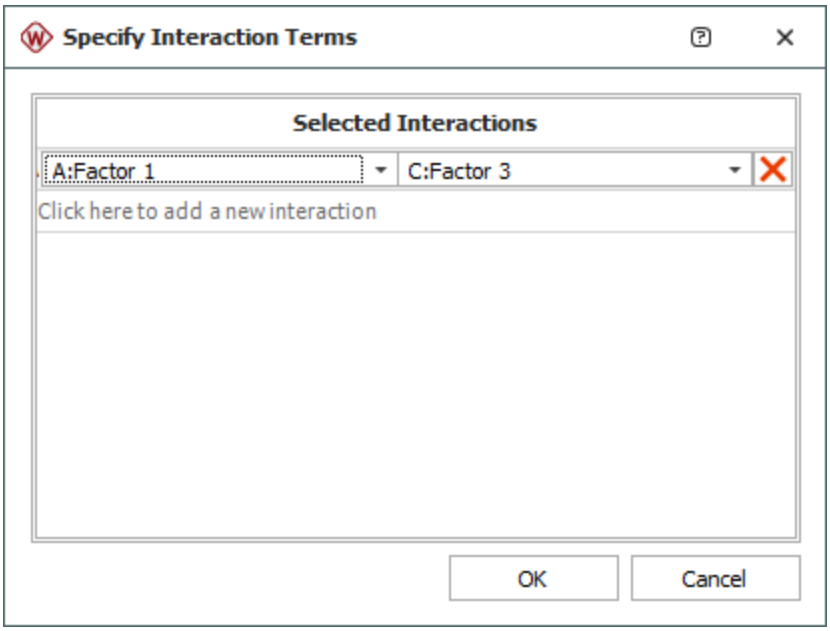
Design an experiment with 3 qualitative factors. Factors A and B have 2 levels; factor C has 4 levels. The experimenters are interested in all the main effects and the interaction effect AC.

Based on this requirement, Taguchi OA L16(2^6*4^3) can be used since it can handle both 2 level and 4 level factors. It has 9 columns. The first 6 columns are used for 2 level factors and the last 3 columns are used for 4 level factors. We need to assign factor A and B to two of the first 6 columns, and assign factor C to one of the last 3 columns.

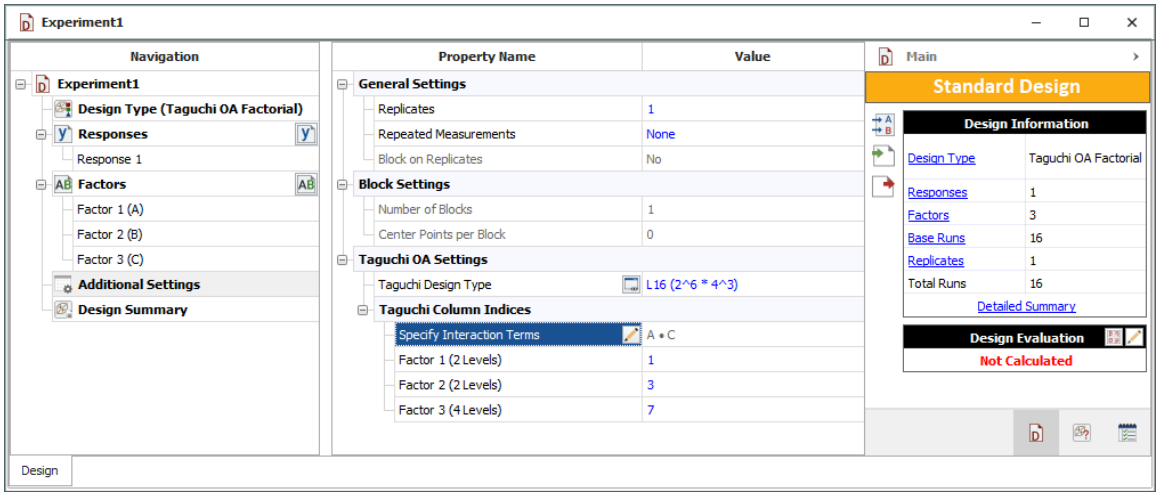
In a Weibull++ DOE folio, we can choose L16(2^6*4^3) in the Additional Settings.



Click **Specify Interaction Terms** to specify the interaction terms that are of interest to the experimenters.



Based on the specified interaction effects, the DOE folio will assign each factor to the appropriate column. In this case, they are column 1, 3, and 7 as shown below.

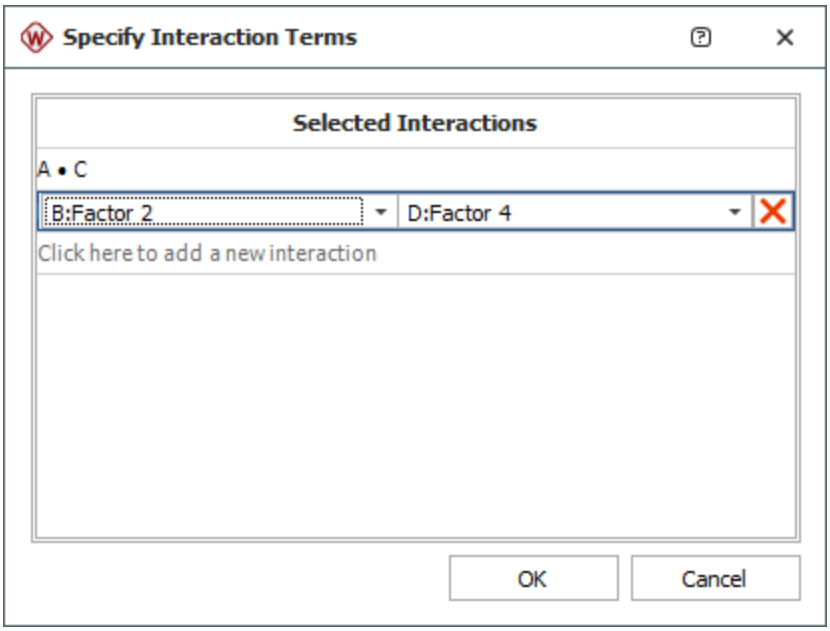


However, for a given Taguchi OA, it may not be possible to estimate all the specified interaction terms. If not all the requirements can be satisfied, the DOE folio will assign factors to columns that result in the least number of aliased effects. In this case, users should either use another Taguchi OA or other design types. The following example is one of these cases.

Example

Design an experiment for a test with 4 qualitative factors. Factors A and B have 2 levels; C and D have 4 levels. We are interested in all the main effects and the interaction effects AC and BD.

Assume again we want to use Taguchi OA L16(2^6*4^3). Click **Specify Interaction Terms** as shown above to specify the interaction effects that you want to estimate in the experiment.



When you click **OK**, you will see a message warning that some of the specified interaction effects are aliased with main effects. These means that it is not possible to clearly estimate all the main effects and the specified interaction effects AC and BD.

This can be explained by checking the alias table of an L16(2^6*4^3) design as given below.

1	2	3	4	5	6	7	8	9
2x7	1x7	1x9	1x5	1x4	1x8	1x2	1x6	1x3
3x9	3x6	2x6	2x8	2x9	2x3	1x8	1x7	1x7
4x5	4x8	4x7	3x7	3x8	4x9	1x9	1x9	1x8
6x8	5x9	5x8	6x9	6x7	5x7	2x8	2x4	2x5
7x8	7x8	7x8	7x8	7x8	7x8	2x9	2x7	2x7
7x9	7x9	7x9	7x9	7x9	7x9	3x8	2x9	2x8
8x9	8x9	8x9	8x9	8x9	8x9	3x9	3x5	3x7
						4x8	3x7	3x8
						4x9	3x9	4x6
						5x6	4x7	4x7
						5x8	4x9	4x8
						5x9	5x7	5x7
						6x8	5x9	5x8

						6x9	6x7	6x7
						8x9	6x9	6x8
							7x9	7x8

From this table, we can see that it is impossible to clearly estimate both AC and BD. Factors C and D can only be assigned to the last three columns since they are 4 level factors. Assume we assign factor C to column 7 and factor D to column 8. Factor B (2 level) will be in one of the columns from 1 to 6. Therefore, effect BD will be one of the effects highlighted in the table, where the first term of the interaction is between 1 and 6 and the last term is 8 (i.e., factor D).

The above alias table shows that factor C is aliased with one of those highlighted effects. Thus, no matter which of the first six columns is assigned to factor B, the main effect C will be aliased with interaction effect BD. This is also true if C is assigned to column 8 or 9. Therefore, if L16(2^6*4^3) is used, there is no way to clearly estimate all the main effects and the interaction effects AC and BD. Another Taguchi OA or other design types such as a general full factorial design should be used. A more efficient way might be to create an optimal custom design that can clearly estimate all the specified terms. For more detail, please refer to the chapter on Optimal Custom Designs.

Response Surface Methods for Optimization

IN THIS CHAPTER

Method of Steepest Ascent **231**
Example 232

Path of Steepest Ascent **235**

RSM Designs **239**
Central Composite Designs 240
Box-Behnken Designs 243

Analysis of the Second Order Model **249**

Multiple Responses **254**
Desirability Functions 255

The experiment designs mentioned in Two Level Factorial Experiments and Highly Fractional Factorial Designs help the experimenter identify factors that affect the response. Once the important factors have been identified, the next step is to determine the settings for these factors that result in the optimum value of the response. The optimum value of the response may either be a maximum value or a minimum value, depending upon the product or process in question. For example, if the response in an experiment is the yield from a chemical process, then the objective might be to find the settings of the factors affecting the yield so that the yield is maximized. On the other hand, if the response in an experiment is the number of defects, then the goal would be to find the factor settings that minimize the number of defects. Methodologies that help the experimenter reach the goal of optimum response are referred to as *response surface methods*. These methods are exclusively used to examine the "surface," or the relationship between the response and the factors affecting the response. Regression models are used for the analysis of the response, as the focus now is on the nature of the relationship between the response and the factors, rather than identification of the important factors.

Response surface methods usually involve the following steps:

1. The experimenter needs to move from the present operating conditions to the vicinity of the operating conditions where the response is optimum. This is done using the *method of steepest ascent* in the case of maximizing the response. The same method can be used to minimize the response and is then referred to as the *method of steepest descent*.

2. Once in the vicinity of the optimum response the experimenter needs to fit a more elaborate model between the response and the factors. Special experiment designs, referred to as *RSM designs*, are used to accomplish this. The fitted model is used to arrive at the best operating conditions that result in either a maximum or minimum response.

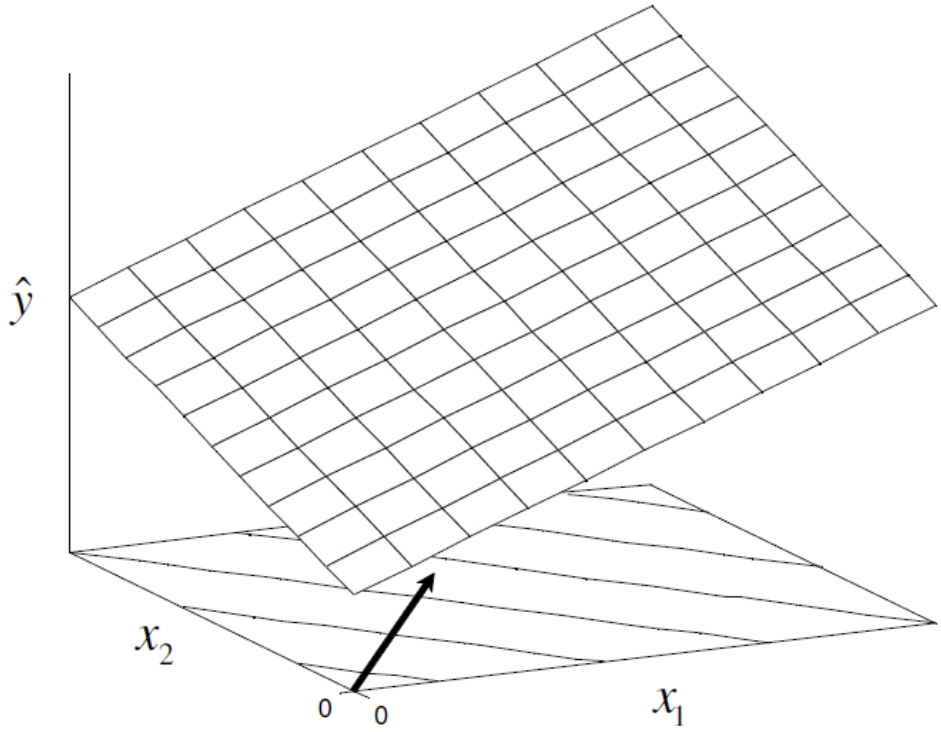
3. It is possible that a number of responses may have to be optimized at the same time. For example, an experimenter may want to maximize strength, while keeping the number of defects to a minimum. The optimum settings for each of the responses in such cases may lead to conflicting settings for the factors. A balanced setting has to be found that gives the most appropriate values for all the responses. Desirability functions are useful in these cases.

Method of Steepest Ascent

The first step in obtaining the optimum response settings, after the important factors have been identified, is to explore the region around the current operating conditions to decide what direction needs to be taken to move towards the optimum region. Usually, a first order regression model (containing just the main effects and no interaction terms) is sufficient at the current operating conditions because the operating conditions are normally far from the optimum response settings. The experimenter needs to move from the current operating conditions to the optimum region in the most efficient way by using the minimum number of experiments. This is done using the method of steepest ascent. In this method, the contour plot of the first order model is used to decide the settings for the next experiment, in order to move towards the optimum conditions. Consider a process where the response has been found to be a function of two factors. To explore the region around the current operating conditions, the experimenter fits the following first order model between the response and the two factors:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

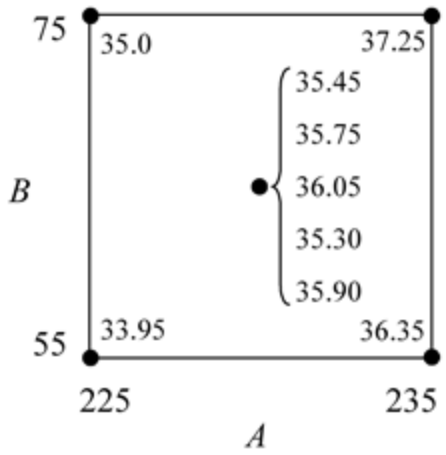
The response surface plot for the model, along with the contours, is shown in the figure below. It can be seen in the figure that in order to maximize the response, the most efficient direction in which to move the experiment is along the line perpendicular to the contours. This line, also referred to as the path of steepest ascent, is the line along which the rate of increase of the response is maximum. The steps along this line to move towards the optimum region are proportional to the regression coefficients, β_j, of the fitted first order model.



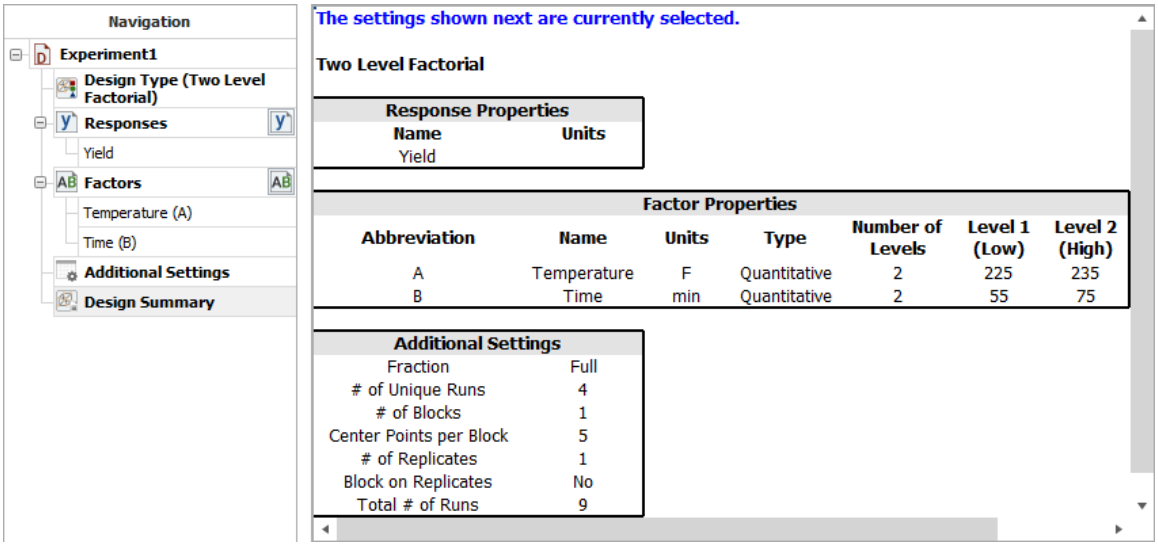
Experiments are conducted along each step of the path of steepest ascent until an increase in the response is not seen. Then, a new first order model is fit at the region of the maximum response. If the first order model shows a lack of fit, then this indicates that the experimenter has reached the vicinity of the optimum. RSM designs are then used explore the region thoroughly and obtain the point of the maximum response. If the first order model does not show a lack of fit, then a new path of steepest ascent is determined and the process is repeated.

Example

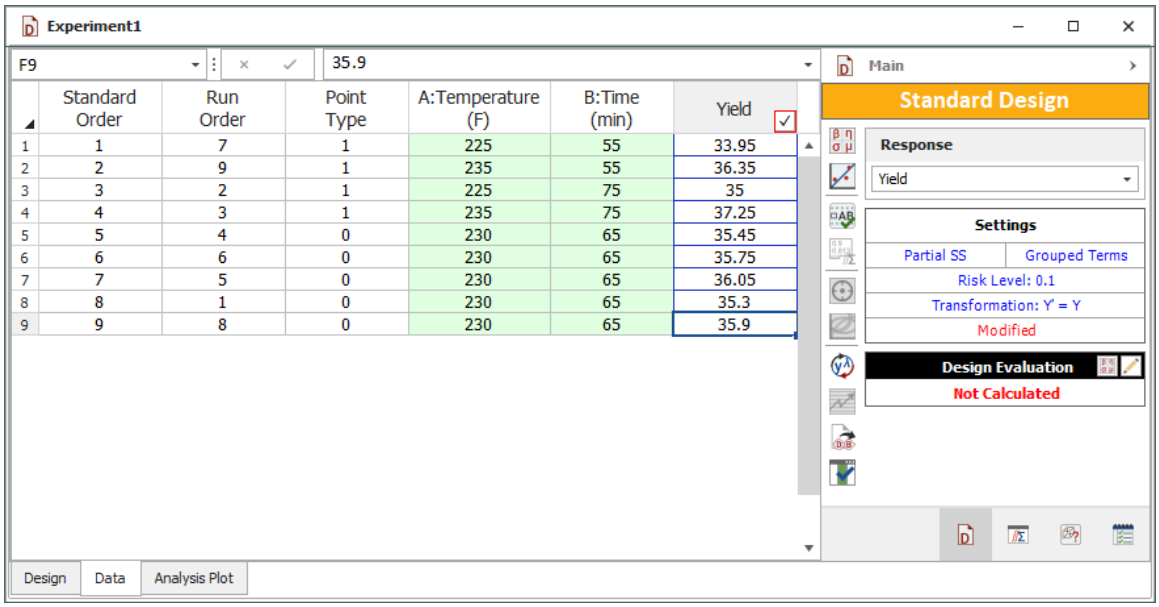
The yield from a chemical process is found to be affected by two factors: reaction temperature and reaction time. The current reaction temperature is 230 F and the reaction time is 65 minutes. The experimenter wants to determine the settings of the two factors such that maximum yield can be obtained from the process. To explore the region around the current operating conditions, the experimenter decides to use a single replicate of the 2^2 design. The range of the factors for this design are chosen to be (225, 235) F for the reaction temperature and (55, 75) minutes for the reaction time. The unreplicated 2^2 design is also augmented with five runs at the center point to estimate the error sum of squares, SS_E, and check for model adequacy. The response values obtained for this design are shown next.



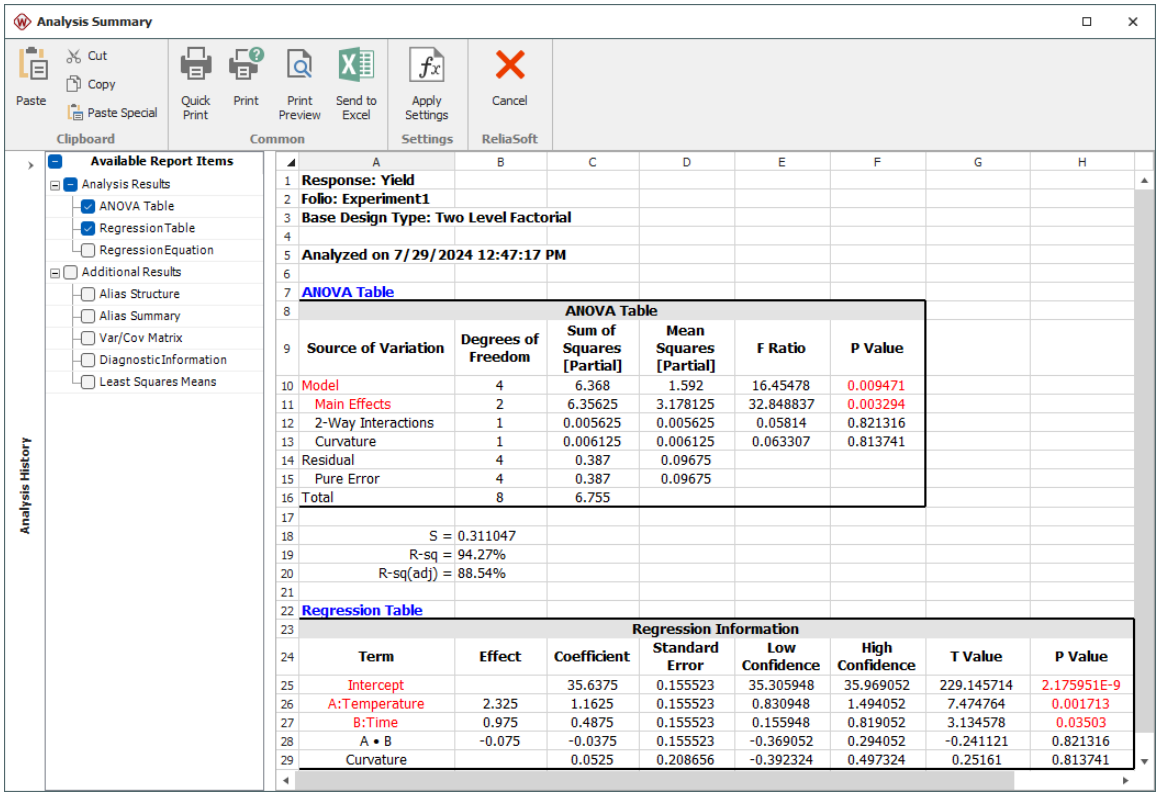
In a DOE folio, this design can be set up using the properties shown next.



The resulting design and the analysis results are shown next.



	Standard Order	Run Order	Point Type	A:Temperature (F)	B:Time (min)	Yield
1	1	7	1	225	55	33.95
2	2	9	1	235	55	36.35
3	3	2	1	225	75	35
4	4	3	1	235	75	37.25
5	5	4	0	230	65	35.45
6	6	6	0	230	65	35.75
7	7	5	0	230	65	36.05
8	8	1	0	230	65	35.3
9	9	8	0	230	65	35.9



Response: Yield
Folio: Experiment1
Base Design Type: Two Level Factorial

Analyzed on 7/29/2024 12:47:17 PM

ANOVA Table

ANOVA Table

Source of Variation	Degrees of Freedom	Sum of Squares [Partial]	Mean Squares [Partial]	F Ratio	P Value
Model	4	6.368	1.592	16.45478	0.009471
Main Effects	2	6.35625	3.178125	32.848837	0.003294
2-Way Interactions	1	0.005625	0.005625	0.05814	0.821316
Curvature	1	0.006125	0.006125	0.063307	0.813741
Residual	4	0.387	0.09675		
Pure Error	4	0.387	0.09675		
Total	8	6.755			

S = 0.311047
R-sq = 94.27%
R-sq(adj) = 88.54%

Regression Table

Regression Information

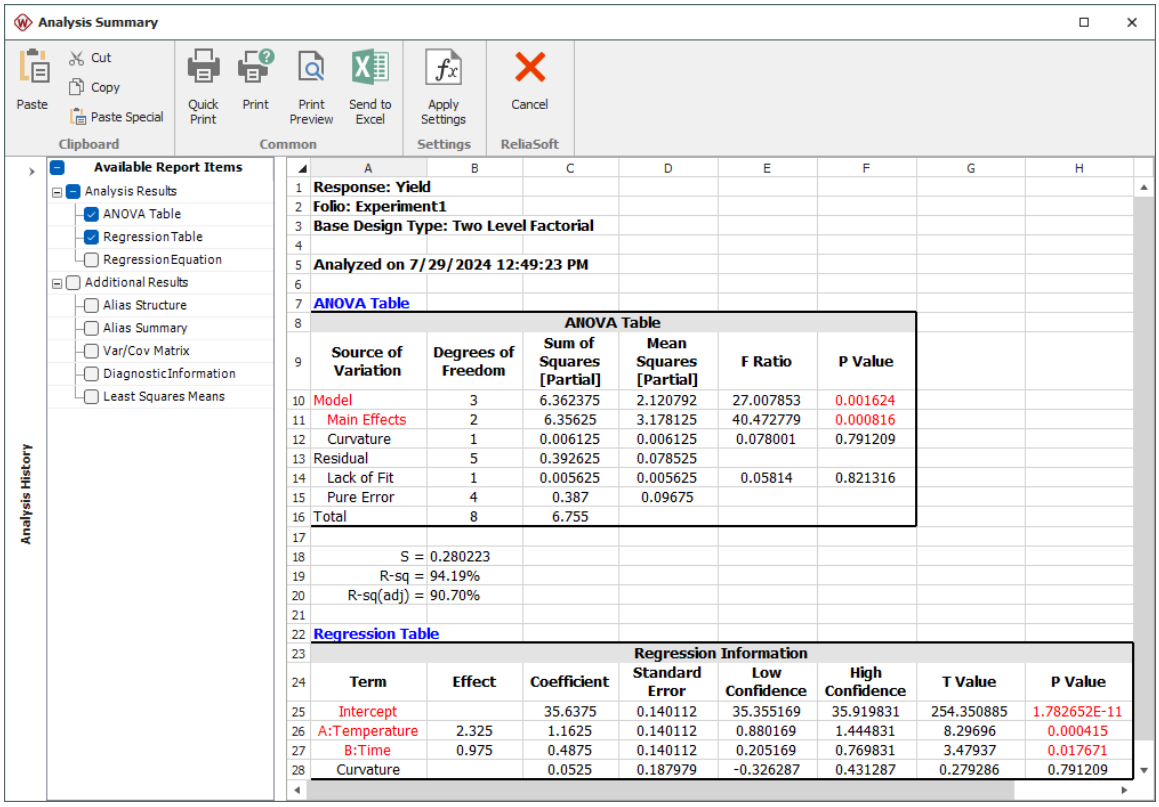
Term	Effect	Coefficient	Standard Error	Low Confidence	High Confidence	T Value	P Value
Intercept		35.6375	0.155523	35.305948	35.969052	229.145714	2.175951E-9
A:Temperature	2.325	1.1625	0.155523	0.830948	1.494052	7.474764	0.001713
B:Time	0.975	0.4875	0.155523	0.155948	0.819052	3.134578	0.03503
A • B	-0.075	-0.0375	0.155523	-0.369052	0.294052	-0.241121	0.821316
Curvature		0.0525	0.208656	-0.392324	0.497324	0.25161	0.813741

Note that the results shown are in terms of the coded values of the factors (taking -1 as the value of the lower settings for reaction temperature and reaction time and +1 as the value for the higher settings for these two factors). The results show that the factors, A (temperature) and B (time), affect the response significantly but their interaction does not affect the response. Therefore the interaction term can be dropped from the model for this experiment. The results also show that Curvature is not a significant factor. This indicates that the first order model is

adequate for the experiment at the current operating conditions. Using these two conclusions, the model for the current operating conditions, in terms the coded variables is:

$$\hat{y} = 36.6375 + 1.1635x_1 + 0.4875x_2$$

where $\hat{y}$ represents the yield and x_1 and x_2 are the predictor variables for the two factors, A and B, respectively. To further confirm the adequacy of the model of the equation given above, the experiment can be analyzed again after dropping the interaction term, AB. The results are shown next.



	A	B	C	D	E	F	G	H
1	**Response: Yield**							
2	**Folio: Experiment1**							
3	**Base Design Type: Two Level Factorial**							
4								
5	**Analyzed on 7/29/2024 12:49:23 PM**							
6								
7	**ANOVA Table**							

ANOVA Table

Source of Variation	Degrees of Freedom	Sum of Squares [Partial]	Mean Squares [Partial]	F Ratio	P Value
Model	3	6.362375	2.120792	27.007853	0.001624
Main Effects	2	6.35625	3.178125	40.472779	0.000816
Curvature	1	0.006125	0.006125	0.078001	0.791209
Residual	5	0.392625	0.078525		
Lack of Fit	1	0.005625	0.005625	0.05814	0.821316
Pure Error	4	0.387	0.09675		
Total	8	6.755			

S = 0.280223
R-sq = 94.19%
R-sq(adj) = 90.70%

Regression Table

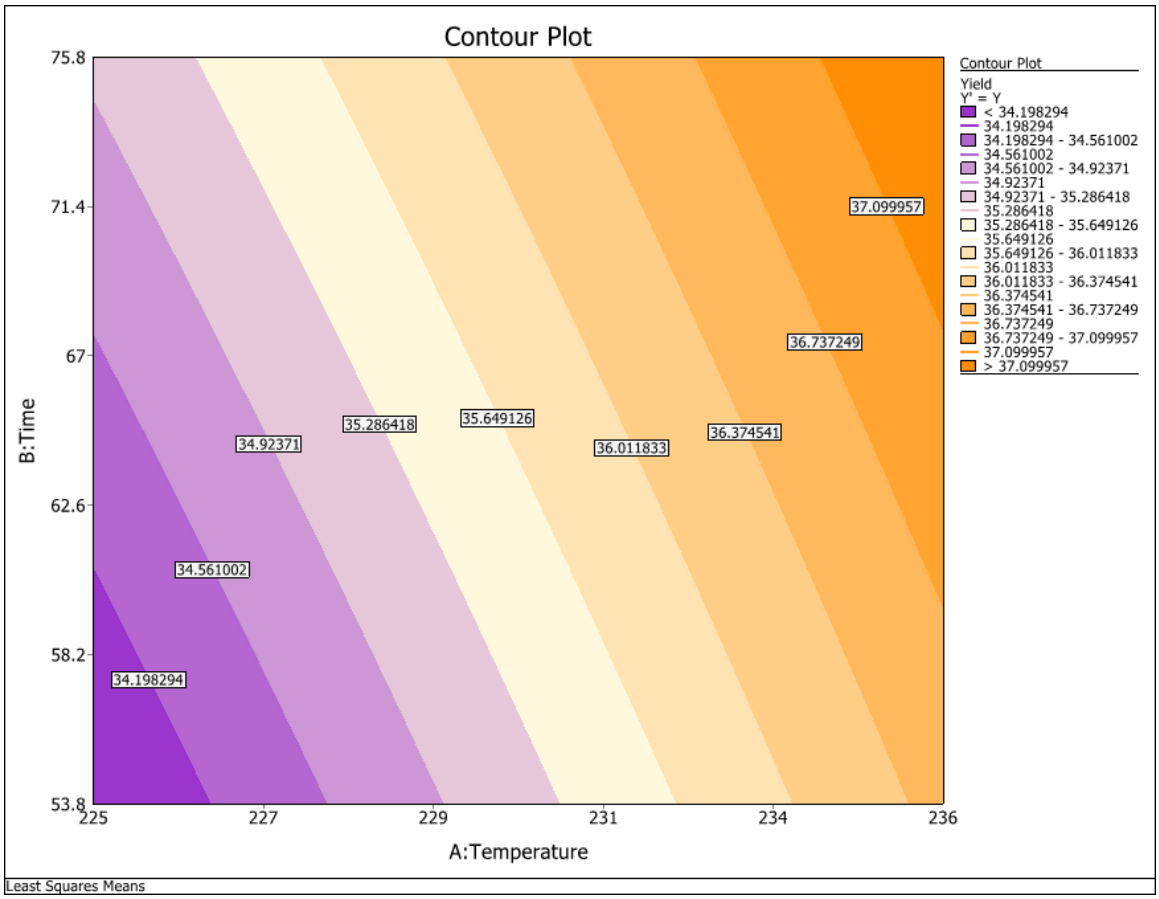
Regression Information

Term	Effect	Coefficient	Standard Error	Low Confidence	High Confidence	T Value	P Value
Intercept		35.6375	0.140112	35.355169	35.919831	254.350885	1.782652E-11
A:Temperature	2.325	1.1625	0.140112	0.880169	1.444831	8.29696	0.000415
B:Time	0.975	0.4875	0.140112	0.205169	0.769831	3.47937	0.017671
Curvature		0.0525	0.187979	-0.326287	0.431287	0.279286	0.791209

The results show that the lack-of-fit for this model (because of the deficiency created in the model by the absence of the interaction term) is not significant, confirming that the model is adequate.

Path of Steepest Ascent

The contour plot for the model used in the above example is shown next.



The regression coefficients for the model are $\hat{\beta}_1 = 1.1625$ and $\hat{\beta}_2 = 0.4875$. To move towards the optimum, the experimenter needs to move along the path of steepest ascent, which lies perpendicular to the contours. This path is the line through the center point of the current operating conditions ($x_1 = 0$, $x_2 = 0$) with a slope of $\hat{\beta}_2/\hat{\beta}_1$. Therefore, in terms of the coded variables, the experiment should be moved 1.1625 units in the x_1 direction for every 0.4875 units in the x_2 direction. To move along this path, the experimenter decides to use a step-size of 10 minutes for the reaction time, x_2. The coded value for this step size can be obtained as follows. Recall from Multiple Linear Regression Analysis that the relationship between coded and actual values is:

$$coded\ value = \frac{actual\ value - mean}{half\ of\ range}$$

or

$$coded\ value = \frac{step\ size\ in\ actual\ value}{half\ of\ range}$$

Thus, for a step-size of 10 minutes, the equivalent step size in coded value for x_2 is:

$$= \frac{10}{half\ of\ range\ of\ x_2}$$
$$= \frac{10}{(75-55)/2}$$
$$= 1$$

In terms of the coded variables, the path of steepest ascent requires a move of 1.1625 units in the x_1 direction for every 0.4875 units in the x_2 direction. The step-size for x_1, in terms of the coded value corresponding to any step-size in x_2, is:

$$= \frac{\hat{\beta}_1}{\hat{\beta}_2} \cdot (step\ size\ in\ x_2)$$

Therefore, the step-size for the reaction temperature, x_1, in terms of the coded variables is:

$$= \frac{\hat{\beta}_1}{\hat{\beta}_2} \cdot \Delta x_2$$
$$= \frac{1.1625}{0.4875} \cdot 1$$
$$= \frac{\hat{\beta}_1}{\hat{\beta}_2} \cdot \Delta x_2$$

This corresponds to a step of approximately 12 F for temperature in terms of the actual value as shown next:

$$\begin{aligned} step\ size\ in\ actual\ value &= (coded\ value) \cdot (half\ of\ the\ range\ of\ x_1) \\ &= 2.39 \cdot 10/2 \\ &= 11.95 \end{aligned}$$

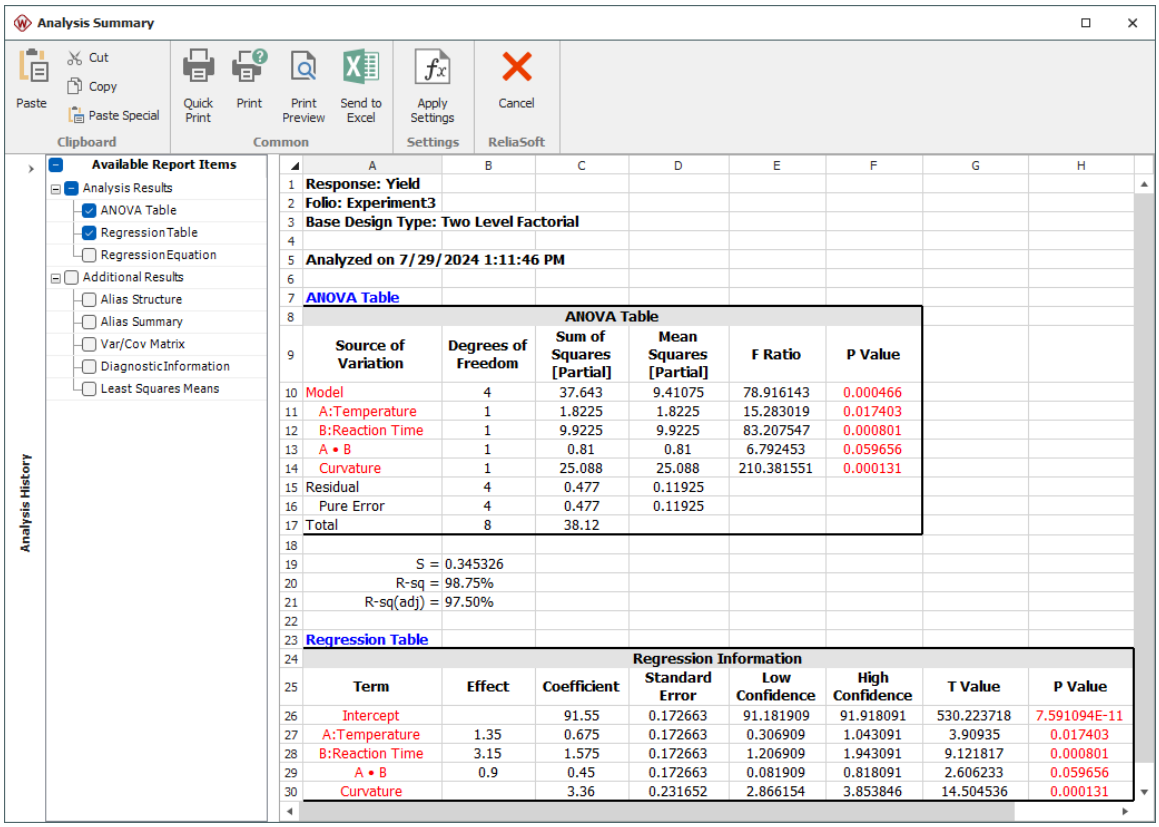
Using a step of 12 F and 10 minutes, the experimenter conducts experiments until no further increase is observed in the yield. The yield values at each step are shown in the table given next.

Step	Factor Levels				Yield
	Coded		Actual		
	A	*B*	*A*	*B*	
Current Operation	0	0	230	65	
1	2.4	1	242	75	36.50
2	4.8	2	254	85	39.35
3	7.2	3	266	95	45.65
4	9.6	4	278	105	49.55
5	12	5	290	115	55.70
6	14.4	6	302	125	64.25
7	16.8	7	314	135	72.50
8	19.2	8	326	145	80.60
9	21.6	9	338	155	91.40
10	**24**	**10**	**350**	**165**	**95.45**
11	26.4	11	362	175	89.30
12	28.8	12	374	185	87.65

The yield starts decreasing after the reaction temperature of 350 F and the reaction time of 165 minutes, indicating that this point may lie close to the optimum region. To analyze the vicinity of this point, a 2^2 design augmented by five center points is selected. The range of exploration is chosen to be 345 to 355 F for reaction temperature and 155 to 175 minutes for reaction time. The response values recorded are shown next.

Factor Levels				Response
Coded		Actual		
A	*B*	*A*	*B*	
-1	-1	345	155	89.75
1	-1	355	155	90.20
-1	1	345	175	92.00
1	1	355	175	94.25
0	0	350	165	94.85
0	0	350	165	95.45
0	0	350	165	95.00
0	0	350	165	94.55
0	0	350	165	94.70

The results for this design are shown next.



	A	B	C	D	E	F	G	H
1	**Response: Yield**							
2	**Folio: Experiment3**							
3	**Base Design Type: Two Level Factorial**							
4								
5	**Analyzed on 7/29/2024 1:11:46 PM**							
6								
7	**ANOVA Table**							
8	**ANOVA Table**							
9	**Source of Variation**	**Degrees of Freedom**	**Sum of Squares [Partial]**	**Mean Squares [Partial]**	**F Ratio**	**P Value**		
10	Model	4	37.643	9.41075	78.916143	0.000466		
11	A:Temperature	1	1.8225	1.8225	15.283019	0.017403		
12	B:Reaction Time	1	9.9225	9.9225	83.207547	0.000801		
13	A • B	1	0.81	0.81	6.792453	0.059656		
14	Curvature	1	25.088	25.088	210.381551	0.000131		
15	Residual	4	0.477	0.11925				
16	Pure Error	4	0.477	0.11925				
17	Total	8	38.12					
18								
19	S = 0.345326							
20	R-sq = 98.75%							
21	R-sq(adj) = 97.50%							
22								
23	**Regression Table**							
24	**Regression Information**							
25	**Term**	**Effect**	**Coefficient**	**Standard Error**	**Low Confidence**	**High Confidence**	**T Value**	**P Value**
26	Intercept		91.55	0.172663	91.181909	91.918091	530.223718	7.591094E-11
27	A:Temperature	1.35	0.675	0.172663	0.306909	1.043091	3.90935	0.017403
28	B:Reaction Time	3.15	1.575	0.172663	1.206909	1.943091	9.121817	0.000801
29	A • B	0.9	0.45	0.172663	0.081909	0.818091	2.606233	0.059656
30	Curvature		3.36	0.231652	2.866154	3.853846	14.504536	0.000131

In the results, Curvature is displayed as a significant factor. This indicates that the first order model is not adequate for this region of the experiment and a higher order model is required. As a result, the methodology of steepest ascent can no longer be used. The presence of curvature indicates that the experiment region may be close to the optimum. Special designs that allow the use of second order models are needed at this point.

RSM Designs

A second order model is generally used to approximate the response once it is realized that the experiment is close to the optimum response region where a first order model is no longer adequate. The second order model is usually sufficient for the optimum region, as third order and higher effects are seldom important. The second order regression model takes the following form for k factors:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \ldots + \beta_{kk} x_k^2 + \beta_{12} x_1 x_2 + \ldots + \beta_{k-1,k} x_{k-1} x_k + \epsilon$$

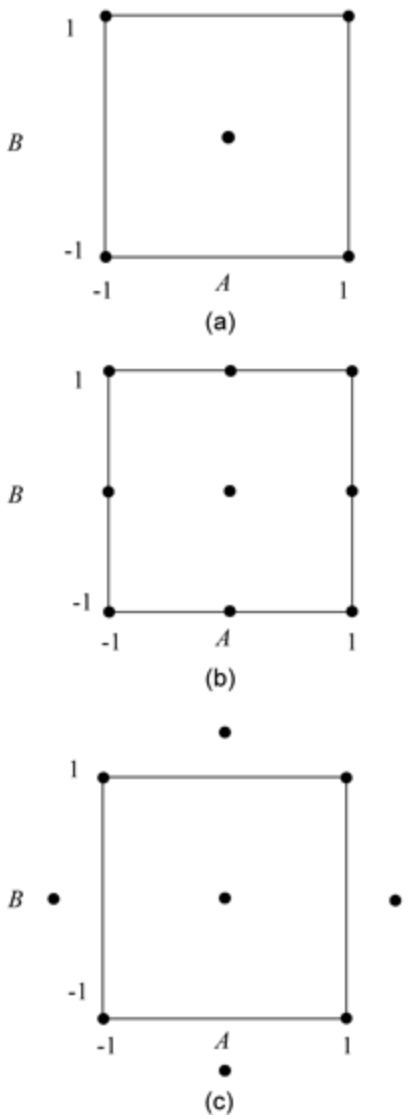
The model contains $p = (k+1)(k+2)/2$ regression parameters that include coefficients for main effects ($\beta_1, \beta_2 \ldots \beta_k$), coefficients for quadratic main effects ($\beta_{11}, \beta_{22} \ldots \beta_{kk}$) and coefficients for two factor interaction effects (β_{12}, β_{13} ... $\beta_{k-1,k}$). A full factorial design with all factors at three

levels would provide estimation of all the required regression parameters. However, full factorial three level designs are expensive to use as the number of runs increases rapidly with the number of factors. For example, a three factor full factorial design with each factor at three levels would require $3^3 = 27$ runs while a design with four factors would require $3^4 = 81$ runs. Additionally, these designs will estimate a number of higher order effects which are usually not of much importance to the experimenter. Therefore, for the purpose of analysis of response surfaces, special designs are used that help the experimenter fit the second order model to the response with the use of a minimum number of runs. Examples of these designs are the central composite and Box-Behnken designs.

Central Composite Designs

Central composite designs are two level full factorial (2^k) or fractional factorial (2^{k-f}) designs augmented by a number of center points and other chosen runs. These designs are such that they allow the estimation of all the regression parameters required to fit a second order model to a given response.

The simplest of the central composite designs can be used to fit a second order model to a response with two factors. The design consists of a 2^2 full factorial design augmented by a few runs at the center point (such a design is shown in figure (a) given below). A central composite design is obtained when runs at points $(-1, 0)$, $(1, 0)$, $(0, -1)$ and $(0, 1)$ are added to this design. These points are referred to as *axial points* or *star points* and represent runs where all but one of the factors are set at their mid-levels. The number of axial points in a central composite design having k factors is $2k$. The distance of the axial points from the center point is denoted by α and is always specified in terms of coded values. For example, the central composite design in figure (b) given below has $\alpha = 1$, while for the design of figure (c) $\alpha = 1.414$.



It can be noted that when $\alpha > 1$, each factor is run at five levels ($-\alpha$, -1, 0, 1 and α) instead of the three levels of -1, 0 and 1. The reason for running central composite designs with $\alpha > 1$ is to have a rotatable design, which is explained next.

Rotatability

A central composite design is said to be *rotatable* if the variance of any predicted value of the response, $\hat{y}_p$, for any level of the factors depends only on the distance of the point from the center of the design, regardless of the direction. In other words, a rotatable central composite design provides constant variance of the estimated response corresponding to all new observation points that are at the same distance from the center point of the design (in terms of the coded variables).

The variance of the predicted response at any point, x_p, is given as follows:

$$V[\hat{y}_p] = \sigma^2 x_p'(X'X)^{-1}x_p$$

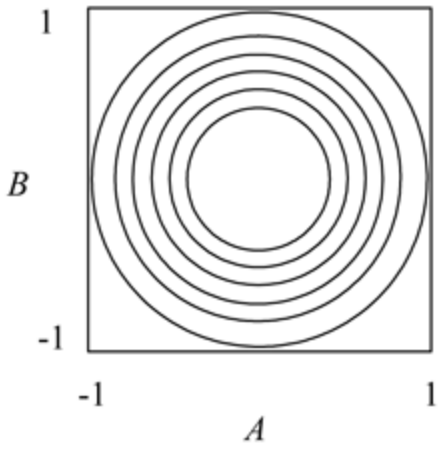
The contours of $V[\hat{y}_p]$ for the central composite design in figure (c) above are shown in the figure below. The contours are concentric circles indicating that the central composite design of figure (c) is rotatable. Rotatability is a desirable property because the experimenter does not have any prior information about the location of the optimum. Therefore, a design that provides equal precision of estimation in all directions would be preferred. Such a design will assure the experimenter that no matter what direction is taken to search for the optimum, he/she will be able to estimate the response value with equal precision. A central composite design is rotatable if the value of α for the design satisfies the following equation:

$$\alpha = \left[\frac{2^{k-f}(n_f)}{n_s}\right]^{1/4}$$

where n_f is the number of replicates of the runs in the original factorial design and n_s is the number of replicates of the runs at the axial points. For example, a central composite design with two factors, having a single replicate of the original factorial design, and a single replicate of all the axial points, would be rotatable for the following α value:

$$\begin{aligned}\alpha &= \left[\frac{2^{k-f}(n_f)}{n_s}\right]^{1/4} \\ &= \left[\frac{2^2(1)}{1}\right]^{1/4} \\ &= \sqrt{2} \\ &= 1.414\end{aligned}$$

Thus, a central composite design in two factors, having a single replicate of the original 2^2 design and axial points, and with $\alpha = 1.414$, is a rotatable design. This design is shown in figure (c) above.

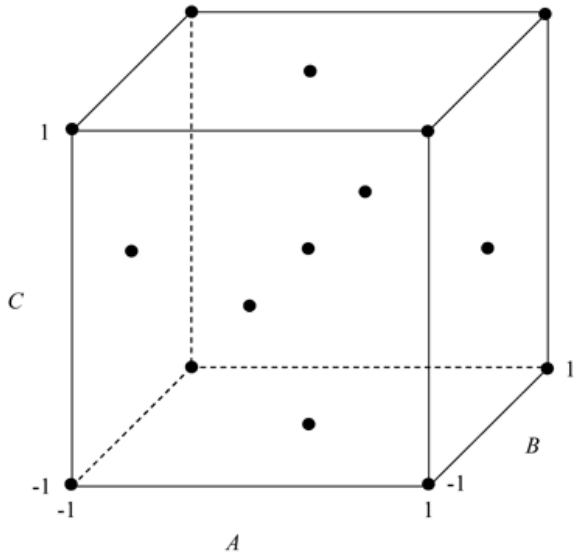


Spherical Design

A central composite design is said to be *spherical* if all factorial and axial points are at same distance from the center of the design. Spherical central composite designs are obtained by setting $\alpha = \sqrt{k}$. For example, the rotatable design in the figure above (c) is also a spherical design because for this design $\alpha = \sqrt{k} = \sqrt{2} = 1.414$.

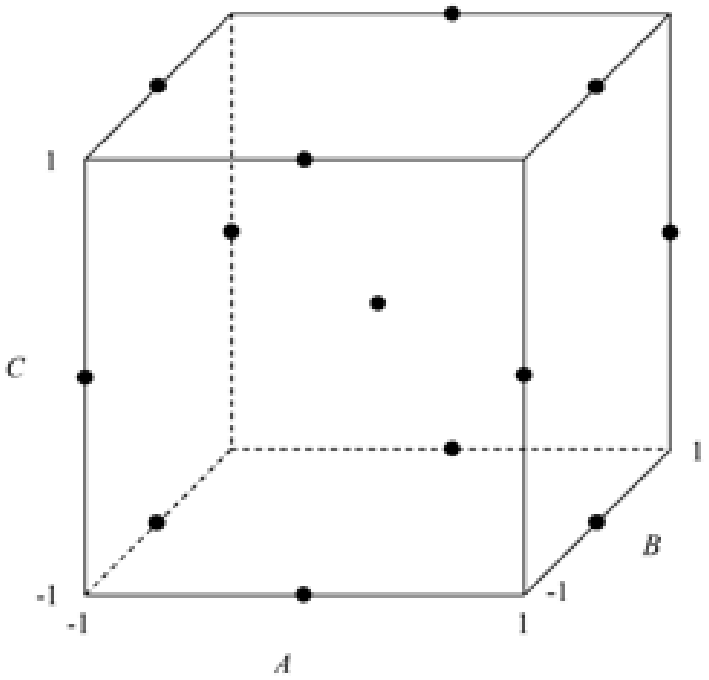
Face-centered Design

Central composite designs in which the axial points represent the mid levels for all but one of the factors are also referred to as *face-centered* central composite designs. For these designs, $\alpha = 1$ and all factors are run at three levels, which are -1, 0 and 1, in terms of the coded values (see the figure below).



Box-Behnken Designs

In Highly Fractional Factorial Designs, highly fractional designs introduced by Plackett and Burman were discussed. Plackett-Burman designs are used to estimate main effects in the case of two level fractional factorial experiments using very few runs. [G. E. P. Box and D. W. Behnken (1960)] introduced similar designs for three level factors that are widely used in response surface methods to fit second-order models to the response. The designs are referred to as Box-Behnken designs. The designs were developed by the combination of two level factorial designs with incomplete block designs. For example, the figure below shows the Box-Behnken design for three factors. The design is obtained by the combination of 2^2 design with a balanced incomplete block design having three treatments and three blocks (for details see [Box 1960, Montgomery 2001]).



(a)

Run	Factors		
	A	*B*	*C*
1	-1	-1	0
2	-1	1	0
3	1	-1	0
4	1	1	0
5	-1	0	-1
6	-1	0	1
7	1	0	-1
8	1	0	1
9	0	-1	-1
10	0	-1	1
11	0	1	-1
12	0	1	1
13	0	0	0

(b)

The advantages of Box-Behnken designs include the fact that they are all spherical designs and require factors to be run at only three levels. The designs are also rotatable or nearly rotatable. Some of these designs also provide orthogonal blocking. Thus, if there is a need to separate runs into blocks for the Box-Behnken design, then designs are available that allow blocks to be used

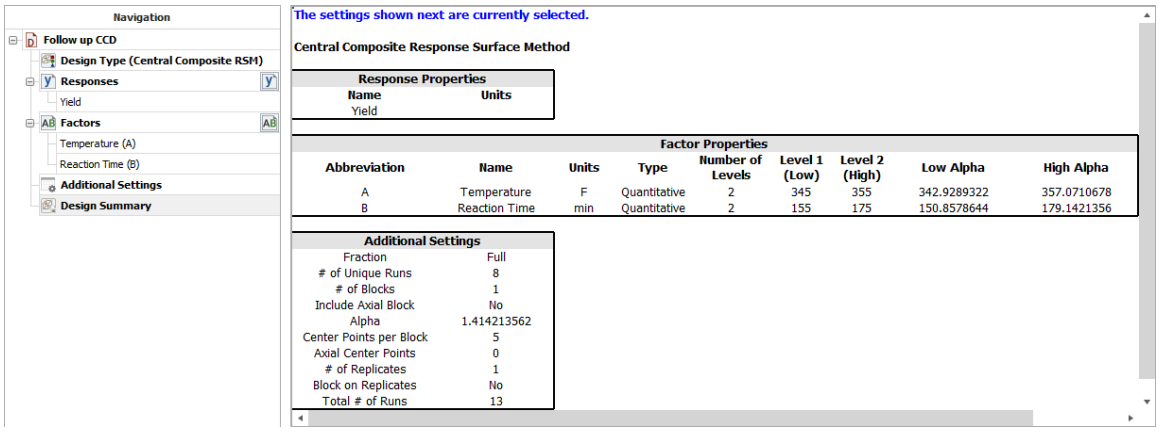
in such a way that the estimation of the regression parameters for the factor effects are not affected by the blocks. In other words, in these designs the block effects are orthogonal to the other factor effects. Yet another advantage of these designs is that there are no runs where all factors are at either the $+1$ or -1 levels. For example, in the figure below the representation of the Box-Behnken design for three factors clearly shows that there are no runs at the corner points. This could be advantageous when the corner points represent runs that are expensive or inconvenient because they lie at the end of the range of the factor levels. A few of the Box-Behnken designs available in a DOE folio are presented in Appendix F.

Example

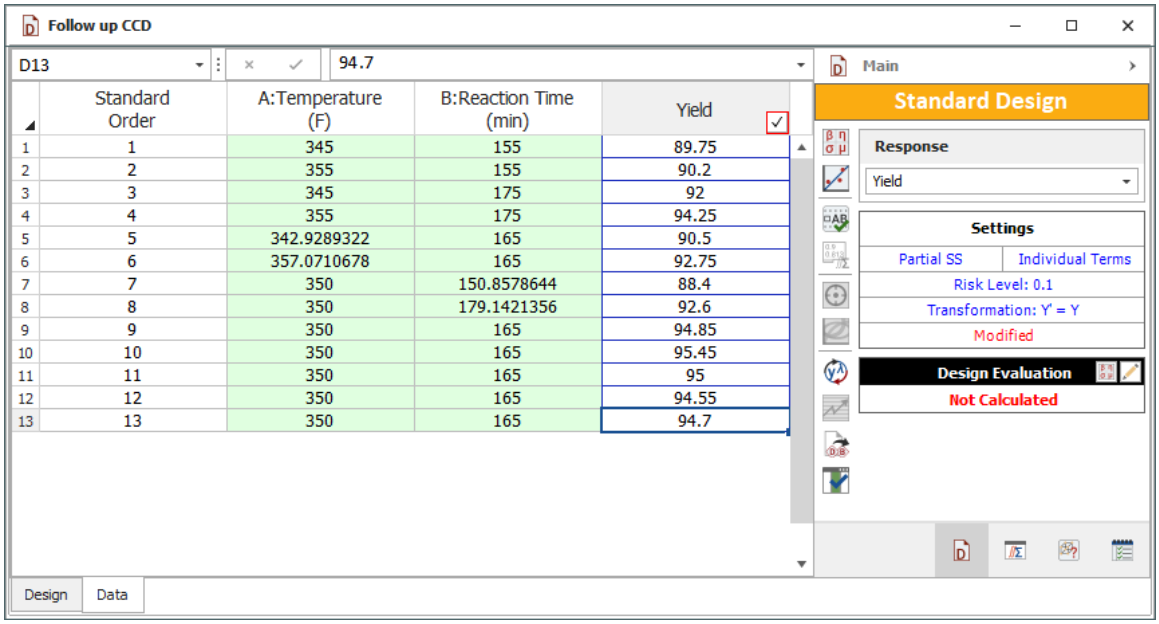
Continuing with the example in Method of Steepest Ascent, the first order model was found to be inadequate for the region near the optimum. Once the experimenter realized that the first order model was not adequate (for the region with a reaction temperature of 350 F and reaction time of 165 minutes), it was decided to augment the experiment with axial runs to be able to complete a central composite design and fit a second order model to the response. Notice the advantage of using a central composite design, as the experimenter only had to add the axial runs to the 2^2 design with center point runs, and did not have to begin a new experiment. The experimenter decided to use $\alpha = 1.4142$ to get a rotatable design. The obtained response values are shown in the figure below.

Factor Levels				Response
Coded		Actual		
A	*B*	*A*	*B*	
-1	-1	345	155	89.75
1	-1	355	155	90.20
-1	1	345	175	92.00
1	1	355	175	94.25
-1.414	0	342.9	165	90.50
1.414	0	357.1	165	92.75
0	-1.414	350	150.9	88.40
0	1.414	350	179.1	92.60
0	0	350	165	94.85
0	0	350	165	95.45
0	0	350	165	95.00
0	0	350	165	94.55
0	0	350	165	94.70

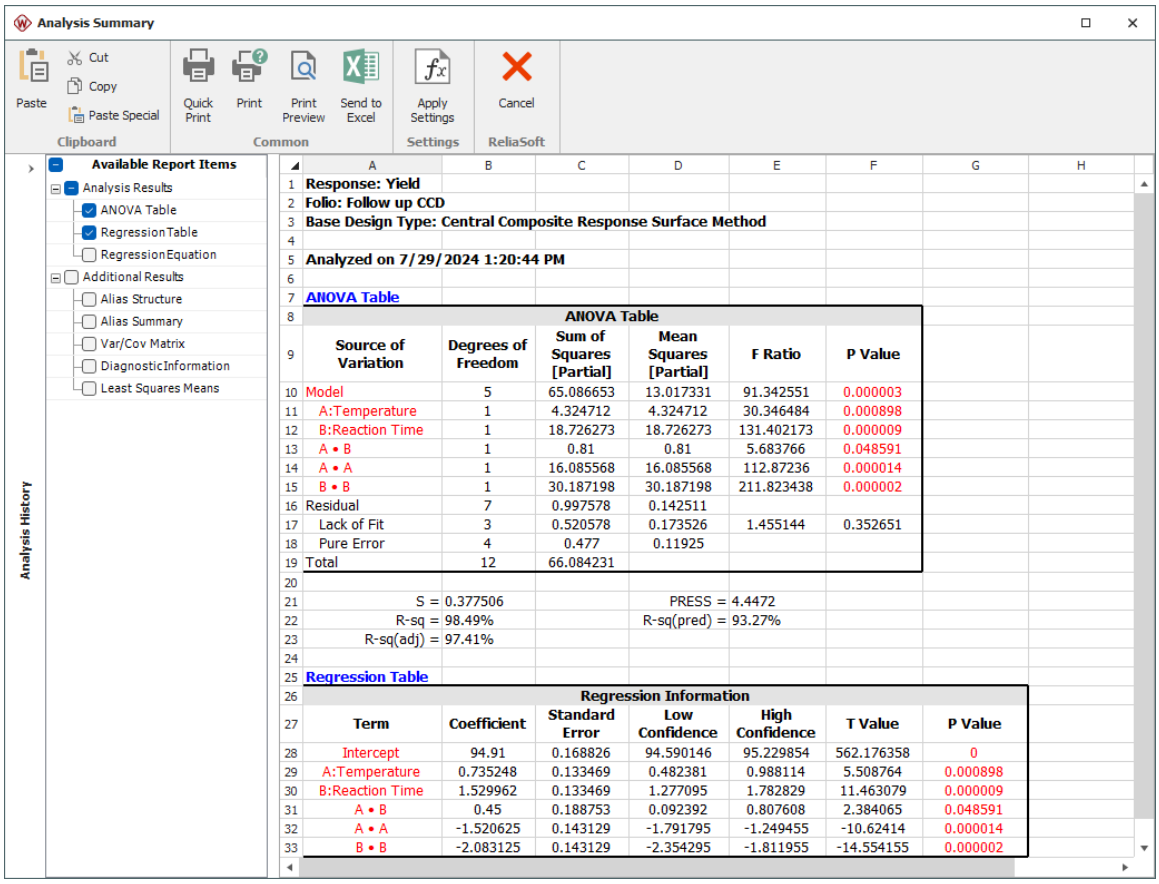
Such a design can be set up in a Weibull++ DOE folio using the properties shown in the figure below.



The resulting design is shown in the figure shown next.



Results from the analysis of the design are shown in the next figure.



Response: Yield
Folio: Follow up CCD
Base Design Type: Central Composite Response Surface Method

Analyzed on 7/29/2024 1:20:44 PM

ANOVA Table

ANOVA Table

Source of Variation	Degrees of Freedom	Sum of Squares [Partial]	Mean Squares [Partial]	F Ratio	P Value
Model	5	65.086653	13.017331	91.342551	0.000003
A:Temperature	1	4.324712	4.324712	30.346484	0.000898
B:Reaction Time	1	18.726273	18.726273	131.402173	0.000009
A • B	1	0.81	0.81	5.683766	0.048591
A • A	1	16.085568	16.085568	112.87236	0.000014
B • B	1	30.187198	30.187198	211.823438	0.000002
Residual	7	0.997578	0.142511		
Lack of Fit	3	0.520578	0.173526	1.455144	0.352651
Pure Error	4	0.477	0.11925		
Total	12	66.084231			

S = 0.377506 PRESS = 4.4472
R-sq = 98.49% R-sq(pred) = 93.27%
R-sq(adj) = 97.41%

Regression Table

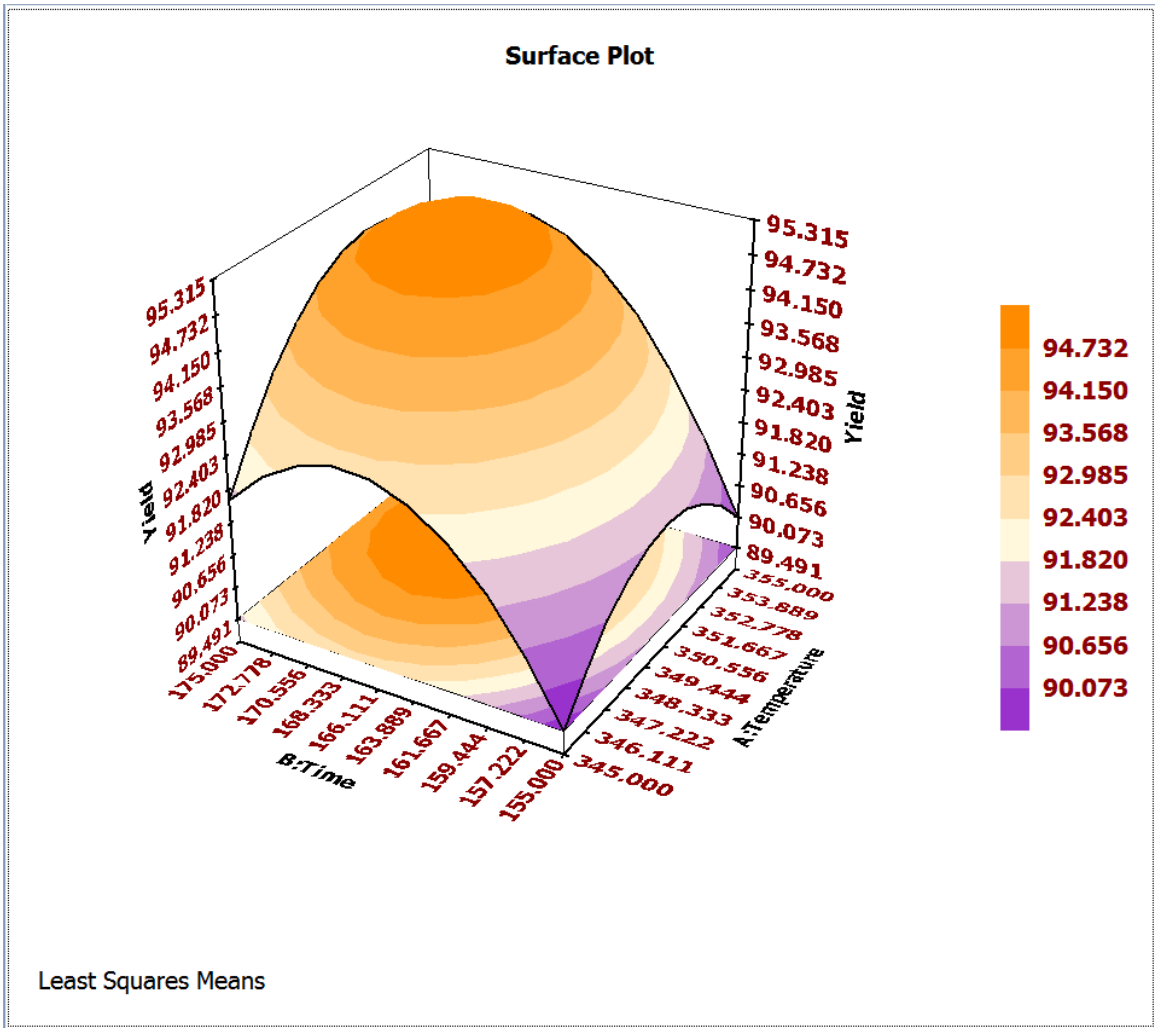
Regression Information

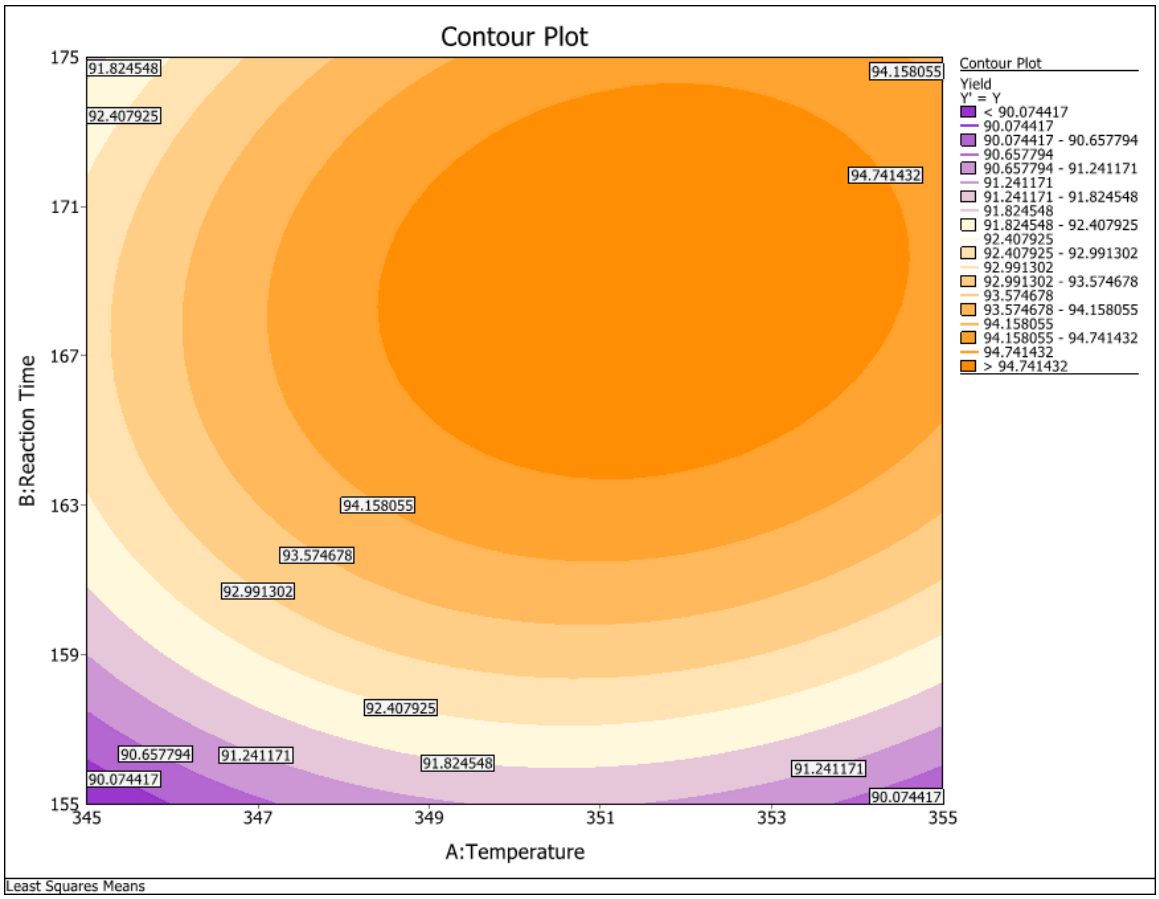
Term	Coefficient	Standard Error	Low Confidence	High Confidence	T Value	P Value
Intercept	94.91	0.168826	94.590146	95.229854	562.176358	0
A:Temperature	0.735248	0.133469	0.482381	0.988114	5.508764	0.000898
B:Reaction Time	1.529962	0.133469	1.277095	1.782829	11.463079	0.000009
A • B	0.45	0.188753	0.092392	0.807608	2.384065	0.048591
A • A	-1.520625	0.143129	-1.791795	-1.249455	-10.62414	0.000014
B • B	-2.083125	0.143129	-2.354295	-1.811955	-14.554155	0.000002

The results in the figure above show that the main effects, A and B, the interaction, AB, and the quadratic main effects, A^2 and B^2, (represented as AA and BB in the figure) are significant. The lack-of-fit test also shows that the second order model with these terms is adequate and a higher order model is not needed. Using these results, the model for the experiment in terms of the coded values is:

$$\hat{y} = 94.91 + 0.74x_1 + 1.53x_2 + 0.45x_1x_2 - 1.52x_1^2 - 2.08x_2^2$$

The response surface and the contour plot for this model, in terms of the actual variables, are shown in the below figures (a) and (b), respectively.


Surface Plot
Yield
95.315
94.732
94.150
93.568
92.985
92.403
91.820
91.238
90.656
90.073
89.491
B:Time
175.000
172.778
170.556
168.333
166.111
163.889
161.667
159.444
157.222
155.000
A:Temperature
355.000
353.889
352.778
351.667
350.556
349.444
348.333
347.222
346.111
345.000
Yield
95.315
94.732
94.150
93.568
92.985
92.403
91.820
91.238
90.656
90.073
89.491
94.732
94.150
93.568
92.985
92.403
91.820
91.238
90.656
90.073
Least Squares Means

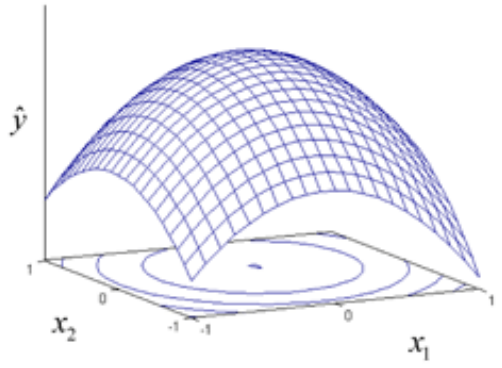


Analysis of the Second Order Model

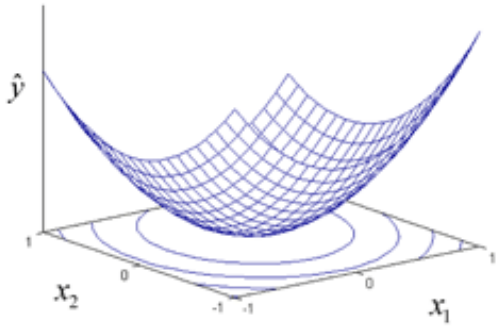
Once a second order model is fit to the response, the next step is to locate the point of maximum or minimum response. The second order model for k factors can be written as:

$$\hat{y} = \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i + \sum_{i=1}^{k} \hat{\beta}_{ii} x_i^2 + \sum_{i=1}^{k} \sum_{j=1}^{i-1} \hat{\beta}_{ij} x_i x_j$$

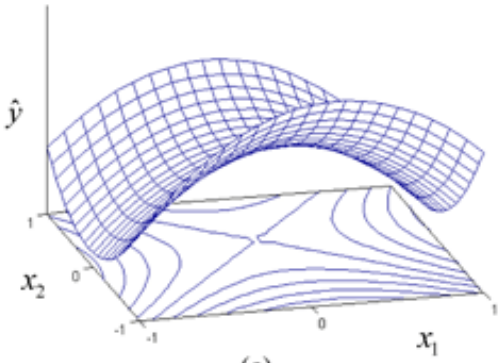
The point for which the response, $\hat{y}$, is optimized is the point at which the partial derivatives, $\partial\hat{y}/\partial x_1$, $\partial\hat{y}/\partial x_2$, $\ldots \partial\hat{y}/\partial x_k$, are all equal to zero. This point is called the *stationary point*. The stationary point may be a point of maximum response, minimum response or a saddle point. These three conditions are shown in the following figures (a), (b) and (c) respectively.



(a)



(b)



(c)

Notice that these conditions are easy to identify, in the case of two factor experiments, by the inspection of the contour plots. However, when more than two factors exist in an experiment, then the general mathematical solution for the location of the stationary point has to be used. The equation given above can be written in matrix notation as:

$$\hat{y} = \hat{\beta}_0 + x'b + x'Bx$$

where:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_k \end{bmatrix} \quad b = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ . \\ . \\ . \\ \hat{\beta}_k \end{bmatrix} \quad \text{and } B = \begin{bmatrix} \hat{\beta}_{11} & \hat{\beta}_{12}/2 & . & . & . & \hat{\beta}_{1k}/2 \\ & \hat{\beta}_{22} & . & . & . & \hat{\beta}_{2k}/2 \\ & & . & . & . & . \\ & & & . & . & . \\ & & & & . & . \\ \text{sym.} & & & & & \hat{\beta}_{kk} \end{bmatrix}$$

Then the stationary point can be determined as follows:

$$\frac{\partial \hat{y}}{\partial x} = b + 2Bx = 0$$

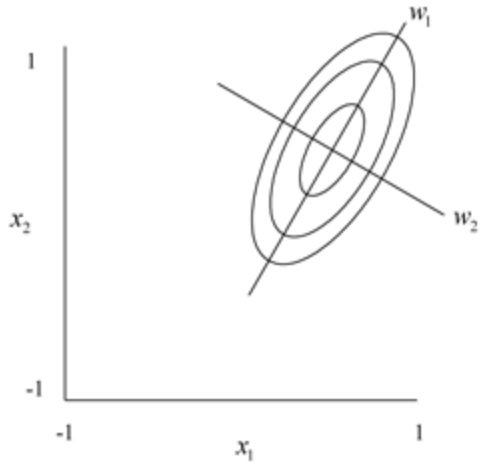
Thus, the stationary point is:

$$x_s = -\frac{1}{2} B^{-1} b$$

The optimum response is the response corresponding to x_s . The optimum response can be obtained to get:

$$\hat{y}_s = \hat{\beta}_0 + \frac{1}{2} x_s' b$$

Once the stationary point is known, it is necessary to determine if it is a maximum or minimum or saddle point. To do this, the second order model has to be transformed to the canonical form. This is done by transforming the model to a new coordinate system such that the origin lies at the stationary point and the axes are parallel to the principal axes of the fitted response surface, shown next.



The resulting model equation then takes the following form:

$$\hat{y} = \hat{y}_s + \lambda_1 w_1^2 + \lambda_2 w_2^2 + \ldots + \lambda_k w_k^2$$

where the w_i s are the transformed independent variables, and λ_i s are constants that are also the eigenvalues of the matrix B. The nature of the stationary point is known by looking at the signs of the λ_i s. If the λ_i s are all negative, then x_s is a point of maximum response. If the λ_i s are all positive then x_s is a point of minimum response. If the λ_i s have different signs, then x_s is a saddle point.

Example

Continuing with the example in Method of Steepest Ascent, the second order model fitted to the response, in terms of the coded variables, was obtained as:

$$\hat{y} = 94.91 + 0.74x_1 + 1.53x_2 + 0.45x_1x_2 - 1.52x_1^2 - 2.08x_2^2$$

Then the b and B matrices for this model are:

$$b = \begin{bmatrix} 0.74 \\ 1.53 \end{bmatrix} \quad B = \begin{bmatrix} -1.52 & 0.45 \\ 0.45 & -2.08 \end{bmatrix}$$

The stationary point is:

$$\begin{aligned} x_s &= -\frac{1}{2}B^{-1}b \\ &= -\frac{1}{2}\begin{bmatrix} -1.52 & 0.45 \\ 0.45 & -2.08 \end{bmatrix}^{-1}\begin{bmatrix} 0.74 \\ 1.53 \end{bmatrix} \\ &= \begin{bmatrix} 0.3 \\ 0.4 \end{bmatrix} \end{aligned}$$

Then, in terms of the actual values, the stationary point can be found as:

$$\text{Reaction Temperature: } 0.3 = \frac{T - 350}{(355 - 345)/2}$$
$$T = 351.5\ F$$

$$\text{Reaction Time: } 0.4 = \frac{t - 165}{(175 - 155)/2}$$
$$t = 169.0 \text{ minutes}$$

To find the nature of the stationary point the eigenvalues of the B matrix can be obtained as follows using the determinant of the matrix $B - \lambda I$:

$$\begin{aligned} |B - \lambda I| &= 0 \\ \begin{vmatrix} -1.52 - \lambda & 0.45 \\ 0.45 & -2.08 - \lambda \end{vmatrix} &= 0 \end{aligned}$$

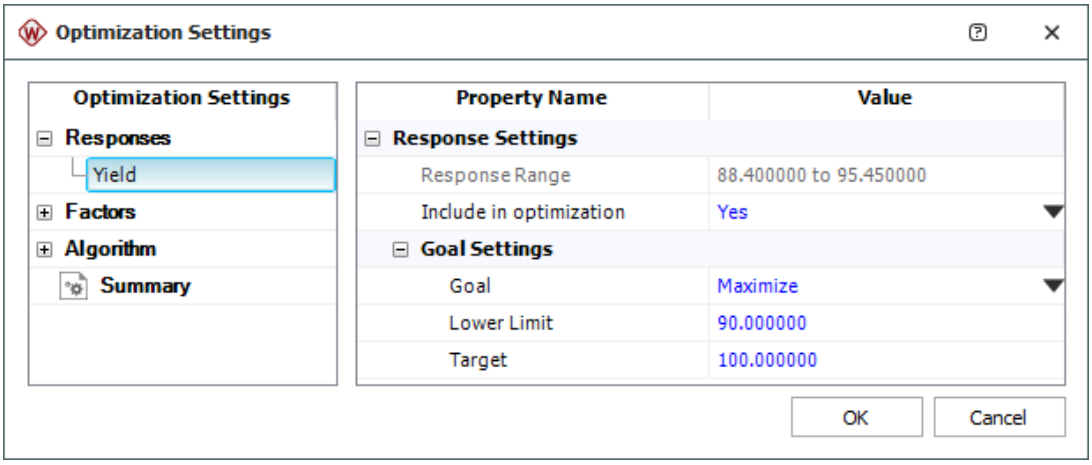
This gives us:

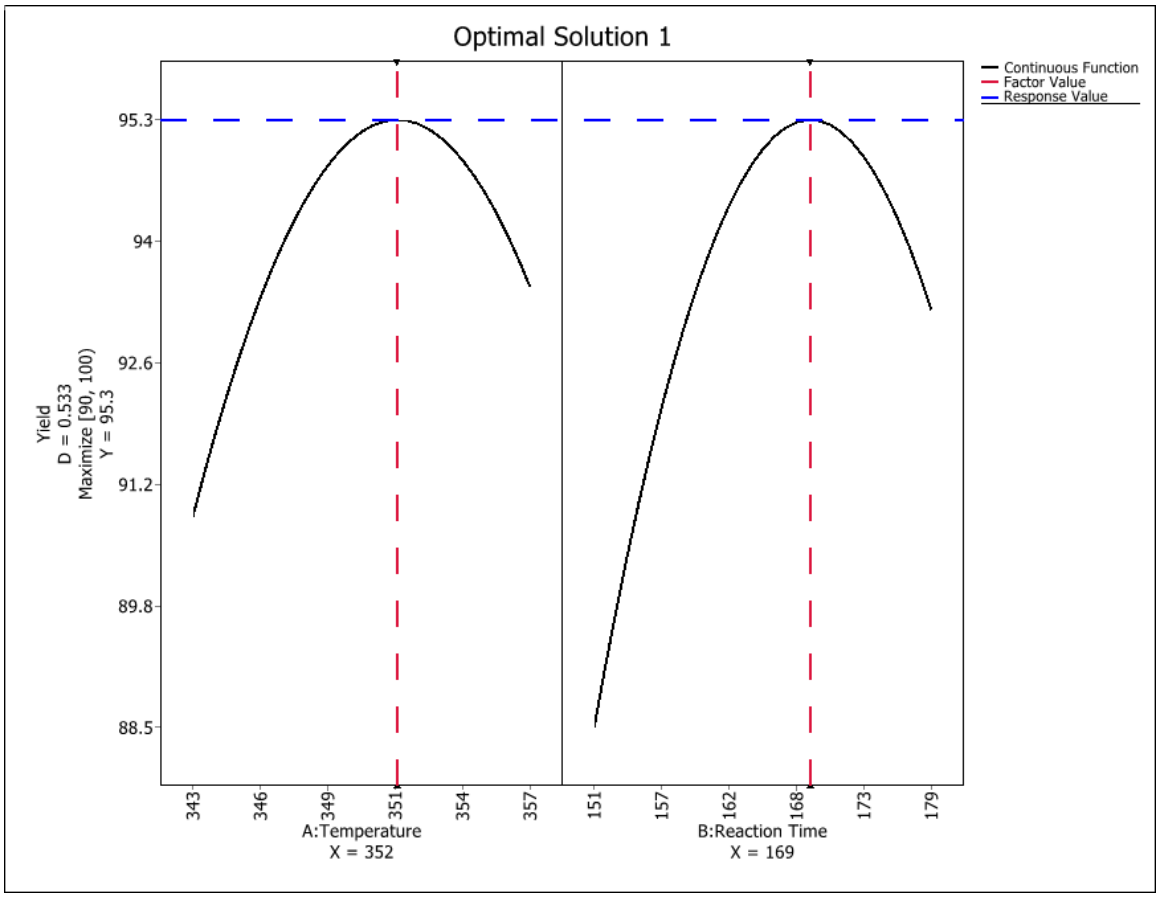
$$\lambda^2 + 3.61\lambda + 2.9743 = 0$$

Solving the quadratic equation in λ returns the eigenvalues $\lambda_1 = -1.2723$ and $\lambda_2 = -2.3377$. Since both the eigenvalues are negative, it can be concluded that the stationary point is a point of maximum response. The predicted value of the maximum response can be obtained as:

$$\begin{aligned}\hat{y}_s =& \hat{\beta}_0 + \frac{1}{2} x_s' b \\ =& 94.91 + \frac{1}{2} \begin{bmatrix} 0.3 \\ 0.4 \end{bmatrix}' \begin{bmatrix} 0.74 \\ 1.53 \end{bmatrix} \\ =& 95.3\end{aligned}$$

In a DOE folio, the maximum response can be obtained by entering the required values as shown in the figure below. In the figure, the goal is to maximize the response and the limits of the search range for maximizing the response are entered as 90 and 100. The value of the maximum response and the corresponding values of the factors obtained are shown in the second figure following. These values match the values calculated in this example.





Multiple Responses

In many cases, the experimenter has to optimize a number of responses at the same time. For the example in Method of Steepest Ascent, assume that the experimenter has to also consider two other responses: cost of the product (which should be minimized) and the pH of the product (which should be close to 7 so that the product is neither acidic nor basic). The data is presented in the figure below.

Factor Levels				Yield	Cost	pH
Coded		Actual				
A	*B*	*A*	*B*			
-1	-1	345	155	89.75	420.50	8.28
1	-1	355	155	90.20	419.60	7.86
-1	1	345	175	92.00	416.00	7.44
1	1	355	175	94.25	411.50	6.38
-1.414	0	342.9	165	90.50	419.00	7.76
1.414	0	357.1	165	92.75	414.50	6.80
0	-1.414	350	150.9	88.40	423.20	7.22
0	1.414	350	179.1	92.60	414.80	6.54
0	0	350	165	94.85	410.30	7.46
0	0	350	165	95.45	409.10	6.90
0	0	350	165	95.00	410.00	7.32
0	0	350	165	94.55	410.90	7.08
0	0	350	165	94.70	410.60	7.50

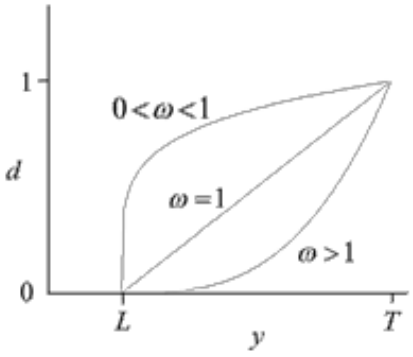
The problem in dealing with multiple responses is that now there might be conflicting objectives because of the different requirements of each of the responses. The experimenter needs to come up with a solution that satisfies each of the requirements as much as possible without compromising too much on any of the requirements. The approach used in a Weibull++ DOE folio to deal with optimization of multiple responses involves the use of desirability functions that are discussed next (for details see [Derringer and Suich, 1980]).

Desirability Functions

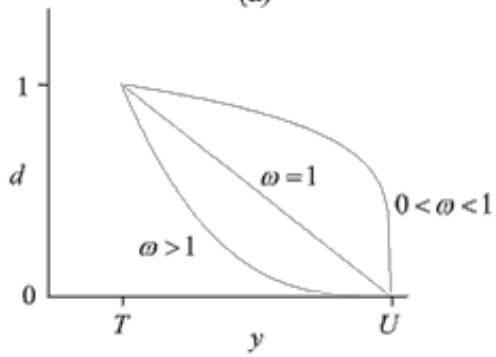
Under this approach, each i th response is assigned a desirability function, d_i , where the value of d_i varies between 0 and 1. The function, d_i, is defined differently based on the objective of the response. If the response is to be maximized, as in the case of the previous example where the yield had to be maximized, then d_i is defined as follows:

$$d_i = \begin{cases} 0 & y_i < L \\ \left(\frac{y_i - L}{T - L}\right)^{\omega} & L \le y_i \le T \\ 1 & y_i > T \end{cases}$$

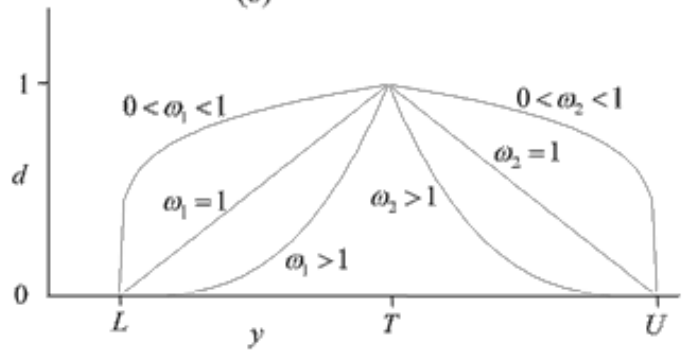
where T represents the target value of the i th response, y_i , L represents the acceptable lower limit value for this response and ω represents the weight. When $\omega = 1$ the function d_i is linear. If $\omega > 1$ then more importance is placed on achieving the target for the response, y_i . When $\omega < 1$, less weight is assigned to achieving the target for the response, y_i . A graphical representation is shown in figure (a) below.



(a)



(b)



(c)

If the response is to be minimized, as in the case when the response is cost, d_i is defined as follows:

$$d_i = \begin{cases} 1 & y_i < T \\ \left(\frac{U-y_i}{U-T}\right)^{\omega} & T \le y_i \le U \\ 0 & y_i > U \end{cases}$$

Here U represents the acceptable upper limit for the response (see figure (b) above).

There may be times when the experimenter wants the response to be neither maximized nor minimized, but instead stay as close to a specified target as possible. For example, in the case where the experimenter wants the product to be neither acidic nor basic, there is a requirement to keep

the pH close to the neutral value of 7. In such cases, the desirability function is defined as follows (see figure (c) above):

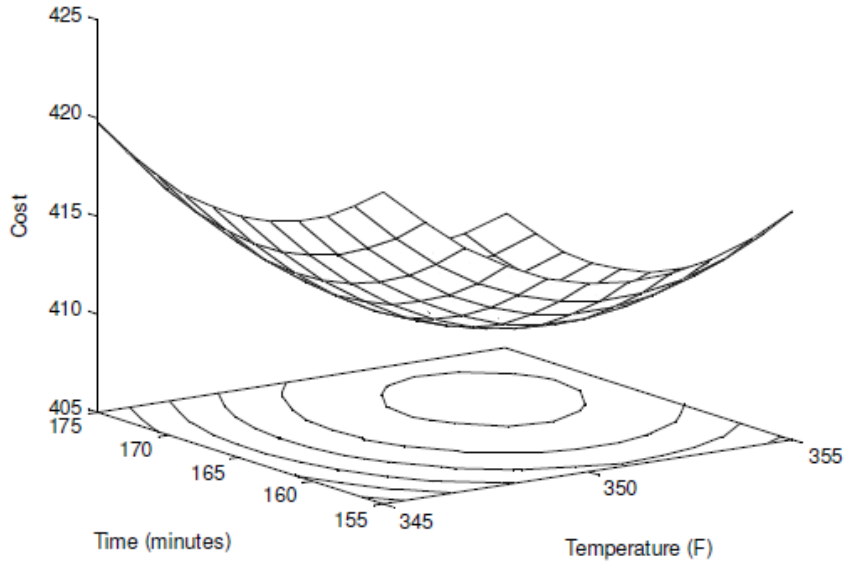
$$d_i = \begin{cases} 0 & y_i < L \\ \left(\frac{y_i - L}{T - L}\right)^{\omega_1} & L \le y_i \le T \\ \left(\frac{U - y_i}{U - T}\right)^{\omega_2} & T \le y_i \le U \\ 0 & y_i > U \end{cases}$$

Once a desirability function is defined for each of the responses, assuming that there are m responses, an overall desirability function is obtained as follows:

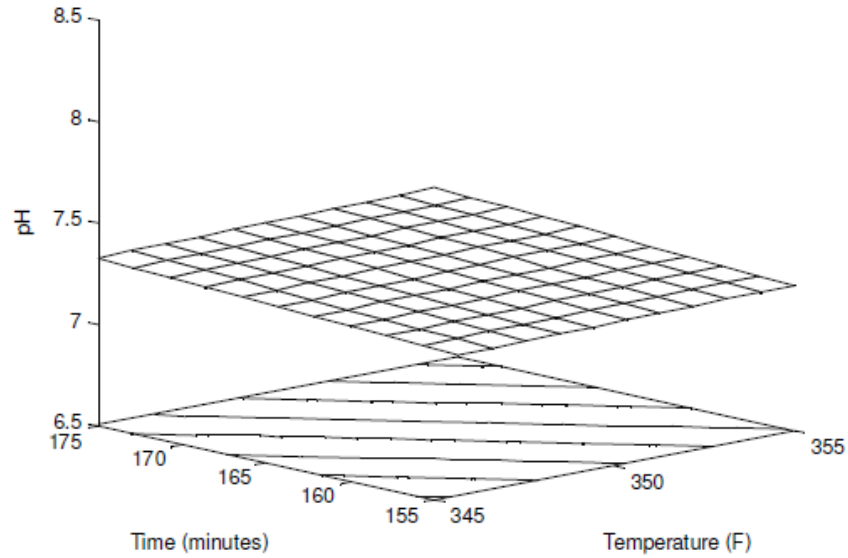
$$D = (d_1^{r_1} \cdot d_2^{r_2} \cdot \ldots \cdot d_m^{r_m})^{1/(r_1 + r_2 + \ldots + r_m)}$$

where the r_i s represent the importance of each response. The greater the value of r_i, the more important the response with respect to the other responses. The objective is to now find the settings that return the maximum value of D.

To illustrate the use of desirability functions, consider the previous example with the three responses of yield, cost and pH. The response surfaces for the two additional responses of cost and pH are shown next in the figures (a) and (b), respectively.



(a)



(b)

In terms of actual variables, the models obtained for all three responses are as shown next:

$$\text{Yield: } \hat{y}_1 = -7480.24 + 41.24x_1 + 3.88x_2 + 0.009x_1x_2 - 0.060825x_1^2 - 0.020831x_2^2$$

$$\text{Cost: } \hat{y}_2 = 15560.49 - 82.48x_1 - 7.76x_2 - 0.02x_1x_2 + 0.12165x_1^2 + 0.041663x_2^2$$

$$\text{pH: } \hat{y}_3 = 38.87 - 0.0709x_1 - 0.041x_2$$

Assume that the experimenter wants to have a target yield value of 95, although any value of yield greater than 94 is acceptable. Then the desirability function for yield is:

$$d_1 = \begin{cases} 0 & y_1 < 94 \\ \left(\frac{y_1 - 94}{95 - 94}\right)^{\omega} & 94 \le y_1 \le 95 \\ 1 & y_1 > 95 \end{cases}$$

For the cost, assume that the experimenter wants to lower the cost to 400, although any cost value below 415 is acceptable. Then the desirability function for cost is:

$$d_2 = \begin{cases} 1 & y_2 < 400 \\ \left(\frac{415 - y_2}{415 - 400}\right)^{\omega} & 400 \le y_2 \le 415 \\ 0 & y_2 > 415 \end{cases}$$

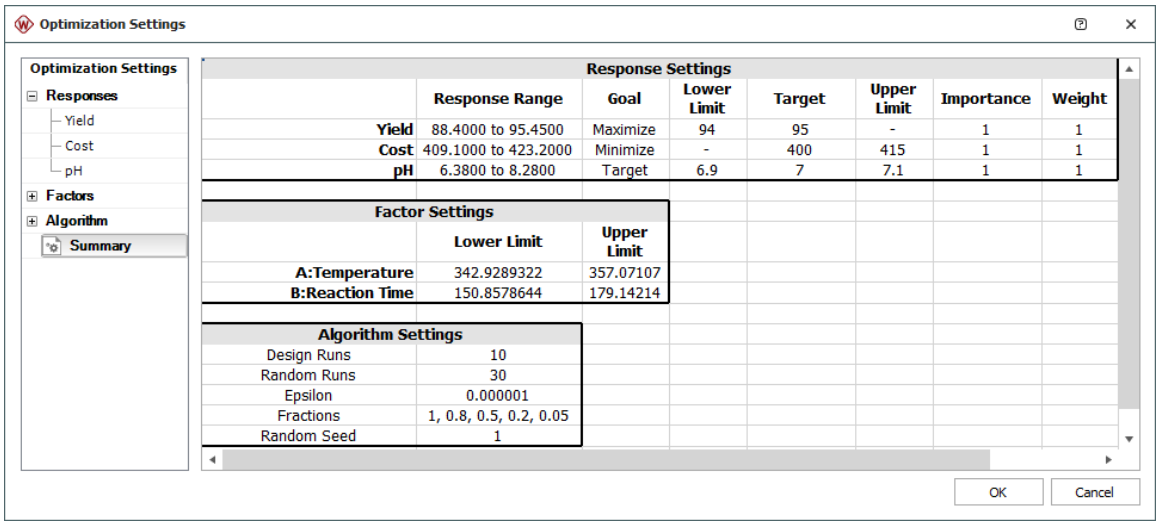
For the pH, a target of 7 is desired but values between 6.9 and 7.1 are also acceptable. Thus, the desirability function here is:

$$d_3 = \begin{cases} 0 & y_3 < L \\ \left(\frac{y_3 - 6.9}{7 - 6.9}\right)^{\omega_1} & 6.9 \le y_3 \le 7 \\ \left(\frac{7.1 - y_3}{7.1 - 7}\right)^{\omega_2} & 7 \le y_3 \le 7.1 \\ 0 & y_3 > 7.1 \end{cases}$$

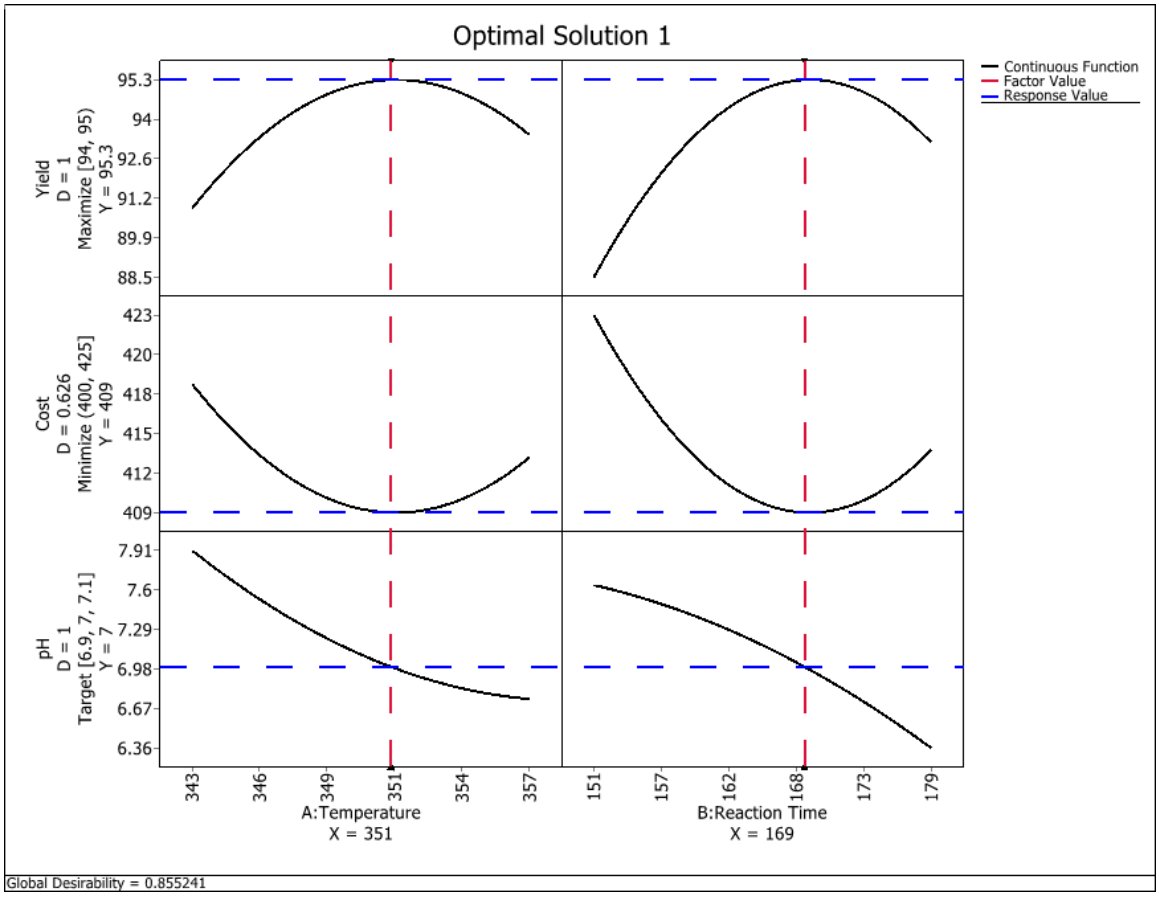
Notice that in the previous equations all weights used (ω_i s) are 1. Thus, all three desirability functions are linear. The overall desirability function, assuming equal importance ($r_1 = r_2 = r_3 = 1$) for all the responses, is:

$$\begin{aligned} D &= (d_1 \cdot d_2 \cdot d_3)^{1/3} \\ &= f(\hat{y}_1, \hat{y}_2, \hat{y}_3) \end{aligned}$$

The objective of the experimenter is to find the settings of x_1 and x_2 such that the overall desirability, D, is maximum. In a DOE folio, the settings for the desirability functions for each of the three responses can be entered as shown in the next figure.



Based on these settings, the Weibull++ DOE folio solves this optimization problem to obtain the following solution:



$$Temperature = 351.5\ F$$

$$Time = 169\ minutes$$

The overall desirability achieved with this solution can be calculated easily. The values of each of the response for these settings are:

$$\begin{aligned}\hat{y}_1 =& -7480.24 + 41.24x_1 + 3.88x_2 + 0.009x_1x_2 \\ & - 0.060825x_1^2 - 0.020831x_2^2 \\ =& -7480.24 + 41.24(351.5) + 3.88(169) + 0.009(351.3)(169) \\ & - 0.060825(351.5)^2 - 0.020831(169)^2 \\ =& 95.3 \\ \hat{y}_2 =& 15560.49 - 82.48x_1 - 7.76x_2 - 0.02x_1x_2 \\ & + 0.12165x_1^2 + 0.041663x_2^2 \\ =& 15560.49 - 82.48(351.5) - 7.76(169) - 0.02(351.5)(169) \\ & + 0.12165(351.5)^2 + 0.041663(169)^2 \\ =& 409.35 \\ \hat{y}_3 =& 38.87 - 0.0709x_1 - 0.041x_2 \\ =& 38.87 - 0.0709(351.5) - 0.041(169) \\ =& 7.00\end{aligned}$$

Based on the response values, the individual desirability functions are:

$$\begin{aligned}d_1 =& 1 \text{ (since } \hat{y}_1 > 95) \\ d_2 =& \left(\frac{415 - \hat{y}_2}{415 - 400}\right) \text{ (since } 400 \le y_2 \le 415) \\ =& \left(\frac{415 - 409.35}{415 - 400}\right) \\ =& 0.3767 \\ d_3 =& 1 \text{ (since } \hat{y}_3 \approx 7)\end{aligned}$$

Then the overall desirability is:

$$\begin{aligned}D =& (d_1 \cdot d_2 \cdot d_3)^{1/3} \\ =& (1 \cdot 0.3767 \cdot 1)^{1/3} \\ =& 0.72\end{aligned}$$

This is the same as the Global Desirability displayed by the Weibull++ DOE folio in the figure above. At times, a number of solutions may be obtained from the DOE folio, and it is up to the experimenter to choose the most feasible one.

Design Evaluation and Power Study

IN THIS CHAPTER

Alias Structure **263**

Orthogonality **265**

Optimality **267**

Power Study **268**
Power Study for Single Factor Designs (One-Way ANOVA) 268
Power Study for 2 Level Factorial Designs 281
Power Study for General Level Factorial Designs 284
Power Study for Response Surface Method Designs 297
Discussion on Power Calculation 305

Conclusion **306**

In general, there are three stages in applying design of experiments (DOE) to solve an issue: designing the experiment, conducting the experiment, and analyzing the data. The first stage is very critical. If the designed experiment is not efficient, you are unlikely to obtain good results. It is very common to evaluate an experiment before conducting the tests. A design evaluation often focuses on the following four properties:

1. **The alias structure**. Are main effects and two-way interactions in the experiment aliased with each other? What is the resolution of the design?
2. **The orthogonality**. An orthogonal design is always preferred. If a design is non-orthogonal, how are the estimated coefficients correlated?
3. **The optimality**. A design is called "optimal" if it can meet one or more of the following criteria:
   - *D*-optimality: minimize the determinant of the variance-covariance matrix.
   - *A*-optimality: minimize the trace of the variance-covariance matrix.

- *V*-optimality: minimize the average prediction variance in the design space.

4. **The power (or its inverse, Type II error).** Power is the probability of detecting an effect through experiments when it is indeed active. A design with low power for main effects is not a good design.

In the following sections, we will discuss how to evaluate a design according to these four properties.

Alias Structure

To reduce the sample size in an experiment, we usually focus only on the main effects and lower-order interactions, while assuming that higher-order interactions are not active. For example, screening experiments are often conducted with a number of runs that barely fits the main effect-only model. However, due to the limited number of runs, the estimated main effects often are actually combined effects of main effects and interaction effects. In other words the estimated main effects are *aliased* with interaction effects. Since these effects are aliased, the estimated main effects are said to be *biased*. If the interaction effects are large, then the bias will be significant. Thus, it is very important to find out how all the effects in an experiment are aliased with each other. A design's alias structure is used for this purpose, and its calculation is given below.

Assume the matrix representation of the true model for an experiment is:

$$\boldsymbol{Y} = \boldsymbol{X}_1\beta_1 + \boldsymbol{X}_2\beta_2 + \varepsilon$$

If the model used in a screening experiment is a reduced one, as given by:

$$\boldsymbol{Y} = \boldsymbol{X}_1\beta_1 + \varepsilon$$

then, from this experiment, the estimated β_1 is biased. This is because the ordinary least square estimator of β_1 is:

$$\hat{\beta}_1 = \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{Y}$$

As discussed in [Wu, 2000], the expected value of this estimator is:

$$\begin{aligned}
&E\left(\hat{\beta}_1\right) = E\left[\left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{Y}\right] \\
&= \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'E(\boldsymbol{Y}) \\
&= \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'E(\boldsymbol{X}_1\beta_1 + \boldsymbol{X}_2\beta_2 + \varepsilon) \\
&= \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_1\beta_1 + \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\beta_2 \\
&= \beta_1 + \boldsymbol{A}\beta_2
\end{aligned}$$

where $A = (X_1' X_1)^{-1} X_1' X_2$ is called the *alias matrix* of the design. For example, for a three factorial screening experiment with four runs, the design matrix is:

A	B	C
-1	-1	1
1	-1	-1
-1	1	-1
1	1	1

If we assume the *true* model is:

$$Y = \beta_0 + \beta_1 A + \beta_2 B + \beta_3 C + \beta_{12} AB + \beta_{13} AC + \beta_{23} BC + \beta_{123} ABC + \varepsilon$$

and the *used* model (i.e., the model used in the experiment data analysis) is:

$$Y = \beta_0 + \beta_1 A + \beta_2 B + \beta_3 C + \varepsilon$$

then $X_1 = [I\ A\ B\ C]$ and $X_2 = [AB\ AC\ BC\ ABC]$. The alias matrix A is calculated as:

	AB	AC	BC	ABC
I	0	0	0	1
A	0	0	1	0
B	0	1	0	0
C	1	0	0	0

Sometimes, we also put X_1 in the above matrix. Then the A matrix becomes:

	I	A	B	C	AB	AC	BC	ABC
I	1	0	0	0	0	0	0	1
A	0	1	0	0	0	0	1	0
B	0	0	1	0	0	1	0	0
C	0	0	0	1	1	0	0	0

For the terms included in the *used* model, the alias structure is:

$[I] = I + ABC$
$[A] = A + BC$
$[B] = B + AC$
$[C] = C + AB$

From the alias structure and the definition of resolution, we know this is a resolution III design. The estimated main effects are aliased with two-way interactions. For example, *A* is aliased with *BC*. If, based on engineering knowledge, the experimenter suspects that some of the interactions are important, then this design is unacceptable since it cannot distinguish the main effect from important interaction effects.

For a designed experiment it is better to check its alias structure before conducting the experiment to determine whether or not some of the important effects can be clearly estimated.

Orthogonality

Orthogonality is a model-related property. For example, for a main effect-only model, if all the coefficients estimated through ordinary least squares estimation are not correlated, then this experiment is an *orthogonal design for main effects*. An orthogonal design has the minimal variance for the estimated model coefficients. Determining whether a design is orthogonal is very simple. Consider the following model:

$$Y = X\beta + \varepsilon$$

The variance and covariance matrix for the model coefficients is:

$$Var\left(\hat{\beta}\right) = \sigma_{\varepsilon}^{2}\left(X'X\right)^{-1}$$

where σ_{ε}^{2} is the variance of the error. When all the factors in the model are quantitative factors or all the factors are 2 levels, $Var\left(\hat{\beta}\right)$ is a regular symmetric matrix . The diagonal elements of it are the variances of model coefficients, and the off-diagonal elements are the covariance among these coefficients. When some of the factors are qualitative factors with more than 2 levels, $Var\left(\hat{\beta}\right)$ is a block symmetric matrix. The block elements in the diagonal represent the variance and covariance matrix of the qualitative factors, and the off-diagonal elements are the covariance among all the coefficients.

Therefore, to check if a design is orthogonal for a given model, we only need to check matrix : $\left(X'X\right)^{-1}$. For the example used in the previous section, if we assume the main effect-only model is used, then $\left(X'X\right)^{-1}$ is:

	I	A	B	C
I	0.25	0	0	0
A	0	0.25	0	0
B	0	0	0.25	0
C	0	0	0	0.25

Since all the off-diagonal elements are 0, the design is an orthogonal design for main effects. For an orthogonal design, it is also true that the diagonal elements are $1/n$, where n is the number of total runs.

When there are qualitative factors with more than 2 levels in the model, $\left(X'X\right)^{-1}$ will be a block symmetric matrix. For example, assume we have the following design matrix.

Run Order	A	B
1	-1	1
2	-1	1
3	-1	1
4	-1	2
5	-1	2
6	-1	2
7	-1	3
8	-1	3
9	-1	3
10	1	1
11	1	1
12	1	1
13	1	2
14	1	2
15	1	2
16	1	3
17	1	3
18	1	3

Factor B has 3 levels, so 2 indicator variables are used in the regression model. The $\left(X'X\right)^{-1}$ matrix for a model with main effects and the interaction is:

	I	A	B[1]	B[2]	AB[1]	AB[2]
I	0.0556	0	0	0	0	0
A	0	0.0556	0	0	0	0
B[1]	0	0	0.1111	-0.0556	0	0
B[2]	0	0	-0.0556	0.1111	0	0
AB[1]	0	0	0	0	0.1111	-0.0556
AB[2]	0	0	0	0	-0.0556	0.1111

The above matrix shows this design is orthogonal since it is a block diagonal matrix.

For an orthogonal design for a given model, all the coefficients in the model can be estimated independently. Dropping one or more terms from the model will not affect the estimation of other coefficients and their variances. If a design is not orthogonal, it means some of the terms in the model are correlated. If the correlation is strong, then the statistical test results for these terms may not be accurate.

VIF (variance inflation factor) is used to examine the correlation of one term with other terms. The VIF is commonly used to diagnose multicollinearity in regression analysis. As a rule of thumb, a VIF of greater than 10 indicates a strong correlation between some of the terms. VIF can be simply calculated by:

$$VIF_i = \frac{n}{\sigma^2}\,\mathbf{var}\left(\hat{\beta}_i\right)$$

For more detailed discussion on VIF, please see Multiple Linear Regression Analysis.

Optimality

Orthogonal design is always ideal. However, due to the constraints on sample size and cost, it is sometimes not possible. If this is the case, we want to get a design that is as orthogonal as possible. The so-called *D*-efficiency is used to measure the orthogonality of a two level factorial design. It is defined as:

$$D\text{-efficiency} = \left(\frac{|X'X|}{n^p}\right)^{1/p}$$

where *p* is the number of coefficients in the model and *n* is the total sample size. *D* represents the determinant.

$X'X$ is the *information matrix* of a design. When you compare two different screening designs, the one with a larger determinant of $X'X$ is usually better. *D-efficiency* can be used for comparing two designs. Other alphabetic optimal criteria are also used in design evaluation. If a model and the number of runs are given, an optimal design can be found using computer algorithms for one of the following optimality criteria:

- *D*-optimality: maximize the determinant of the information matrix $X'X$. This is the same as minimizing the determinant of the variance-covariance matrix $(X'X)^{-1}$.
- *A*-optimality: minimize the trace of the variance-covariance matrix $(X'X)^{-1}$. The trace of a matrix is the sum of all its diagonal elements.
- *V*-optimality (or *I*-optimality): minimize the average prediction variance within the design space.

The determinant of $X'X$ and the trace of $(X'X)^{-1}$ are given in the design evaluation in the DOE folio. *V*-optimality is not yet included.

Power Study

Power calculation is another very important topic in design evaluation. When designs are balanced, calculating the power (which, you will recall, is the probability of detecting an effect when that effect is active) is straightforward. However, for unbalanced designs, the calculation can be very complicated. We will discuss methods for calculating the power for a given effect for both balanced and unbalanced designs.

Power Study for Single Factor Designs (One-Way ANOVA)

Power is related to Type II error in hypothesis testing and is commonly used in statistical process control (SPC). Assume that at the normal condition, the output of a process follows a normal distribution with a mean of 10 and a standard deviation of 1.2. If the 3-sigma control limits are used and the sample size is 5, the control limits (assuming a normal distribution) for the X-bar chart are:

$$UCL = \bar{x} + 3\frac{\sigma}{\sqrt{n}} = 10 + 3\frac{1.2}{\sqrt{5}} = 11.61$$
$$LCL = \bar{x} - 3\frac{\sigma}{\sqrt{n}} = 10 - 3\frac{1.2}{\sqrt{5}} = 8.39$$

If a calculated mean value from a sampling group is outside of the control limits, then the process is said to be out of control. However, since the mean value is from a random process following a normal distribution with a mean of 10 and standard derivation of $\sigma/\sqrt{n}$, even when the process is under control, the sample mean still can be out of the control limits and cause a false alarm. The probability of causing a false alarm is called *Type I error* (or significance level or risk level). For this example, it is:

$$\text{Type I Err} = 2 \times (1 - \Phi(3)) = 0.0027$$

Similarly, if the process mean has shifted to a new value that means the process is indeed out of control (e.g., 12), applying the above control chart, the sample mean can still be within the control limits, resulting in a failure to detect the shift. The probability of causing a misdetection is called *Type II error*. For this example, it is:

$$\text{Type II Err} = \Pr\{LCL < \bar{x} < UCL | \mu = 12\} = \Phi\left(\frac{UCL - 12}{1.2/\sqrt{5}}\right) - \Phi\left(\frac{LCL - 12}{1.2/\sqrt{5}}\right)$$
$$= \Phi(-0.72672) - \Phi(-6.72684)$$
$$= 0.2337$$

Power is defined as 1-Type II error. In this case, it is 0.766302. From this example, we can see that Type I and Type II errors are affected by sample size. Increasing sample size can reduce both errors. Engineers usually determine the sample size of a test based on the power requirement for a given effect. This is called the *Power and Sample Size* issue in design of experiments.

Power Calculation for Comparing Two Means

For one factor design, or one-way ANOVA, the simplest case is to design an experiment to compare the mean values at two different levels of a factor. Like the above control chart example, the calculated mean value at each level (in control and out of control) is a random variable. If the two means are different, we want to have a good chance to detect it. The difference of the two means is called the *effect* of this factor. For example, to compare the strength of a similar rope from two different manufacturers, 5 samples from each manufacturer are taken and tested. The test results (in newtons) are given below.

M1	M2
123	99
134	103
132	100

100	105
98	97

For this data, the ANOVA results are:

ANOVA Table

ANOVA Table					
Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F Ratio	P Value
Treatment	1	688.9	688.9	4.444516	0.068063
Error	8	1240	155		
Total	9	1928.9			
S =	12.4499				
R-sq =	35.714656%				

The standard deviation of the error is 12.4499 as shown in the above screenshot. and the *t*-test results are:

Mean Comparisons						
Contrast	Mean Difference	Pooled Standard Error	Low CI	High CI	T Value	P Value
M1 - M2	16.6	7.874	-1.5575	34.7575	2.1082	0.0681

Since the *p* value is 0.0681, there is no significant difference between these two vendors at a significance level of 0.05 (since .0681 > 0.05). However, since the samples are randomly taken from the two populations, if the true difference between the two vendors is 30, what is the power of detecting this amount of difference from this test?

To answer this question: first, from the significance level of 0.05, let's calculate the critical limits for the *t*-test. They are:

$$L = t^{-1}_{0.025,v=8} = -2.306$$
$$U = t^{-1}_{0.975,v=8} = 2.306$$

Define the mean of each vendor as μ_i and $d = \mu_1 - \mu_2$. Then the difference between the estimated sample means is:

$$\hat{d} = \hat{\mu}_1 - \hat{\mu}_2$$

Under the null hypothesis (the two vendors are the same), the *t* statistic is:

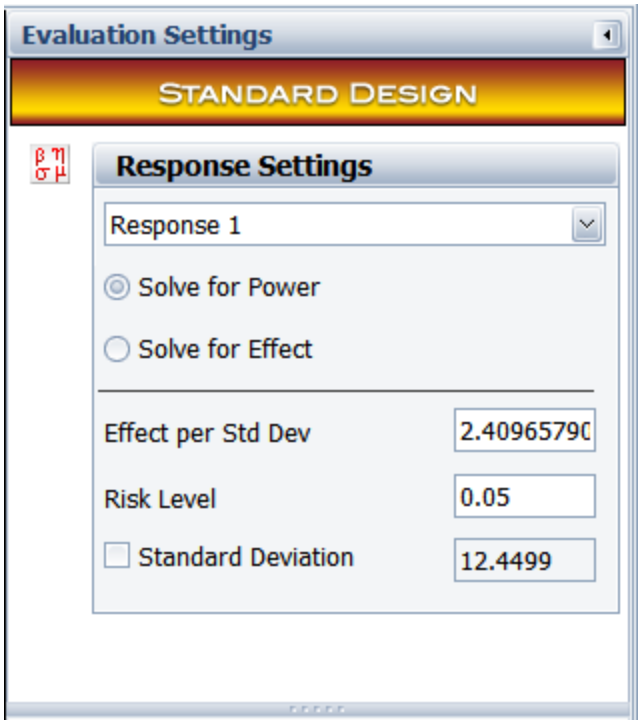
$$t_0 = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}$$

Under the alternative hypothesis when the true difference is 30, the calculated *t* statistic is from a non-central *t* distribution with non-centrality parameter δ of:

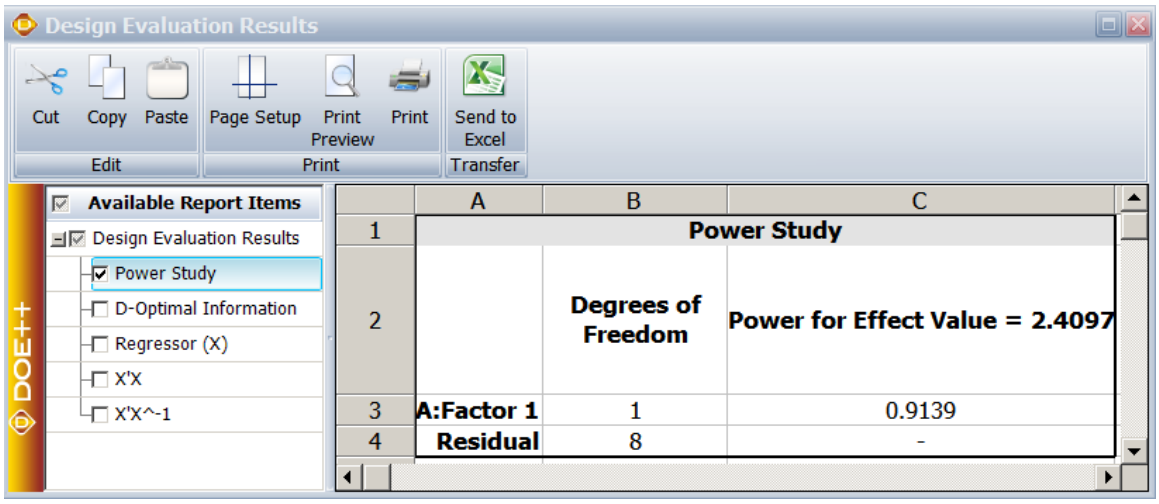
$$\delta = \frac{30}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} = 3.81004$$

The Type II error is $\Pr\{L < t_0 < U | d = 30\} = 0.08609$. So the power is 1-0.08609 =0.91391.

In a DOE folio, the ***Effect*** for the power calculation is entered as the multiple of the standard deviation of error. So effect of 30 is $30/S = 30/12.4499 = 2.409658$ standard deviation. This information is illustrated below.



and the calculated power for this effect is:



	A	B	C
1	Power Study		
2		Degrees of Freedom	Power for Effect Value = 2.4097
3	A:Factor 1	1	0.9139
4	Residual	8	-

As we know, the square of a *t* distribution is an *F* distribution. The above ANOVA table uses the *F* distribution and the above "mean comparison" table uses the *t* distribution to calculate the *p* value. The ANOVA table is especially useful when conducting multiple level comparisons. We will illustrate how to use the *F* distribution to calculate the power for this example.

At a significance level of 0.05, the critical value for the *F* distribution is:

$$U = f^{-1}_{0.05,v1=1,v2=8} = 5.317655$$

Under the alternative hypothesis when the true difference of these 2 vendors is 30, the calculated *f* statistic is from a non-central *F* distribution with non-centrality parameter $\phi = \delta^2 = 14.5161$.

The Type II error is $\Pr\{f < U | d = 30\} = F_{v1=1,v2=8,\phi=14.5161}(f < U) = 0.08609$. So the power is 1-0.08609 = 0.91391. This is the same as the value we calculated using the non-central t distribution.

Power Calculation for Comparing Multiple Means: Balanced Designs

When a factor has only two levels, as in the above example, there is only one effect of this factor, which is the difference of the means at these two levels. However, when there are multiple levels, there are multiple paired comparisons. For example, if there are *r* levels for a factor, there are $\binom{r}{2}$ paired comparisons. In this case, what is the power of detecting a given difference among these comparisons?

In a DOE folio, the power for a multiple level factor is defined as follows: given the largest difference among all the level means is Δ*, power is the smallest probability of detecting this difference at a given significance level.*

For example, if a factor has 4 levels and Δ is 3, there are many scenarios that the largest difference among all the level means will be 3. The following table gives 4 possible scenarios.

Case	*M1*	*M2*	*M3*	*M4*
1	24	27	25	26
2	25	25	26	23
3	25	25	25	28
4	25	25	26.5	23.5

For all 4 cases, the largest difference among the means is the same: 3. The probability of detecting $\Delta = 3$ (individual power) can be calculated using the method in the previous section for each case. It has been proven in [Kutner etc 2005, Guo etc 2012] that when the experiment is balanced, case 4 gives the lowest probability of detecting a given amount of effect. Therefore, the individual power calculated for case 4 is also the power for this experiment. In case 4, all but two factor level means are at the grand mean, and the two remaining factor level means are equally spaced around the grand mean. Is this a general pattern? Can the conclusion from this example be applied to general cases of balanced design?

To answer these questions, let's illustrate the power calculation mathematically. In one factor design or one-way ANOVA, a level is also traditionally called a *treatment*. The following linear regression model is used to model the data:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \ldots + \beta_{r-1} X_{ij,r-1} + \varepsilon_{ij}$$

where Y_{ij} is the jth observation at the ith treatment and

$$X_{ij1} = \begin{cases} 1 \text{ if case from factor level } 1 \\ -1 \text{ if case from factor level } r \\ 0 \text{ otherwise} \end{cases}$$

$$\vdots$$

$$X_{ij,r-1} = \begin{cases} 1 \text{ if case from factor level } r\text{-}1 \\ -1 \text{ if case from factor level } r \\ 0 \text{ otherwise} \end{cases}$$

First, let's define the problem of power calculation.

The power calculation of an experiment can be mathematically defined as:

$$
\begin{aligned}
&min\ P\{f_{critical} < F(1-\alpha; r-1, n_T - r)\,|\phi\} \\
&subject\ to \\
&\max_{i \neq j}\ (|\mu_i - \mu_j|) = \Delta,\ i, j = 1, \ldots, r
\end{aligned}
$$

where r is the number of levels, n_T is the total samples, α is the significance level of the hypothesis testing, and $f_{critical}$ is the critical value. The obtained minimal of the objective function in the above optimization problem is the power. The above optimization is the same as minimizing ϕ, the non-centrality parameter, since all the other variables in the non-central F distribution are fixed.

Second, let's relate the level means with the regression coefficients.

Using the regression model, the mean response at the ith factor level is:

$$
\begin{cases}
\mu_i = \beta_0 + \beta_i \text{ for } i < r \\
\mu_r = \beta_0 - \sum_{i=1}^{r-1} \beta_i \text{ for } i = r
\end{cases}
$$

The difference of level means can also be defined using the β values. For example, let $\Delta_{ij} = \mu_i - \mu_j$, then:

$$
\Delta_{ij} = \begin{cases}
|\beta_i - \beta_j \mid\ i < j,\ j \neq r \\
\left| 2\beta_i + \sum_{l \neq i} \beta_l \right| \ i = r
\end{cases}
$$

Using β, the non-centrality parameter ϕ can be calculated as:

$$
\phi = \beta \Sigma_\beta^{-1} \beta^T
$$

where $\beta = (\beta_1, \beta_2, \ldots, \beta_{r-1})$ and Σ_β is the variance and covariance matrix for β. When the design is balanced, we know:

$$
\Sigma_\beta^{-1} = \frac{1}{\sigma^2} X_\beta^T X_\beta = \frac{1}{\sigma^2} \begin{pmatrix}
2n & n & & \ldots & n \\
n & 2n & \ldots & n & \\
\ldots & \ldots & \ldots & \ldots & \\
n & n & \ldots & 2n &
\end{pmatrix}
$$

where n is the sample size at each level.

Third, let's solve the optimization problem for balanced designs.

The power is calculated when ϕ is at its minimum. Therefore, for balanced designs, the optimization issue becomes:

$$\min \ \phi = \frac{2n}{\sigma^2}\left[\sum_{i=1}^{r}\beta_i^2 + \sum_{i=1}^{r}\sum_{i\neq j}^{r}\beta_i\beta_j\right]$$

subject to

$$\max \left\{ \underset{i<j, j\neq r}{|\mu_i - \mu_j|}, \underset{i<r,}{|\mu_i - \mu_r|} \right\}$$
$$= \max \left\{ \underset{i<j, j\neq r}{|\beta_i - \beta_j|}, |2\beta_i + \sum_{l\neq i}\beta_l| \right\} = \Delta$$

The two equations in the constraint represent two cases. Without losing generality, Δ is set to 1 in the following discussion.

Case 1: $\Delta = \mu_k - \mu_l$, that is, the last level of the factor does not appear in the difference of level means. For example, let $\Delta_{kl} = \mu_k - \mu_l = \beta_k - \beta_l = 1$. $k, l \neq r$. The optimal solution is $\beta_k = 0.5$, $\beta_l = -0.5$, $\beta_i = 0$ for $i \neq k, l$. This result means that at the optimal solution, $\mu_k = 0.5$, $\mu_l = -0.5$, $\mu_i = 0$, $i \neq k, l$.

Case 2: In this case, one level in the comparisons is the last level of the factor in the largest difference of $\Delta = 1$.

For example, let $\Delta_{kr} = \mu_k - \mu_r = \begin{matrix} 2\beta_k + \sum_{l\neq i}\beta_l = 1 \end{matrix}$.

The optimal solution is $\beta_k = 0.5$, $\beta_i = 0$ for $i \neq k$. This result means that at the optimal solution, $\mu_k = 0.5$, $\mu_r = -0.5$, and $\mu_i = 0$, $i \neq k, r$.

The proof for Case 1 and Case 2 is given in [Guo IEEM2012]. The results for Case 1 and Case 2 show that when one of the level means (adjusted by the grand mean) is $\Delta/2$, another level mean is -$\Delta/2$ and the rest level means are 0, the calculated power is the smallest power among all the possible scenarios. This result is the same as the observation for the 4-case example given at the beginning at this section.

Let's use the above optimization method to solve the example given in the previous section. In that example, the factor has 2 levels; the sample size is 5 at each level; the estimated $\sigma^2 = 155$; and $\Delta = 30$. The regression model is:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij1} + \varepsilon_{ij}$$

Since the sample size is 5, $\Sigma_{\beta}^{-1} = \frac{2n}{\sigma^2} = \frac{10}{155} = 0.064516$. From the above discussion, we know that when $\beta_1 = 0.5\Delta$, we get the minimal non-centrality parameter $\phi = \beta_1 \Sigma_{\beta}^{-1} \beta_1 = 14.51613$. This value is the same as what we got in the previous section using the non-central t and F distributions. Therefore, the method discussed in this section is a general method and can be used for cases with 2 level and multiple level factors. The previous non-central t and F distribution method is only for cases with 2 level factors.

A 4 level balanced design example

Assume an engineer wants to compare the performance of 4 different materials. Each material is a level of the factor. The sample size for each level is 15 and the standard deviation σ is 10. The engineer wants to calculate the power of this experiment when the largest difference among the materials is 15. If the power is less than 80%, he also wants to know what the sample size should be in order to obtain a power of 80%. Assume the significant level is 5%.

Step 1: Build the linear regression model. Since there are 4 levels, we need 3 indicator variables. The model is:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ij3} + \varepsilon_{ij}$$

Step 2: Since the sample size is 15 and σ is 10:

$$\Sigma_{\beta}^{-1} = \frac{1}{\sigma^2} \begin{pmatrix} 30 & 15 & 15 \\ 15 & 30 & 15 \\ 15 & 15 & 30 \end{pmatrix} = \begin{pmatrix} 0.30 & 0.15 & 0.15 \\ 0.15 & 0.30 & 0.15 \\ 0.15 & 0.15 & 0.30 \end{pmatrix}$$

Step 3: Since there are 4 levels, there are 6 paired comparisons. For each comparison, the optimal β is:

ID	Paired Comparison	beta1	beta2	beta3
1	Level 1- Level 2 0.5	-0.5	0	
2	Level 1- Level 3	0.5	0	-0.5
3	Level 1- Level 4	0.5	0	0
4	Level 2- Level 3	0	0.5	-0.5
5	Level 2- Level 4	0	0.5	0
6	Level 3- Level 4	0	0	0.5

Step 4: Calculate the non-centrality parameter for each of the 6 solutions:

$$\Phi = \beta\Sigma_{\beta}^{-1}\beta'\Delta^2 = \begin{pmatrix} 16.875 & 8.4375 & 8.4375 & -8.4375 & -8.4375 & 0 \\ & 16.875 & 8.4375 & 8.4375 & 0 & -8.4375 \\ & & 16.875 & 0 & 8.4375 & 8.4375 \\ & & & 16.875 & 8.4375 & -8.4375 \\ & & & & 16.875 & 8.4375 \\ & & & & & 16.875 \end{pmatrix}$$

The diagonal elements are the non-centrality parameter from each paired comparison. Denoting them as ϕ_i , the power should be calculated using $\phi = \min(\phi_i)$. Since the design is balanced, we see here that all the ϕ_i are the same.

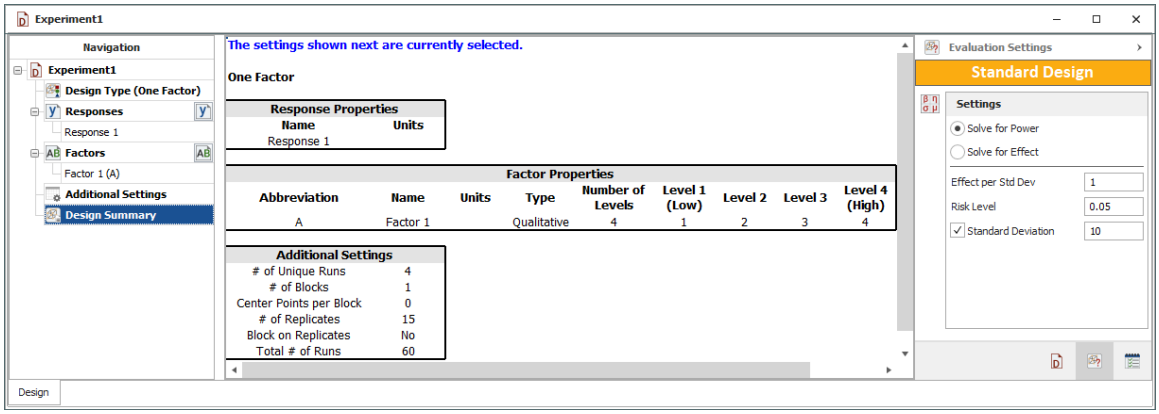
Step 5: Calculate the critical F value.

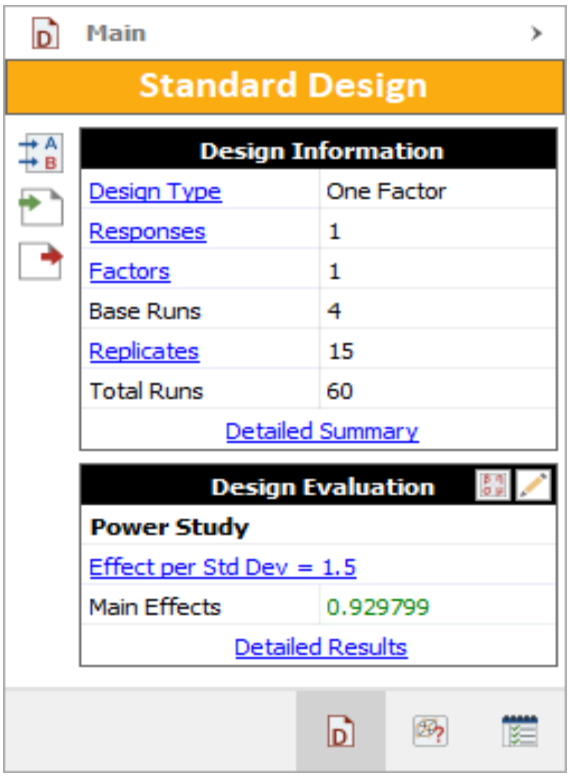
$$f_{critical} = F_{3,56}^{-1}(0.05) = 2.7694$$

Step 6: Calculate the power for this design using the non-central F distribution.

$$Power = 1 - F_{3,56}\left(f_{critical} | \phi = 16.875\right) = 0.9298$$

Since the power is greater than 80%, the sample size of 15 is sufficient. Otherwise, the sample size should be increased in order to achieve the desired power requirement. The settings and results in the DOE folio are given below.





Power Calculation for Comparing Multiple Means: Unbalanced Designs

If the design is not balanced, the non-centrality parameter does not have the simple expression of $\phi=\frac{2n}{\sigma^2}\left[\sum_{i=1}^{r}\beta_i^2+\sum_{i=1}^{r}\sum_{i\neq j}^{r}\beta_i\beta_j\right]$, since Σ_β^{-1} will not have the simpler format seen in balanced designs. The optimization thus becomes more complicated. For each paired comparison, we need to solve an optimization problem by assuming this comparison has the largest difference. For example, assuming the ith comparison $\Delta=\max_{i<j}|\mu_i-\mu_j|$ has the largest difference, we need to solve the following problem:

$$\min\ \phi_i=\beta\Sigma_\beta^{-1}\beta^T$$

$$subject to$$

$$\Delta=\max_{i<j}|\mu_i-\mu_j|$$

$$and$$

$$\left\{\max_{i<k,k\neq j}|\mu_i-\mu_k|\right\}\leq\Delta$$

In total, we need to solve $\binom{r}{2}$ optimization problems and use the smallest $\min(\phi_i)$ among all the solutions to calculate the power of the experiment. Clearly, the calculation will be very expensive.

In a DOE folio, instead of calculating the exact solution, we use the optimal β for a balanced design to calculate the approximated power for an unbalanced design. It can be seen that the optimal β for a balanced design also can satisfy all the constraints for an unbalanced design. Therefore, the approximated power is always higher than the unknown true power when the design is unbalanced.

A 3-level unbalanced design example: exact solution

Assume an engineer wants to compare the performance of three different materials. 4 samples are available for material A, 5 samples for material B and 13 samples for material C. The responses of different materials follow a normal distribution with a standard deviation of $\sigma = 1$. The engineer is required to calculate the power of detecting difference of 1 σ among all the level means at a significance level of 0.05.

From the design matrix of the test, $X^T X$ and Σ_β^{-1} are calculated as:

$$X^T X = \begin{pmatrix} 22 & -9 & -8 \\ -9 & 17 & 13 \\ -8 & 13 & 18 \end{pmatrix},$$

$$\Sigma_\beta^{-1} = \begin{pmatrix} 13.31818 & 9.727273 \\ 9.727273 & 15.09091 \end{pmatrix}$$

There are 3 paired comparisons. They are $\mu_1 - \mu_2$, $\mu_1 - \mu_3$ and $\mu_2 - \mu_3$.

If the first comparison $\mu_1 - \mu_2$ has the largest level mean difference of 1 σ, then the optimization problem becomes:

$$\begin{aligned} &\min \ \phi_1 = \beta \Sigma_\beta^{-1} \beta^T \\ &\textit{subject to } \beta_1 - \beta_2 = 1;\ |2\beta_1 + \beta_2| \le 1;\ |2\beta_2 + \beta_1| \le 1 \end{aligned}$$

The optimal solution is $\beta_1 = 0.51852;\ \beta_2 = -0.48148$, and the optimal $\phi_1 = 2.22222$.

If the second comparison $\mu_1 - \mu_3$ has the largest level mean difference, then the optimization is similar to the above problem. The optimal solution is $\beta_1 = 0.588235$; $\beta_2 = -0.17647$ and the optimal $\phi_2 = 3.058824$.

If the third comparison $\mu_2 - \mu_3$ has the largest level mean difference, then the optimal solution is $\beta_1 = -0.14815$; {{\beta }_{2}}=0.57407\,\!</math> and the optimal $\phi_3 = 3.61111$.

From the definition of power, we know that the power of a design should be calculated using the smallest non-centrality parameter of all possible outcomes. In this example, it is $\phi = \min(\phi_i) = 2.22222$. Since the significance level is 0.05, the critical value for the F test is $f_{citical} = F_{2,19}^{-1}(0.05) = 3.52189$. The power for this example is:

$$Power = 1 - F_{2,19}(f_{critical} | \phi = 2.22222) = 0.2161$$

A 3-level unbalanced design example: approximated solution

For the above example, we can get the approximated power by using the optimal β of a balanced design. If the design is balanced, the optimal solution will be:

Solution ID	Paired Comparison	β1	β2
1	u1-u2	0.5	-0.5
2	u1-u3	0.5	0
3	u2-u3	0	0.5

Therefore:

$$\mathbf{B} = \begin{pmatrix} 0.5 & -0.5 \\ 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$$

Since the design is unbalanced, use Σ_β^{-1} from the above example to get:

$$\Phi = \beta \Sigma_\beta^{-1} \beta' \Delta^2 = \begin{pmatrix} 2.238636 & 0.897727 & -1.34091 \\ 0.897727 & 3.329545 & 2.431818 \\ -1.34091 & 2.431818 & 3.772727 \end{pmatrix}$$

The smallest ϕ_i is 2.238636. For this example, it is very close to the exact solution 2.22222 given in the previous calculation. The approximated power is:

$$Power = 1 - F_{2,19}(f_{critical} | \phi = 2.238636) = 0.2174$$

This result is a little larger than the exact solution of 0.2162.

In practical cases, the above method can be applied to quickly check the power of a design. If the calculated power cannot meet the required value, the true power definitely will not meet the

requirement, since the calculated power using this procedure is always equal to (for balanced designs) or larger than (for unbalanced designs) the true value.

The result in the DOE folio for this example is given as:

Power Study		
	Degrees of Freedom	**Power for Max Difference = 1**
A:Factor 1	2	0.2174
Residual	19	-

Power Study for 2 Level Factorial Designs

For 2 level factorial designs, each factor (effect) has only one coefficient. The linear regression model is:

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \beta_3 X_{i,3} + \ldots + \beta_{12} X_{i,1} X_{i,2} + \ldots + \varepsilon_i$$

The model can include main effect terms and interaction effect terms. Each X_i can be -1 (the low level) or +1 (the high level). The effect of a main effect term is defined as the difference of the mean value of Y at $X_i = +1$ and $X_i = -1$. Please notice that all the factor values here are coded values. For example, the effect of X_1 is defined by:

$$Y(X_1 = 1) - Y(X_1 = -1) = 2\beta_1$$

Similarly, the effect of an interaction term is also defined as the difference of the mean values of Y at the interaction terms of +1 and -1. For example, the effect of $X_1 X_2$ is:

$$Y(X_1 X_2 = 1) - Y(X_1 X_2 = -1) = 2\beta_{12}$$

Therefore, if the effect of a term that we want to calculate the power for is Δ_i , then the corresponding coefficient β_i must be $\Delta_i/2$. Therefore, the non-centrality parameter for each term in the model for a 2 level factorial design can be calculated as

$$\phi_i = \frac{\beta_i^2}{Var(\beta_i)} = \frac{\Delta_i^2}{4Var(\beta_i)}$$

Once ϕ_i is calculated, we can use it to calculate the power. If the design is balanced, the power of terms with the same order will be the same. In other words, all the main effects have the same power and all the k-way (k=2, 3, 4, ...) interactions have the same power.

Example: Due to the constraints of sample size and cost, an engineer can run only the following 13 tests for a 4 factorial design:

Run	A	B	C	D
1	1	1	1	1
2	1	1	-1	-1
3	1	-1	1	-1
4	-1	1	1	-1
5	-1	1	-1	1
6	-1	-1	1	1
7	-1	-1	-1	-1
8	0	0	0	0
9	0	0	0	0
10	0	0	0	0
11	0	0	0	0
12	0	0	0	0
13	0	0	0	0

Before doing the tests, he wants to evaluate the power for each main effect. Assume the amount of effect he wants to perform a power calculation for is 2 σ . The significance level is 0.05.

Step 1: Calculate the variance and covariance matrix for the model coefficients. The main effect-only model is:

$$Y_i = \beta_0 + \beta_1 A_i + \beta_2 B_i + \beta_3 C_i + \beta_4 D_i + \varepsilon_i$$

For this model:

$$Var(\beta) = \sigma^2 (X'X)^{-1}$$

The value for $(X'X)^{-1}$ is

	beta0	beta1	beta2	beta3	beta4
beta0	0.083333	0.020833	-0.02083	-0.02083	0.020833
beta1	0.020833	0.161458	-0.03646	-0.03646	0.036458

beta2	-0.02083	-0.03646	0.161458	0.036458	-0.03646
beta3	-0.02083	-0.03646	0.036458	0.161458	-0.03646
beta4	0.020833	0.036458	-0.03646	-0.03646	0.161458

The diagonal elements are the variances for the coefficients.

Step 2: Calculate the non-centrality parameter for each term. In this example, all the main effect terms have the same variance, so they have the same non-centrality parameter value.

$$\phi_i = \frac{\Delta_i^2}{4Var(\beta_i)} = \frac{1}{0.161458} = 6.19355$$

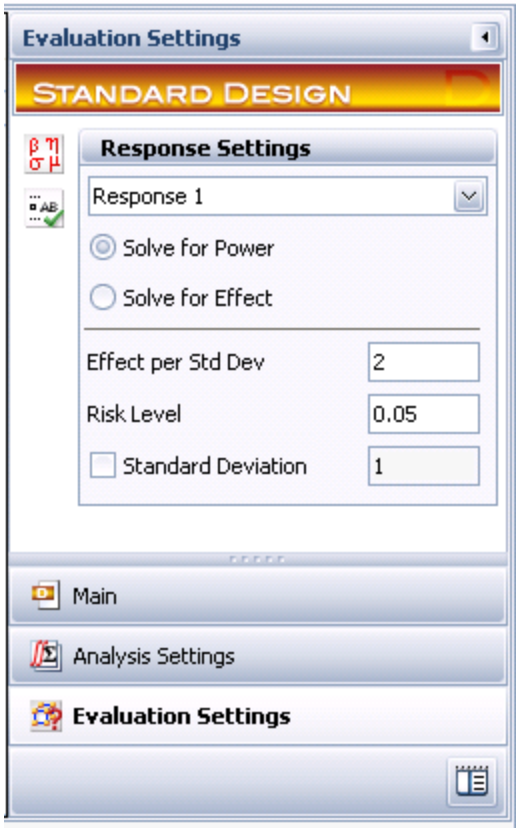
Step 3: Calculate the critical value for the F test. It is:

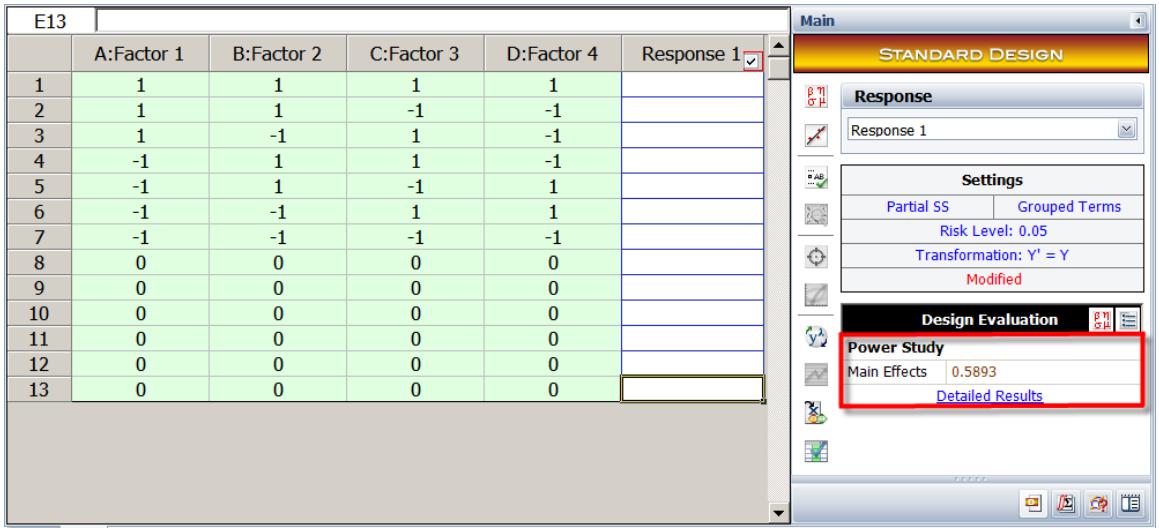
$$f_{citical} = F_{1,8}^{-1}(0.05) = 5.31766$$

Step 4: Calculate the power for each main effect term. For this example, the power is the same for all of them:

$$Power = 1 - F_{1,8}\left(f_{critical} | \phi = 6.19355\right) = 0.58926$$

The settings and results in the DOE folio are given below.





In general, the calculated power for each term will be different for unbalanced designs. However, the above procedure can be applied for both balanced and unbalanced 2 level factorial designs.

Power Study for General Level Factorial Designs

For a quantitative factor X with more than 2 levels, its effect is defined as:

$$Y(X_i = 1) - Y(X_i = -1) = 2\beta_i$$

This is the difference of the expected Y values at its defined high and low level. Therefore, a quantitative factor can always be treated as a 2 level factor mathematically, regardless of its defined number of levels. A quantitative factor has only 1 term in the regression equation.

For a qualitative factor with more than 2 levels, it has more than 1 term in the regression equation. Like in the multiple level 1 factor designs, a qualitative factor with r levels will have r-1 terms in the linear regression equation. Assume there are 2 factors in a design. Factor A has 3 levels and factor B has 3 levels, the regression equation for this design is:

$$Y = \beta_0 + \beta_1 A[1] + \beta_2 A[2] + \beta_3 B[1] + \beta_4 B[2] + \beta_{11} A[1]B[1] + \beta_{12} A[1]B[2] + \beta_{21} A[2]B[1] + \beta_{22} A[2]B[2]$$

There are 2 regression terms for each main effect, and 4 regression terms for the interaction effect. We will use the above equation to explain how the power for the main effects and interaction effects is calculated in the DOE folio. The following balanced design is used for the calculation:

Run	A	B	Run	A	B
1	1	1	14	2	2
2	1	2	15	2	3
3	1	3	16	3	1
4	2	1	17	3	2
5	2	2	18	3	3
6	2	3	19	1	1
7	3	1	20	1	2
8	3	2	21	1	3
9	3	3	22	2	1
10	1	1	23	2	2
11	1	2	24	2	3
12	1	3	25	3	1
13	2	1	26	3	2
			27	3	3

Power Study for Main Effects

Let's use factor A to show how the power is defined and calculated for the main effects. For the above design, if we ignore factor B, then it becomes a 1 factor design with 9 samples at each level. Therefore, the same linear regression model and power calculation method as discussed for 1 factor designs can be used to calculate the power for the main effects for this multiple level factorial design. Since A has 3 levels, it has 3 paired comparisons: $\Delta_{12} = \mu_1 - \mu_2$; $\Delta_{13} = \mu_1 - \mu_3$ and $\Delta_{23} = \mu_2 - \mu_3$. μ_i is the average of the responses at the ith level. However, these three contrasts are not independent, since $\Delta_{12} = \Delta_{13} - \Delta_{23}$. We are interested in the largest difference among all the contrasts. Let $\Delta = \max(\Delta_{ij})$. Power is defined as the probability of detecting a given Δ in an experiment. Using the linear regression equation, we get:

$$\Delta_{12} = \beta_1 - \beta_2;\ \Delta_{13} = 2\beta_1 + \beta_2;\ \Delta_{23} = 2\beta_2 + \beta_1$$

Just as for the 1 factor design, we know the optimal solutions are: $\beta_1 = 0.5\Delta, \beta_2 = -0.5\Delta$ when Δ_{12} is the largest difference Δ ; $\beta_1 = 0.5\Delta, \beta_2 = 0$ when Δ_{13} is the largest difference Δ and $\beta_1 = 0, \beta_2 = 0.5\Delta$ when Δ_{23} is the largest difference Δ . For each of the solution, a non-centrality parameter can be calculated using $\phi_i = \beta\Sigma_\beta^{-1}\beta^T$. Here $\beta = (\beta_1, \beta_2)$, and Σ_β^{-1} is the inverse of the variance and covariance matrix obtained from the linear regression model when all the terms are included. For this example, we have the coefficient matrix for the optimal solution:

$$B = \Delta \begin{pmatrix} 0.5 & -0.5 \\ 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$$

The standard variance matrix $(X'X)^{-1}$ for all the coefficients is:

I	A[1]	A[2]	B[1]	B[2]	A[1]B[1]	A[1]B[2]	A[2]B[1]	A[2]B[2]
0.0370	0	0	0	0	0	0	0	0
0	0.0741	-0.0370	0	0	0	0	0	0
0	-0.0370	0.0741	0	0	0	0	0	0
0	0	0	0.0741	-0.0370	0	0	0	0
0	0	0	-0.0370	0.0741	0	0	0	0
0	0	0	0	0	0.1481	-0.0741	-0.0741	0.0370
0	0	0	0	0	-0.0741	0.1481	0.0370	-0.0741

0	0	0	0	0	-0.0741	0.0370	0.1481	-0.0741
0	0	0	0	0	0.0370	-0.0741	-0.0741	0.1481

Clearly the design is balanced for all the terms since the above matrix is a block diagonal matrix.

From the above table, we know the variance and covariance matrix Σ_β of A is:

$$\Sigma_\beta = \sigma^2 \begin{pmatrix} 0.0741 & -0.0370 \\ -0.0370 & 0.0741 \end{pmatrix}$$

Its inverse Σ_β^{-1} for factor A is:

$$\Sigma_\beta^{-1} = \frac{1}{\sigma^2}\begin{pmatrix} 0.0741 & -0.0370 \\ -0.0370 & 0.0741 \end{pmatrix}^{-1} = \frac{1}{\sigma^2}\begin{pmatrix} 18 & 9 \\ 9 & 18 \end{pmatrix}$$

Assuming that the Δ we are interested in is σ, then the calculated non-centrality parameters are:

$$\Phi = \beta\Sigma_\beta^{-1}\beta'\Delta^2$$

=

$$\begin{matrix} 4.5 & 2.25 & -2.25 \\ 2.25 & 4.5 & 2.25 \\ -2.25 & 2.25 & 4.5 \end{matrix}$$

The power is calculated using the smallest value at the diagonal of the above matrix. Since the design is balanced, all the 3 non-centrality parameters are the same in this example (i.e., they are 4.5).

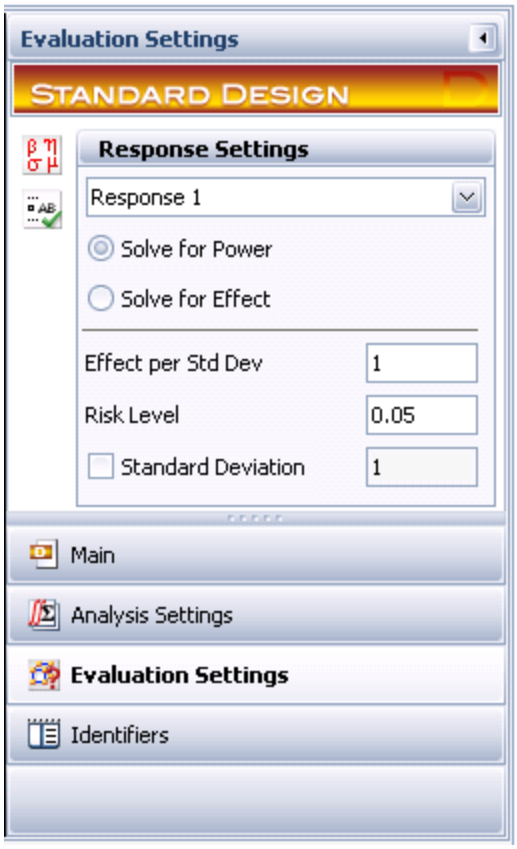
The critical value for the F test is:

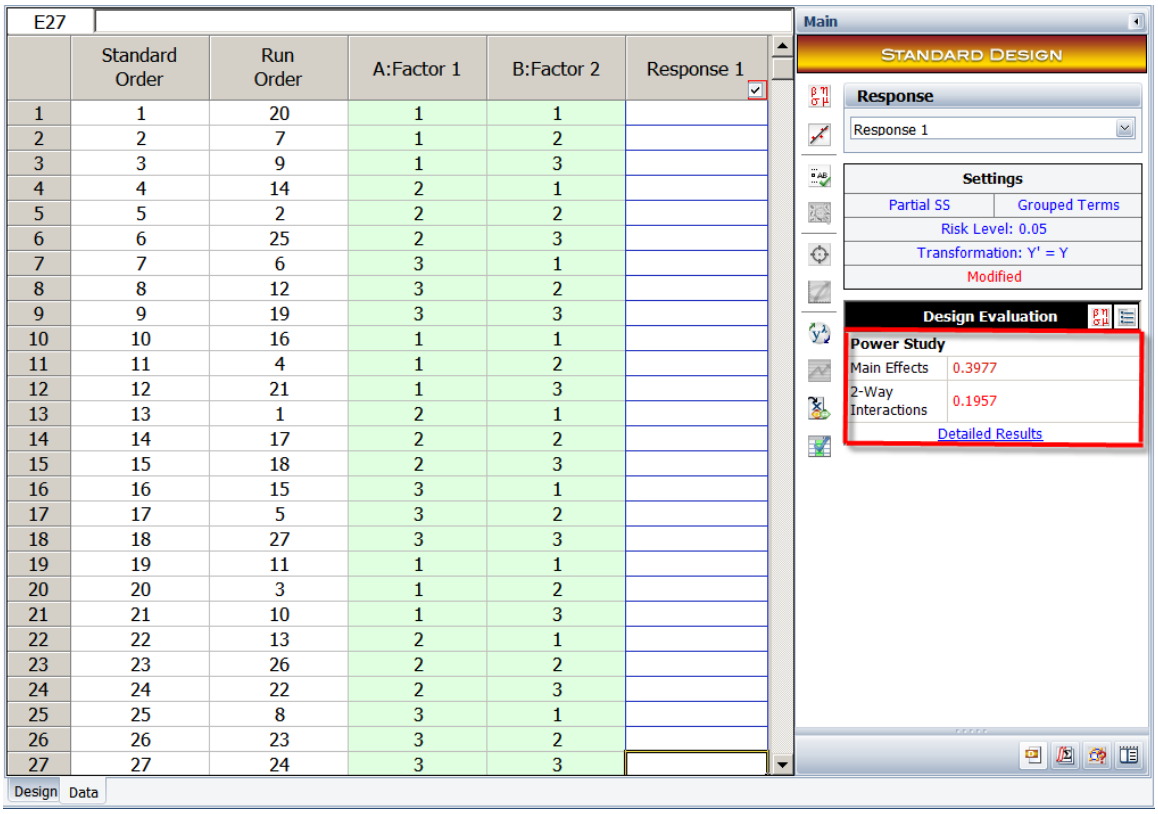
$$f_{citical} = F_{2,18}^{-1}(0.05) = 3.55456$$

Please notice that for the F distribution, the first degree of freedom is 2 (the number of terms for factor A in the regression model) and the 2nd degree of freedom is 18 (the degrees of freedom of error).

The power for main effect A is:

$$Power = 1 - F_{2,18}\left(f_{critical} | \phi = 4.5\right) = 0.397729$$





If the Δ we are interested in is 2σ, then the non-centrality parameter will be 18. The power for main effect A is:

$$Power = 1 - F_{2,18}\left(f_{critical} | \phi = 18\right) = 0.9457$$

The power is greater for a larger Δ. The above calculation also can be used for unbalanced designs to get the approximated power.

Power Study for Interaction Effects

First, we need to define what an "interaction effect" is. From the discussion for 2 level factorial designs, we know the interaction effect AB is defined by:

$$Y\left(AB = 1\right) - Y\left(AB = -1\right) = 2\beta_{12}$$

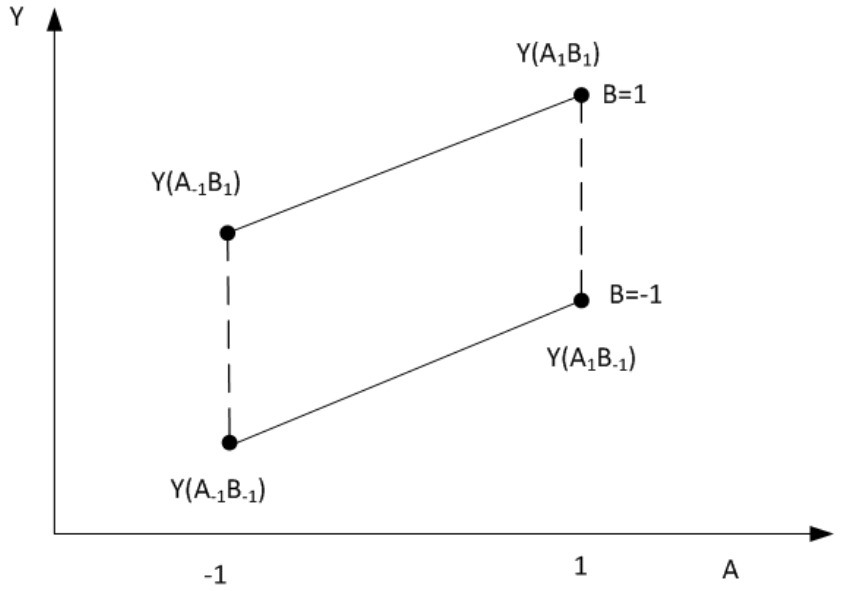
It is the difference between the average response at AB=1 and AB=-1. The above equation also can be written as:

$$\frac{Y\left(A_1B_1\right) + Y\left(A_{-1}B_{-1}\right)}{2} - \frac{Y\left(A_{-1}B_1\right) + Y\left(A_1B_{-1}\right)}{2}$$

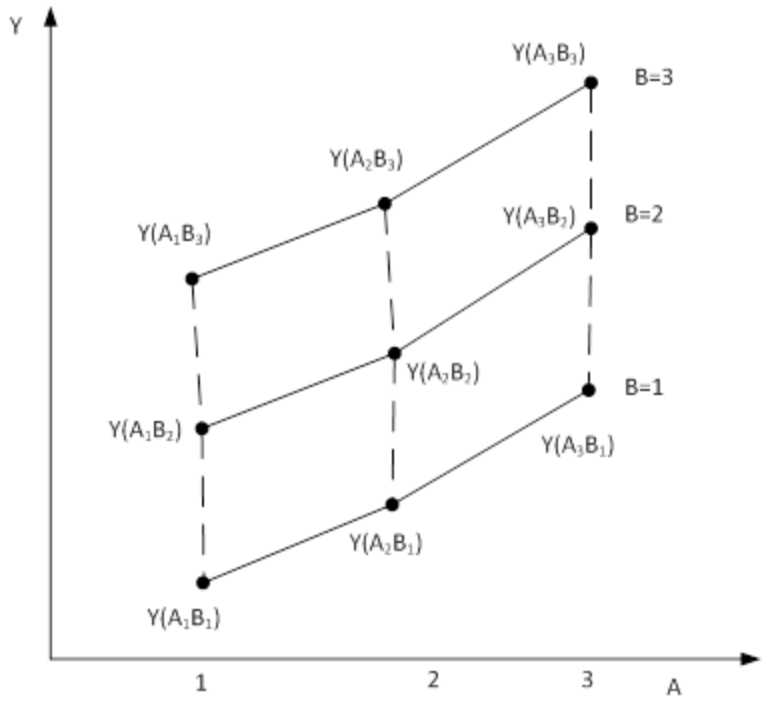
or:

$$\frac{Y(A_1B_1) - Y(A_1B_{-1})}{2} - \frac{Y(A_{-1}B_1) - Y(A_{-1}B_{-1})}{2}$$
$$= \frac{\text{Effect of B at A=1}}{2} - \frac{\text{Effect of B at A=-1}}{2}$$

From here we can see that the effect of AB is half of the difference of the effect of B when A is fixed at 1 and the effect of B when A is fixed at -1. Therefore, a two-way interaction effect is calculated using 4 points as shown in the above equation. This is illustrated in the following figure.



As we discussed before, a main effect is defined by two points. For example, the main effect of B at A=1 is defined by $Y(A_1B_1)$ and $Y(A_1B_{-1})$. The above figure clearly shows that a two-way interaction effect of two-level factors is defined by the 4 vertex of a quadrilateral. How can we define the two-way interaction effects of factorials with more than two levels? For example, for the design used in the previous section, A and B are both three levels. What is the interaction effect AB? For this example, the 9 design points are shown in the following figure.



Notice that there are 9 quadrilaterals in the above figure. These 9 contrasts define the interaction effect AB. This is similar to the paired comparisons in a one factorial design with multiple levels, where a main effect is defined by a group of contrasts (or paired comparisons). For the design in the above figure, to construct a quadrilateral, we need to choose 2 levels from A and 2 levels from B. There are $\binom{3}{2} \times \binom{3}{2} = 9$ combinations. Therefore, we see the following 9 contrasts.

Contrast ID	A	B
1	(1, 2)	(1, 2)
2	(1, 2)	(1, 3)
3	(1, 2)	(2, 3)
4	(1, 3)	(1, 2)
5	(1, 3)	(1, 3)
6	(1, 3)	(2, 3)
7	(2, 3)	(1, 2)
8	(2, 3)	(1, 3)
9	(2, 3)	(2, 3)

Let's use the first contrast to explain the meaning of a contrast. (1, 2) in column A means the selected levels from A are 1 and 2. (1, 2) in column B means the selected levels from B are 1 and 2. They form 4 points: $Y(A_1B_1)$, $Y(A_1B_2)$, $Y(A_2B_1)$ and $Y(A_2B_2)$. We can denote the AB effect defined by this contrast as Δ_1.

$$\Delta_1 = \frac{Y(A_1B_1) + Y(A_2B_2)}{2} - \frac{Y(A_1B_2) + Y(A_2B_1)}{2}$$
$$= \frac{Y(A_1B_1) - Y(A_1B_2) - Y(A_2B_1) + Y(A_2B_2)}{2}$$

In general, if a contrast is defined by A (i, j) and B(i', j'), then the effect is calculated by:

$$\Delta_{AB} = \frac{Y(A_iB_j) - Y(A_iB_{j'}) - Y(A_{i'}B_j) + Y(A_{i'}B_{j'})}{2}$$

From the above two equations we can see that the two-way interaction effect AB is defined as the difference of the main effect of B at A = i and the main effect of B at A = j. This logic can be easily extended to three-way interactions. For example ABC can be defined as the difference of AB at C=k and AC at C=k'. Its calculation is:

$$\Delta_{ABC} = \frac{Y(A_iB_jC_k) - Y(A_iB_{j'}C_k) - Y(A_{i'}B_jC_k) + Y(A_{i'}B_{j'}C_k)}{4}$$
$$- \frac{Y(A_iB_jC_{k'}) - Y(A_iB_{j'}C_{k'}) - Y(A_{i'}B_jC_{k'}) + Y(A_{i'}B_{j'}C_{k'})}{4}$$

For a design with A, B, and C with 3 levels, there are $\binom{3}{2} \times \binom{3}{2} \times \binom{3}{2} = 27$ contrast for the three-way interaction ABC.

Similarly, the above method can be extended for higher order interactions. By now, we know the main effect and interactions for multiple level factorial designs are defined by a group of contrasts. We will discuss how the power of these effects is calculated in the following section.

The power for an effect is defined as follows: when the largest value of a contrast group for an effect is Δ, power is the smallest probability of detecting this Δ among all the possible outcomes at a given significance level.

To calculate the power for an effect, as in the previous sections, we need to relate a contrast with model coefficients. The 9 contrasts in the above table can be expressed using model coefficients. For example:

$$\Delta_1 = \frac{\beta_{11} + \beta_{22}}{2} - \frac{\beta_{12} + \beta_{21}}{2}$$

If this contrast has the largest value Δ, the power is calculated from the following optimization problem:

$$\min \ \phi_1 = \beta \Sigma_\beta^{-1} \beta^T$$
$$subject\ to\ \frac{\beta_{11}+\beta_{22}}{2} - \frac{\beta_{12}+\beta_{21}}{2} = \Delta_1 = \Delta; |\Delta_i| \le \Delta, i = 2, \ldots, 9$$

where $\beta = (\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22})$, and Σ_β^{-1} is the variance and covariance matrix of β.

For a balanced general level factorial design such as this example, the optimal solution for the above optimization issue is:

$$\beta = (\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}) = (0.5, -0.5, -0.5, 0.5)$$

For all the 9 contrasts, by assuming each of the contrasts has the largest value Δ one by one, we can get 9 optimal solutions and 9 non-centrality parameters ϕ_i. The power for the interaction effect AB is calculated using the min(ϕ_i). The 9 optimal solutions are:

Contrast ID	A	B	β_{11}	β_{12}	β_{21}	β_{22}
1	(1, 2)	(1, 2)	0.5	-0.5	-0.5	0.5
2	(1, 2)	(1, 3)	0.5	0	-0.5	0
3	(1, 2)	(2, 3)	0	0.5	0	-0.5
4	(1, 3)	(1, 2)	0.5	-0.5	0	0
5	(1, 3)	(1, 3)	0.5	0	0	0
6	(1, 3)	(2, 3)	0	0.5	0	0
7	(2, 3)	(1, 2)	0	0	0.5	-0.5
8	(2, 3)	(1, 3)	0	0	0.5	0
9	(2, 3)	(2, 3)	0	0	0	0.5

In the regression equation for this example, there are 4 terms for AB effect. Therefore there are 4 independent contrasts in the above table. These are contrasts 5, 6, 8, and 9. The rest of the contrasts are linear combinations of these 4 contrasts. Based on the calculation in the main effect section, we know that the standard variance matrix $(X'X)^{-1}$ for all the coefficients is:

I	A[1]	A[2]	B[1]	B[2]	A[1]B[1]	A[1]B[2]	A[2]B[1]	A[2]B[2]
0.0370	0	0	0	0	0	0	0	0

0	0.0741	-0.0370	0	0	0	0	0	0
0	-0.0370	0.0741	0	0	0	0	0	0
0	0	0	0.0741	-0.0370	0	0	0	0
0	0	0	-0.0370	0.0741	0	0	0	0
0	0	0	0	0	0.1481	-0.0741	-0.0741	0.0370
0	0	0	0	0	-0.0741	0.1481	0.0370	-0.0741
0	0	0	0	0	-0.0741	0.0370	0.1481	-0.0741
0	0	0	0	0	0.0370	-0.0741	-0.0741	0.1481

The variance and covariance matrix Σ_β of AB is:

$$\Sigma_\beta = \sigma^2 \begin{pmatrix} 0.1481 & -0.0741 & -0.0741 & 0.0370 \\ -0.0741 & 0.1481 & 0.0370 & -0.0741 \\ -0.0741 & 0.0370 & 0.1481 & -0.0741 \\ 0.0370 & -0.0741 & -0.0741 & 0.1481 \end{pmatrix}$$

Then its inverse matrix Σ_β^{-1} is:

$$\Sigma_\beta^{-1} = \frac{1}{\sigma^2} \begin{pmatrix} 0.1481 & -0.0741 & -0.0741 & 0.0370 \\ -0.0741 & 0.1481 & 0.0370 & -0.0741 \\ -0.0741 & 0.0370 & 0.1481 & -0.0741 \\ 0.0370 & -0.0741 & -0.0741 & 0.1481 \end{pmatrix}^{-1} = \frac{1}{\sigma^2} \begin{pmatrix} 12.0256 & 6.0250 & 6.0250 & 3.0247 \\ 6.0250 & 12.0256 & 3.0247 & 6.0250 \\ 6.0250 & 3.0247 & 12.0256 & 6.0250 \\ 3.0247 & 6.0250 & 6.0250 & 12.0256 \end{pmatrix}$$

Assuming that the Δ we are interested in is σ , then the calculated non-centrality parameters for all the contrasts are the diagonal elements of the following matrix.

$$\Phi = \beta \Sigma_\beta^{-1} \beta' \Delta^2$$

=

3.0003	1.5002	-1.5002	1.5002	0.7501	-0.7501	-1.5002	-0.7501	0.7501
1.5002	3.0003	1.5002	0.7501	1.5002	0.7501	-0.7501	-1.5002	-0.7501
1.5002	1.5002	3.0003	-0.7501	0.7501	1.5002	0.7501	-0.7501	-1.5002

1.5002	0.7501	-0.7501	3.0003	1.5002	-1.5002	1.5002	0.7501	-0.7501
0.7501	1.5002	0.7501	1.5002	3.0064	1.5062	0.7501	1.5062	0.7562
0.7501	0.7501	1.5002	-1.5002	1.5062	3.0064	-0.7501	0.7562	1.5062
1.5002	-0.7501	0.7501	1.5002	0.7501	-0.7501	3.0003	1.5002	-1.5002
0.7501	-1.5002	-0.7501	0.7501	1.5062	0.7562	1.5002	3.0064	1.5062
0.7501	-0.7501	-1.5002	-0.7501	0.7562	1.5062	-1.5002	1.5062	3.0064

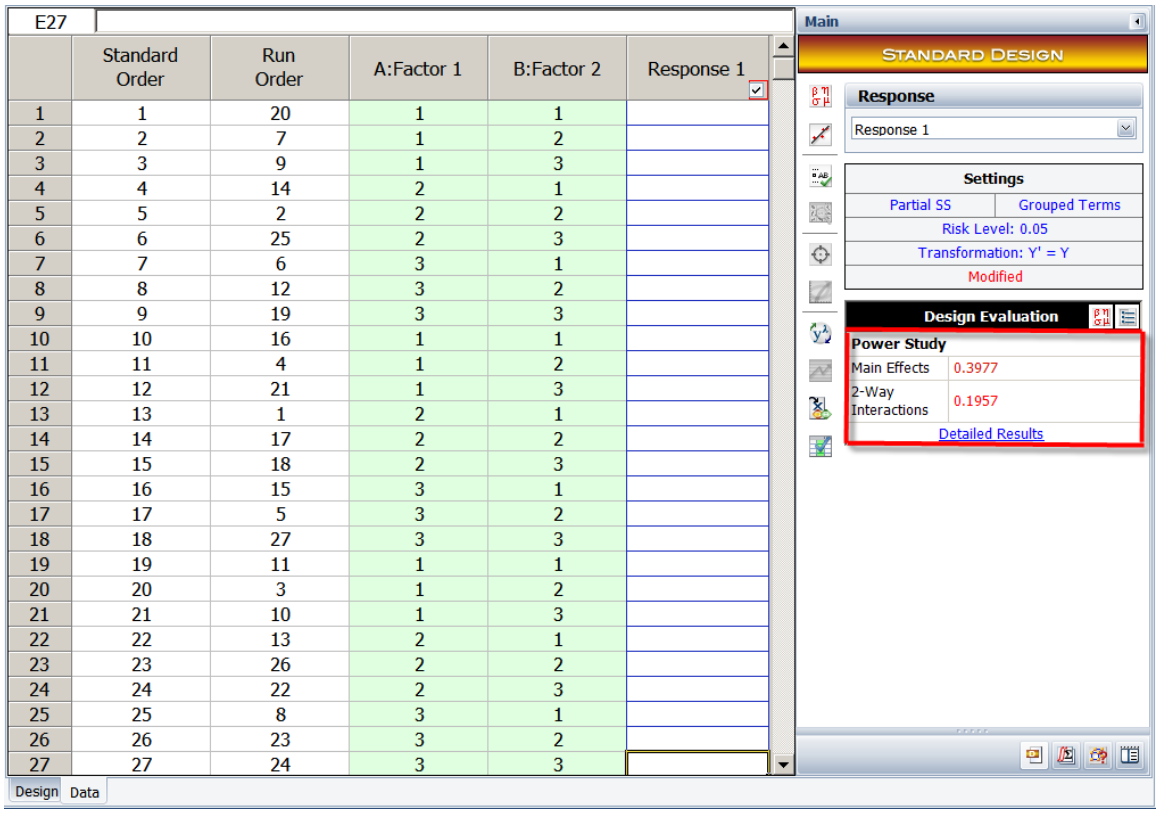
The power is calculated using the smallest value at the diagonal of the above matrix (i.e., 3.0003). The critical value for the F test is:

$$f_{citical} = F_{4,18}^{-1}(0.05) = 2.927744$$

Please notice that for the F distribution, the first degree of freedom is 4 (the number of terms for effect AB in the regression model) and the 2nd degree of freedom is 18 (the degree of freedom of error).

The power for AB is:

$$Power = 1 - F_{4,18}\left(f_{critical} | \phi = 3.0003\right) = 0.1957$$

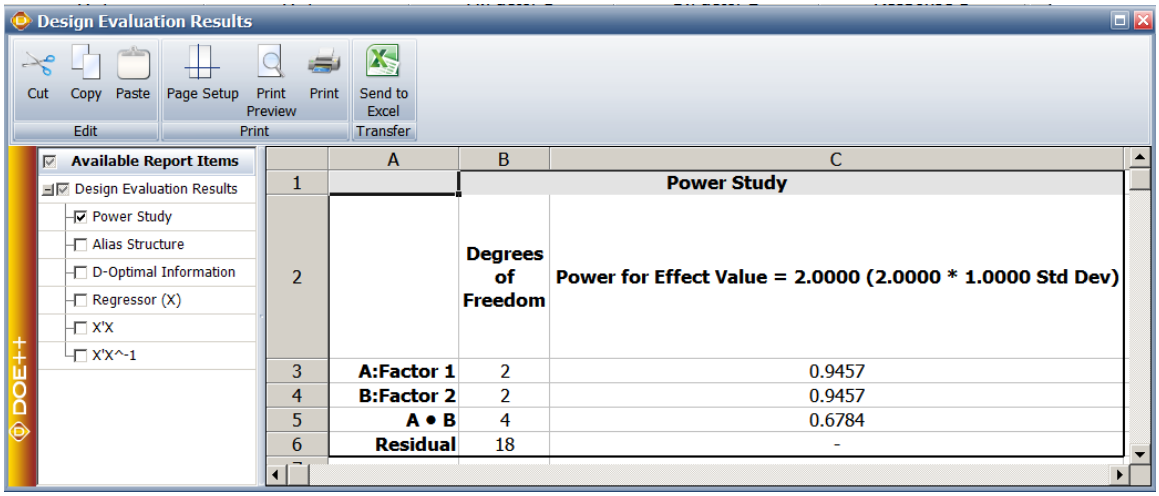


	Standard Order	Run Order	A:Factor 1	B:Factor 2	Response 1
1	1	20	1	1	
2	2	7	1	2	
3	3	9	1	3	
4	4	14	2	1	
5	5	2	2	2	
6	6	25	2	3	
7	7	6	3	1	
8	8	12	3	2	
9	9	19	3	3	
10	10	16	1	1	
11	11	4	1	2	
12	12	21	1	3	
13	13	1	2	1	
14	14	17	2	2	
15	15	18	2	3	
16	16	15	3	1	
17	17	5	3	2	
18	18	27	3	3	
19	19	11	1	1	
20	20	3	1	2	
21	21	10	1	3	
22	22	13	2	1	
23	23	26	2	2	
24	24	22	2	3	
25	25	8	3	1	
26	26	23	3	2	
27	27	24	3	3	

If the Δ we are interested in is 2σ, then the non-centrality parameter will be 12.0012. The power for main effect A is:

$$Power = 1 - F_{4,18}\left(f_{critical} | \phi = 12.0012\right) = 0.6784$$

The power values for all the effects in the model are:



	A	B	C
1		Power Study	
2		Degrees of Freedom	Power for Effect Value = 2.0000 (2.0000 * 1.0000 Std Dev)
3	A:Factor 1	2	0.9457
4	B:Factor 2	2	0.9457
5	A • B	4	0.6784
6	Residual	18	-

For balanced designs, the above calculation gives the exact power. For unbalanced design, the above method will give the approximated power. The true power is always less than the approximated value.

This section explained how to use a group of contrasts to represent the main and interaction effects for multiple level factorial designs. Examples for main and 2nd order interactions were provided. The power calculation for higher order interactions is the same as the above example. Therefore, it is not repeated here.

Power Study for Response Surface Method Designs

For response surface method designs, the following linear regression model is used:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2 + \ldots + \varepsilon$$

The above equations can have both qualitative and quantitative factors. As we discussed before, for each effect (main or quadratic effect) of a quantitative factor, there is only one term in the regression model. Therefore, the power calculation for a quantitative factor is the same as treating this factor as a 2 level factor, no matter how many levels are defined for it. If qualitative factors are used in the design, they do not have quadratic effects in the model. The power calculation for qualitative factors is the same as discussed in the previous sections.

First we need to define what the "effect" is for each term in the above linear regression equation. The definition for main effects and interaction effects is the same as for 2 level factorial designs. The effect is defined as the difference of the average response at the +1 of the term and at the -1 of the term. For example, the main effect of X_i is:

$$\Delta_i = Y(X_i = 1) - Y(X_i = -1) = 2\beta_i$$

The interaction effect of $X_i X_j$ is:

$$\Delta_{ij} = Y(X_i X_j = 1) - Y(X_i X_j = -1) = 2\beta_{ij}$$

For a quadratic term X_i^2, its range is from 0 to 1. Therefore, its effect is:

$$\Delta_{ii} = Y\left(X_i^2 = 1\right) - Y\left(X_i^2 = 0\right) = \beta_{ii}$$

The quadratic term also can be thought of as:

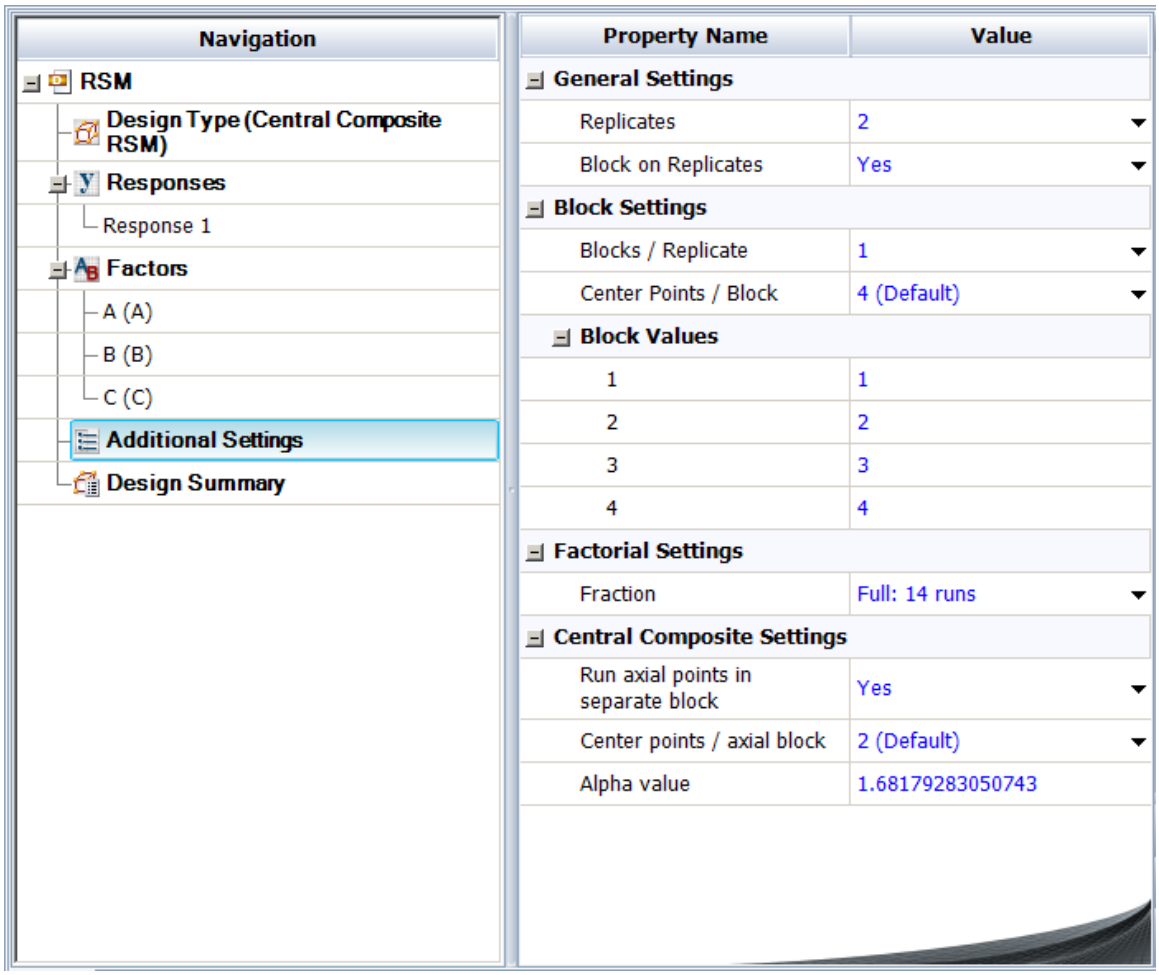
$$\Delta_{ii} = \frac{Y(X_i = 1) + Y(X_i = -1)}{2} - Y(X_i = 0) = \beta_{ii}$$

Since there are no grouped contrasts for each effect, the power can be calculated using either the non-central t distribution or the non-central F distribution. They will lead to the same results. Let's use the following design to illustrate the calculation.

Run	Block	A	B	C
1	1	-1	-1	-1
2	1	1	-1	-1
3	1	-1	1	-1
4	1	1	1	-1
5	1	-1	-1	1
6	1	1	-1	1
7	1	-1	1	1
8	1	1	1	1
9	1	0	0	0
10	1	0	0	0
11	1	0	0	0
12	1	0	0	0
13	2	-1.68179	0	0
14	2	1.681793	0	0
15	2	0	-1.68179	0
16	2	0	1.681793	0
17	2	0	0	-1.68179
18	2	0	0	1.681793
19	2	0	0	0
20	2	0	0	0
21	3	-1	-1	-1
22	3	1	-1	-1
23	3	-1	1	-1
24	3	1	1	-1
25	3	-1	-1	1

26	3	1	-1	1
27	3	-1	1	1
28	3	1	1	1
29	3	0	0	0
30	3	0	0	0
31	3	0	0	0
32	3	0	0	0
33	4	-1.68179	0	0
34	4	1.681793	0	0
35	4	0	-1.68179	0
36	4	0	1.681793	0
37	4	0	0	-1.68179
38	4	0	0	1.681793
39	4	0	0	0
40	4	0	0	0

The above design can be created in a DOE folio using the following settings:



The model used here is:

$$Y = \beta_0 + \beta_{b1} BLK[1] + \beta_{b2} BLK[2] + \beta_{b3} BLK[3] + \beta_1 A + \beta_2 B + \beta_3 C + \beta_{12} AB + \beta_{13} AC + \beta_{23} BC + \beta_{11} A^2 + \beta_{22} B^2 + \beta_{33} C^2$$

Blocks are included in the model. Since there are four blocks, three indicator variables are used. The standard variance and covariance matrix $(X'X)^{-1}$ is

Con-st	BL-K1	BL-K2	BL-K3	A	B	C	AB	AC	BC	AA	BB	CC
0.08-5018	-0.00-694	0.00-6944	-0.00-694	0	0	0	0	0	0	-0.02-862	-0.02-862	-0.02-862

0.00-694	0.06-7759	-0.02-609	-0.01-557	0	0	0	0	0	0	0.00-0843	0.00-0843	0.00-0843
0.00-6944	-0.02-609	0.08-8593	-0.02-609	0	0	0	0	0	0	-0.00-084	-0.00-084	-0.00-084
0.00-694	-0.01-557	-0.02-609	0.06-7759	0	0	0	0	0	0	0.00-0843	0.00-0843	0.00-0843
0	0	0	0	0.03-6612	0	0	0	0	0	0	0	0
0	0	0	0	0	0.03-6612	0	0	0	0	0	0	0
0	0	0	0	0	0	0.03-6612	0	0	0	0	0	0
0	0	0	0	0	0	0	0.0-625	0	0	0	0	0
0	0	0	0	0	0	0	0	0.0-625	0	0	0	0
0	0	0	0	0	0	0	0	0	0.0-625	0	0	0
0.02-862	0.00-0843	-0.00-084	0.00-0843	0	0	0	0	0	0	0.03-4722	0.00-3472	0.00-3472
0.02-862	0.00-0843	-0.00-084	0.00-0843	0	0	0	0	0	0	0.00-3472	0.03-4722	0.00-3472

The variances for all the coefficients are the diagonal elements in the above matrix. These are:

Term	Var(σ^2)
A	0.036612
B	0.036612
C	0.036612
AB	0.0625

AC	0.0625
BC	0.0625
AA	0.034722
BB	0.034722
CC	0.034722

Assume the value for each effect we are interested in is Δ . Then, to get this Δ the corresponding value for each model coefficient is:

Term	Coefficient
A	0.5Δ
B	0.5Δ
C	0.5Δ
AB	0.5Δ
AC	0.5Δ
BC	0.5Δ
AA	1Δ
BB	1Δ
CC	1Δ

The degrees of freedom used in the calculation are:

Source	Degree of Freedom
Block	3
A:A	1
B:B	1
C:C	1
AB	1
AC	1
BC	1
AA	1

BB	1
CC	1
Residual	27
Lack of Fit	19
Pure Error	8
Total	39

The above table shows all the factor effects have the same degree of freedom, therefore they have the same critical F value. For a significance level of 0.05, the critical value is:

$$f_{citical} = F_{1,27}^{-1}(0.05) = 4.210008$$

When $\Delta = 1\sigma$, the non-centrality parameter for each main effect is calculated by:

$$\phi_i = \frac{\beta_i^2}{Var(\beta_i)} = \frac{(0.5\Delta)^2}{Var(\beta_i)} = \frac{\sigma^2}{4Var(\beta_i)}$$

The non-centrality parameter for each interaction effect is calculated by:

$$\phi_{ij} = \frac{\beta_{ij}^2}{Var(\beta_{ij})} = \frac{(0.5\Delta)^2}{Var(\beta_{ij})} = \frac{\sigma^2}{4Var(\beta_{ij})}$$

The non-centrality parameter for each quadratic effect is calculated by:

$$\phi_{ii} = \frac{\beta_{ii}^2}{Var(\beta_{ii})} = \frac{(\Delta)^2}{Var(\beta_{ii})} = \frac{\sigma^2}{Var(\beta_{ii})}$$

All the non-centrality parameters are given in the following table:

Term	**Non-centrality parameter** (ϕ)
A	6.828362
B	6.828362
C	6.828362
AB	4
AC	4
BC	4
AA	28.80018

BB	28.80018
CC	28.80018

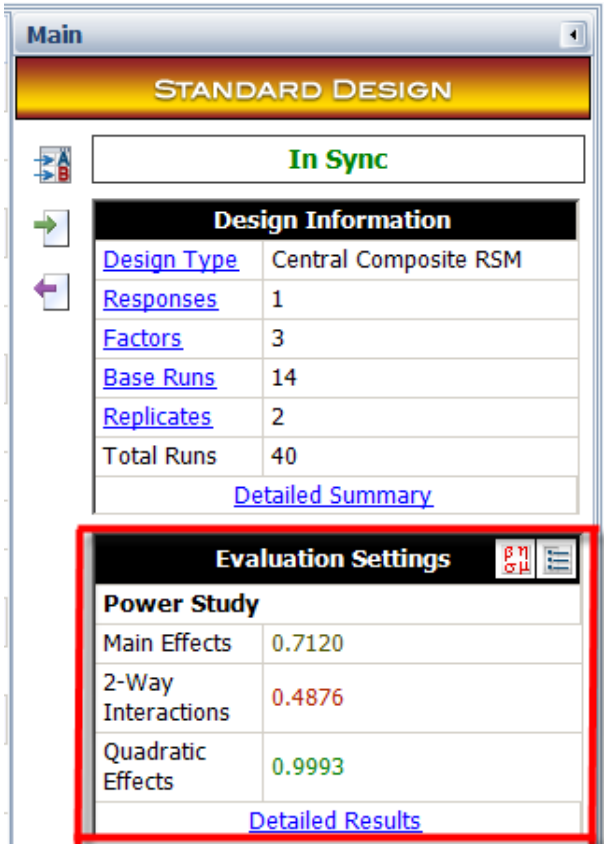
The power for each term is calculated by:

$$Power = 1 - F_{1,27}\left(f_{critical} | \phi_i\right)$$

They are:

Source	Power $(\Delta = 1\sigma)$
A:A	0.712033
B:B	0.712033
C:C	0.712033
AB	0.487574
AC	0.487574
BC	0.487574
AA	0.999331
BB	0.999331
CC	0.999331

The results in a DOE folio can be obtained from the design evaluation.



Discussion on Power Calculation

All the above examples show how to calculate the power for a given amount of effect. When a power value is given, using the above method we also can calculate the corresponding effect. If the power is too low for an effect of interest, the sample size of the experiment must be increased in order to get a higher power value.

We discussed in detail how to define an "effect" for quantitative and qualitative factors, and how to use model coefficients to represent a given effect. The power in a DOE folio is calculated based on this definition. Readers may find that power is calculated directly based on model coefficients (instead of the contrasts) in other software packages or books. However, for some cases, such as for the main and interaction effects of qualitative factors with multiple levels, the meaning of model coefficients is not very straightforward. Therefore, it is better to use the defined effect (or contrast) shown here to calculate the power, even though this calculation is much more complicated.

Conclusion

In this chapter, we discussed how to evaluate an experiment design. Although the evaluation can be conducted either before or after conducting the experiment, it is always recommended to evaluate an experiment before performing it. A bad design will waste time and money. Readers should check the alias structure, the orthogonality and the power for important effects for an experiment before the tests.

Optimal Custom Designs

IN THIS CHAPTER

Regression Model-Based Optimal Designs **307**
Select a D-Optimal Custom Design from a Standard Design 308
Algorithms for Selecting Model-Based D-Optimal Designs 314

Although two level fractional factorial designs, Plackett-Burman designs, Taguchi orthogonal array and other predefined designs are enough for most applications, sometimes these designs may not be sufficient due to constraints on available resources such as time, cost and factor values. Therefore, in this chapter, we will discuss how to create an optimal custom design. A DOE folio has two types of optimal custom designs: regression model-based and distance-based.

Regression Model-Based Optimal Designs

Regression model-based optimal designs are optimal for a selected regression model. Therefore, a regression model must first be specified. The regression model should include all the effects that the experimenters are interested in.

As discussed in the linear regression chapter, the following linear equation is used in DOE data analysis.

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon$$

where:

- y is the response
- x_1 , ..., x_p are the factors
- β_0 , β_1 , ...,
- β_p are model coefficients
- ε is the error term

For each run, the above equation becomes:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots \beta_p x_{ip} + \varepsilon_i$$

It can be written in matrix notation as:

$$Y = X\beta + \varepsilon$$

where:

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

n is the total number of samples or runs. As discussed in the design evaluation chapter, the information matrix for an experiment is:

$$I = X'X$$

The variance and covariance matrix for the regression coefficients is:

$$\Sigma = \sigma^2 (X'X)^{-1}$$

where σ^2 is the variance of the error ε. It can be either specified by experimenters from their engineering knowledge or estimated from the data analysis. If the number of available samples is given, we need to choose the value of x_{ij} in matrix X to minimize the determinant of Σ. A small determinate means less uncertainty of the estimated coefficients β. This is the same as maximizing the determinant of $X'X$. A design that uses the determinant as the objective is called D-optimal design. A D-optimal design can be either selected from a standard design or created based on the values of factors without creating a standard design first. In this chapter, we discuss how to select a D-optimal design from a standard factorial design.

Select a D-Optimal Custom Design from a Standard Design

Assume that we need to design an experiment to investigate five factors with two levels each. To run a full factorial design, a total of $2^5 = 32$ factor combinations are needed. However, only 11 samples are available. To obtain the most information, which of the 11 factor combinations from the total of 32 should be applied to these 11 samples? To answer this question, we need to first decide what information we want to obtain.

The 32 runs required for a full factorial design are given in the table below.

Order	A	B	C	D	E
1	-1	-1	-1	-1	-1
2	1	-1	-1	-1	-1
3	-1	1	-1	-1	-1
4	1	1	-1	-1	-1
5	-1	-1	1	-1	-1
6	1	-1	1	-1	-1
7	-1	1	1	-1	-1
8	1	1	1	-1	-1
9	-1	-1	-1	1	-1
10	1	-1	-1	1	-1
11	-1	1	-1	1	-1
12	1	1	-1	1	-1
13	-1	-1	1	1	-1
14	1	-1	1	1	-1
15	-1	1	1	1	-1
16	1	1	1	1	-1
17	-1	-1	-1	-1	1
18	1	-1	-1	-1	1
19	-1	1	-1	-1	1
20	1	1	-1	-1	1
21	-1	-1	1	-1	1
22	1	-1	1	-1	1
23	-1	1	1	-1	1
24	1	1	1	-1	1
25	-1	-1	-1	1	1
26	1	-1	-1	1	1
27	-1	1	-1	1	1
28	1	1	-1	1	1

29	-1	-1	1	1	1
30	1	-1	1	1	1
31	-1	1	1	1	1
32	1	1	1	1	1

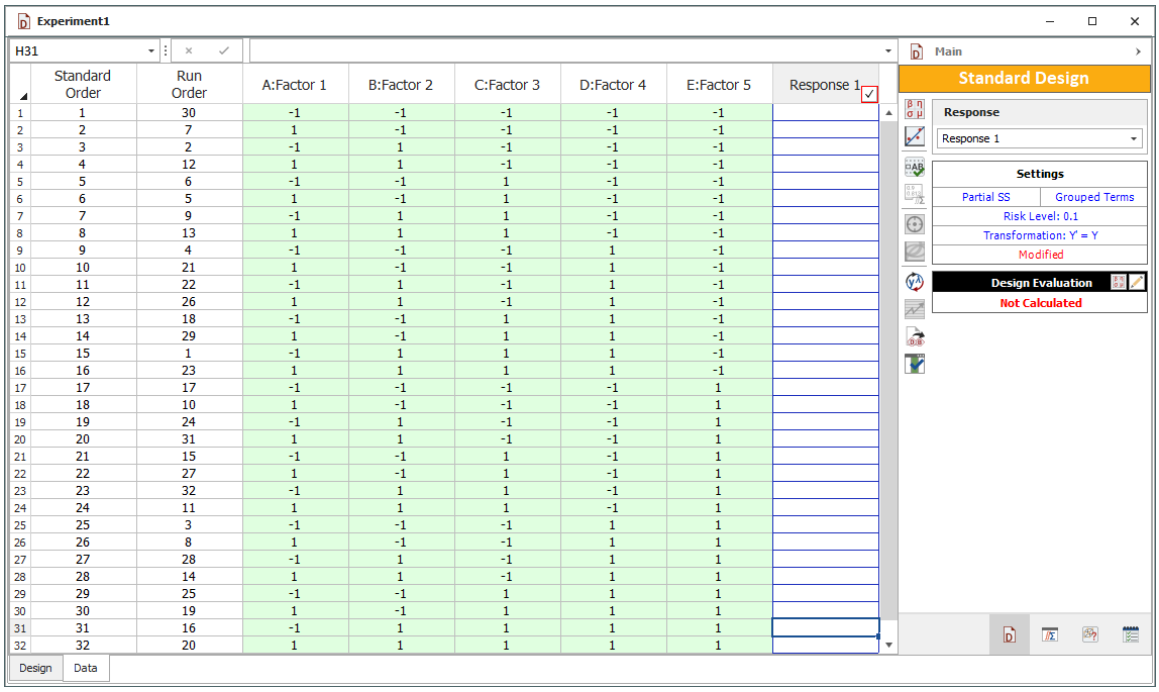
Since only 11 test samples are available, we must choose 11 factor value combinations from the above table. The experiment also has the following constraints:

- Due to safety concerns, it is not allowed to set all the factors at their high level at the same time. Therefore, we cannot use run number 32 in the experiment.
- The engineers are very interested in checking the response at A=D=1, and B=C=E=-1. Therefore, run number 10 must be included in the experiment.
- The engineers need to check the main effects and the interaction effect of AE. Therefore, the model for designing the optimal customer design is:

$$y = \beta_0 + \beta_1 A + \beta_2 B + \beta_3 C + \beta_4 D + \beta_5 E + \beta_{15} AE + \varepsilon$$

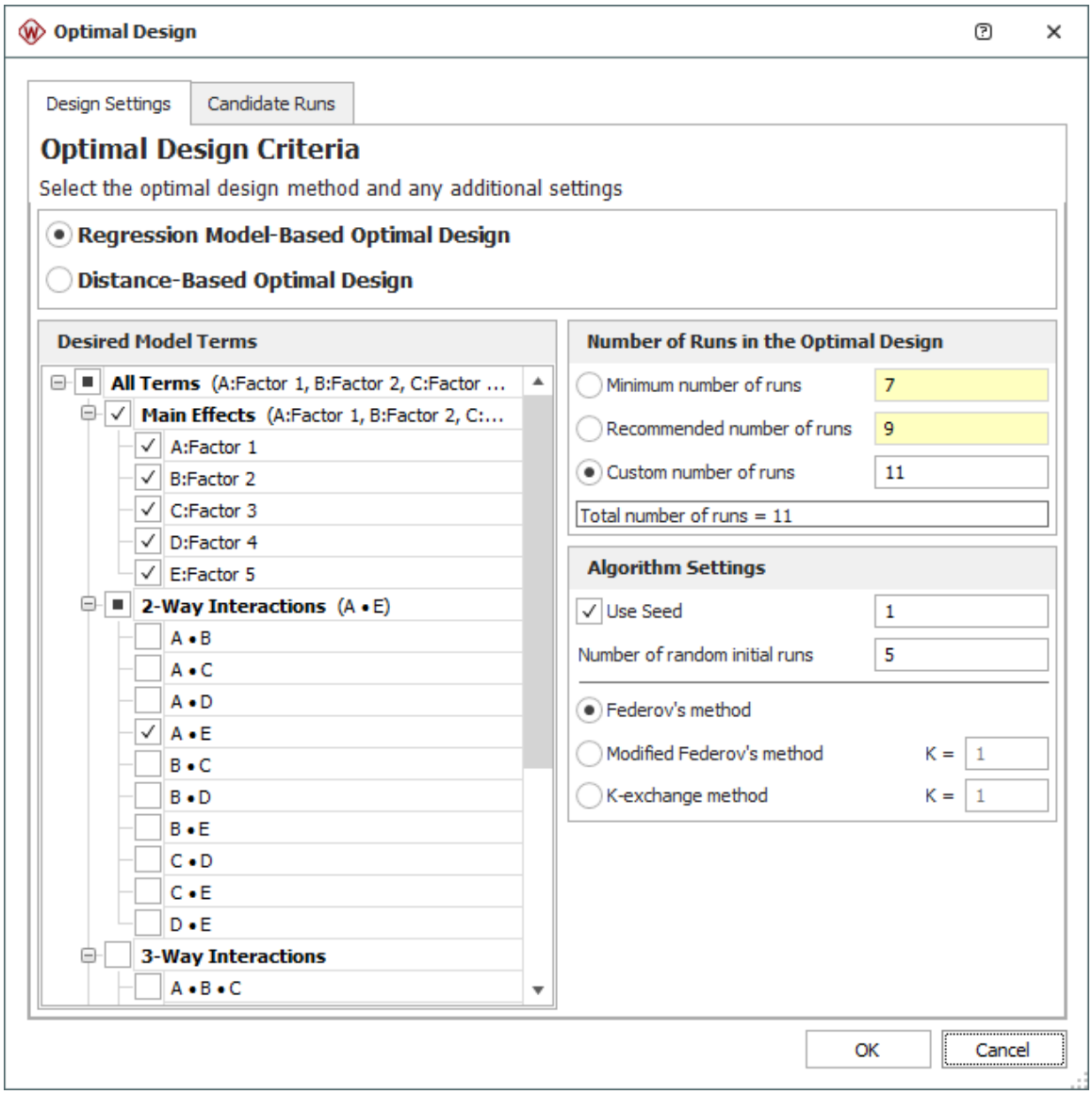
Since run number 10 must be included, there are only 10 runs left to choose from the above table. Many algorithms can be used to choose these 10 runs to maximize the determinant of the information matrix.

For this example, we first need to create a regular two level factorial design with five factors, as shown next.



	Standard Order	Run Order	A:Factor 1	B:Factor 2	C:Factor 3	D:Factor 4	E:Factor 5	Response 1
1	1	30	-1	-1	-1	-1	-1	
2	2	7	1	-1	-1	-1	-1	
3	3	2	-1	1	-1	-1	-1	
4	4	12	1	1	-1	-1	-1	
5	5	6	-1	-1	1	-1	-1	
6	6	5	1	-1	1	-1	-1	
7	7	9	-1	1	1	-1	-1	
8	8	13	1	1	1	-1	-1	
9	9	4	-1	-1	-1	1	-1	
10	10	21	1	-1	-1	1	-1	
11	11	22	-1	1	-1	1	-1	
12	12	26	1	1	-1	1	-1	
13	13	18	-1	-1	1	1	-1	
14	14	29	1	-1	1	1	-1	
15	15	1	-1	1	1	1	-1	
16	16	23	1	1	1	1	-1	
17	17	17	-1	-1	-1	-1	1	
18	18	10	1	-1	-1	-1	1	
19	19	24	-1	1	-1	-1	1	
20	20	31	1	1	-1	-1	1	
21	21	15	-1	-1	1	-1	1	
22	22	27	1	-1	1	-1	1	
23	23	32	-1	1	1	-1	1	
24	24	11	1	1	1	-1	1	
25	25	3	-1	-1	-1	1	1	
26	26	8	1	-1	-1	1	1	
27	27	28	-1	1	-1	1	1	
28	28	14	1	1	-1	1	1	
29	29	25	-1	-1	1	1	1	
30	30	19	1	-1	1	1	1	
31	31	16	-1	1	1	1	1	
32	32	20	1	1	1	1	1	

On the Design Settings tab of the Optimal Design window, you can select which terms to include in the model and specify the number of runs to be used in the experiment. Only main effects and AE are selected. The number of runs is set to 11. In the Candidate Runs tab of the window, run number 10 was set to be included and run number 32 was set to be excluded.


Optimal Design
Design Settings
Candidate Runs
Optimal Design Criteria
Select the optimal design method and any additional settings
Regression Model-Based Optimal Design
Distance-Based Optimal Design
Desired Model Terms
All Terms (A:Factor 1, B:Factor 2, C:Factor ...
Main Effects (A:Factor 1, B:Factor 2, C:...
A:Factor 1
B:Factor 2
C:Factor 3
D:Factor 4
E:Factor 5
2-Way Interactions (A • E)
A • B
A • C
A • D
A • E
B • C
B • D
B • E
C • D
C • E
D • E
3-Way Interactions
A • B • C
Number of Runs in the Optimal Design
Minimum number of runs 7
Recommended number of runs 9
Custom number of runs 11
Total number of runs = 11
Algorithm Settings
Use Seed 1
Number of random initial runs 5
Federov's method
Modified Federov's method K = 1
K-exchange method K = 1
OK
Cancel

Optimal Design

Design Settings | Candidate Runs

Optimal Design Runs

Select the runs to be considered in the optimal design

# of reps in sheet	Included in ptimal design	# of reps optimal desig	ID	Block	A:Factor 1	B:Factor 2	C:Factor 3	D:Factor 4	E:Factor 5
1	Consider	0	1	1	1	1	1	1	1
1	Consider	0	2	1	2	1	1	1	1
1	Consider	0	3	1	1	2	1	1	1
1	Consider	0	4	1	2	2	1	1	1
1	Consider	0	5	1	1	1	2	1	1
1	Consider	0	6	1	2	1	2	1	1
1	Consider	0	7	1	1	2	2	1	1
1	Consider	0	8	1	2	2	2	1	1
1	Consider	0	9	1	1	1	1	2	1
1	Include	1	10	1	2	1	1	2	1
1	Consider	0	11	1	1	2	1	2	1
1	Consider	0	12	1	2	2	1	2	1
1	Consider	0	13	1	1	1	2	2	1
1	Consider	0	14	1	2	1	2	2	1
1	Consider	0	15	1	1	2	2	2	1
1	Consider	0	16	1	2	2	2	2	1
1	Consider	0	17	1	1	1	1	1	2
1	Consider	0	18	1	2	1	1	1	2
1	Consider	0	19	1	1	2	1	1	2
1	Consider	0	20	1	2	2	1	1	2
1	Consider	0	21	1	1	1	2	1	2
1	Consider	0	22	1	2	1	2	1	2
1	Consider	0	23	1	1	2	2	1	2
1	Consider	0	24	1	2	2	2	1	2
1	Consider	0	25	1	1	1	1	2	2
1	Consider	0	26	1	2	1	1	2	2
1	Consider	0	27	1	1	2	1	2	2
1	Consider	0	28	1	2	2	1	2	2
1	Consider	0	29	1	1	1	2	2	2
1	Consider	0	30	1	2	1	2	2	2
1	Consider	0	31	1	1	2	2	2	2
1	Exclude	0	32	1	2	2	2	2	2

Delete | Add

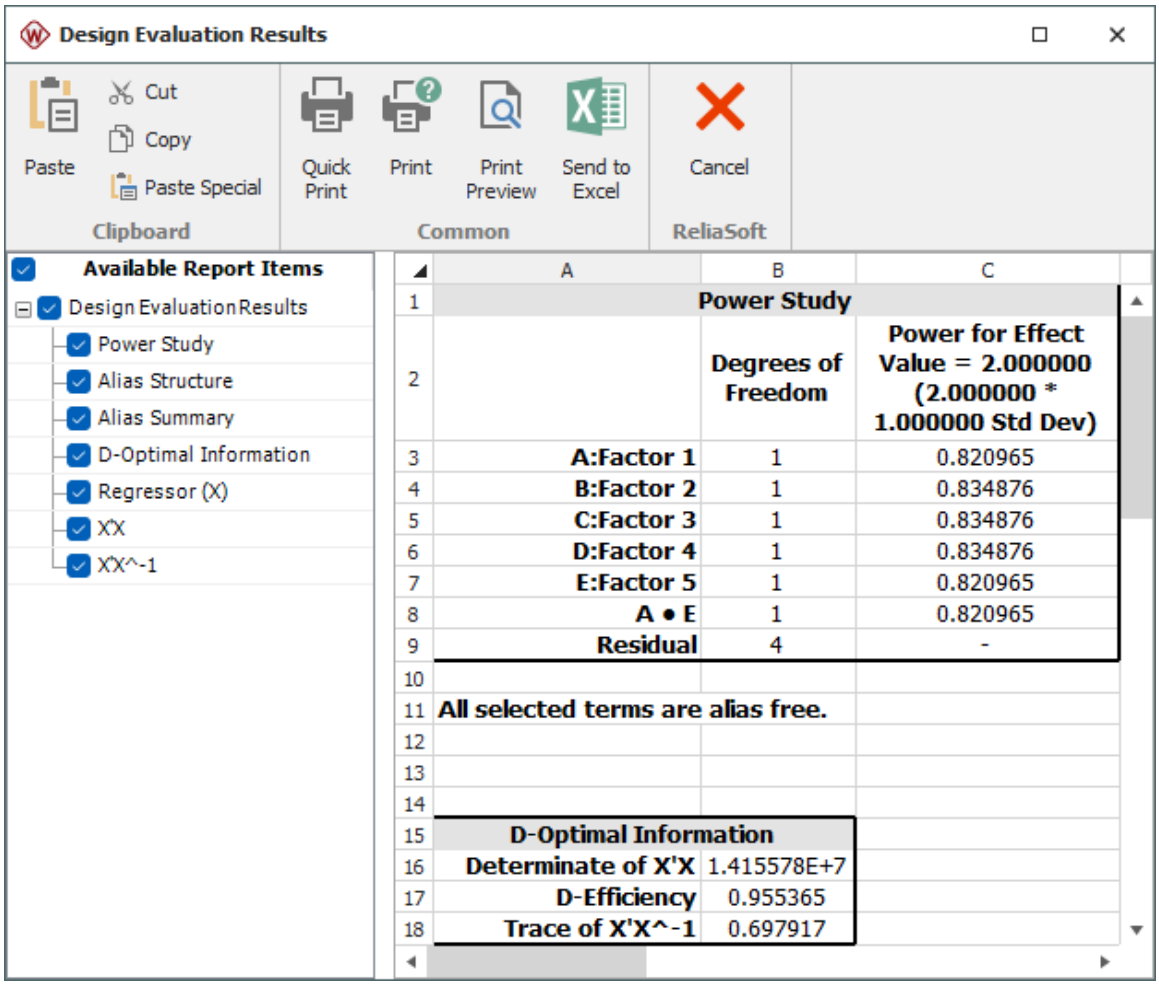
OK | Cancel

The resulting optimal custom design is shown next.

A7 | 10

	Standard Order	Run Order	A:Factor 1	B:Factor 2	C:Factor 3	D:Factor 4	E:Factor 5	Response 1
1	2	1	1	-1	-1	-1	-1	
2	13	2	-1	-1	1	1	-1	
3	27	3	-1	1	-1	1	1	
4	18	4	1	-1	-1	-1	1	
5	16	5	1	1	1	1	-1	
6	28	6	1	1	-1	1	1	
7	10	7	1	-1	-1	1	-1	
8	30	8	1	-1	1	1	1	
9	24	9	1	1	1	-1	1	
10	21	10	-1	-1	1	-1	1	
11	3	11	-1	1	-1	-1	-1	

The design evaluation results for this design are given below.



	A	B	C
1	Power Study		
2		Degrees of Freedom	Power for Effect Value = 2.000000 (2.000000 * 1.000000 Std Dev)
3	A:Factor 1	1	0.820965
4	B:Factor 2	1	0.834876
5	C:Factor 3	1	0.834876
6	D:Factor 4	1	0.834876
7	E:Factor 5	1	0.820965
8	A • E	1	0.820965
9	Residual	4	-
10			
11	All selected terms are alias free.		
12			
13			
14			
15	D-Optimal Information		
16	Determinate of X'X	1.415578E+7	
17	D-Efficiency	0.955365	
18	Trace of X'X^-1	0.697917	

From the design evaluation results, we can see the generated design can clearly estimate all the main effects and the AE interaction. The determinate of the information matrix X'X is 1.42E+7 and the D-efficiency is 0.9554. The power for an effect of 2-standard deviation for each of the terms of interest is also given in the design evaluation.

Algorithms for Selecting Model-Based D-Optimal Designs

In a DOE folio, the Federov's method, the modified Federov's method and the k-exchange method are used to select test runs from all the candidates. They are briefly explained below.

The Federov algorithm [Fedorov, 1972]

Assume there is an initial optimal design with number of runs of n. The initial optimal design can be obtained using the sequential optimization method given in [Dykstra 1971, Galil and Kiefer 1980]. We need to exchange one of the rows in the initial design with one of the rows from the candidate runs. Let's call the initial design X_{old} and call the design after row exchange X_{new}. The determinant of the new information matrix is:

$$\left|X'_{new}X_{new}\right| = \left|X'_{old}X_{old}\right| \times (1 + \Delta(x_i, x_j))$$

where x_i is the row from the current optimal design. It needs to be exchanged with x_j, a candidate run from the candidate set. $\Delta(x_j, x_j)$ is the amount of change in the determinant of the information matrix. It is calculated by:

$$\Delta(x_i, x_j) = d(x_j) - \left[d(x_i)\,d(x_j) - d(x_i, x_j)^2\right] - d(x_i)$$

where:

- $d(x_i, x_j) = x_i'\left(X'_{old}X_{old}\right)^{-1}x_j$ is the covariance for x_i and x_j
- $d(x_i)$ and $d(x_j)$ are the variance of x_i and x_j calculated using the current optimal design X_{old}

The basic idea behind the Fedorov algorithm is to calculate the delta-value for all the possible exchange pairs from the current design and the candidate runs, and then select the pair with the highest value. At each iteration, it calculates $n \times N$ deltas (where n is the number of runs in the current design matrix and N is the number of runs in the candidate run matrix) and chooses the best one for exchange. The algorithm stops when the change of the determinate is less than a pre-defined small value.

The Modified Federov Algorithm [Cook and Nachtsheim, 1980]

The above Federov algorithm is very slow since it only conducts one exchange after calculating $n \times N$ deltas. The modified Federov algorithm tries to improve the speed. It is a simplified version of the Fedorov method. Assume the current design matrix is X_{old}. The algorithm starts from the 1st row in the design matrix and uses it to calculate $1 \times N$ deltas (deltas of this design run with all the candidate runs). An exchange is conducted if the largest delta is a positive value. The above steps are repeated until the increase of the determinant is less than a pre-defined small value. Therefore, the modified Federov algorithm results in one exchange after calculating N deltas.

The K-Exchange Algorithm [Johnson and Nachtsheim, 1983]

This is a variation of the modified Fedorov algorithm. Instead of calculating the deltas for all the design runs in one iteration, it calculates only the deltas for the k-worst runs. First, the algorithm uses the current design matrix to calculate the variance of each run in the design matrix. The k runs with the lowest variances are the runs that need to exchange. Then for each of the k worst runs, it calculates N deltas with all the N candidate runs. If the largest delta is greater than a pre-defined small value, then this row is exchanged with the candidate row, resulting in the largest positive delta. Once all the k points are exchanged, a new design matrix is

obtained and the above steps are repeated until no exchange can be conducted. Usually k is set to be:$k \leq n/4$ where n is the number of runs in the optimal design. 2 Distance-Based Optimal Designs

Sometimes, experimenters want the design points (runs) in an experiment to cover as large of a design space as possible. In other words, the distance between design points is intended to be as far as possible. Distance-based optimal designs are used for this purpose.

To create a distance-based optimal design, the candidate runs should be available. First, the average value of each factor is calculated. This average is called the "center" of the design space. For qualitative factors, the average is calculated for the indicator variables in the regression model. For quantitative factors, the average is calculated based on the coded values. The distance of each candidate run to the center is calculated and sorted. The run with the largest distance is selected to be in the optimal design. If there are multiple runs with the same largest distance, a run is randomly selected from them. The "center" of the runs in the current optimal design is then calculated. The distances of all the available runs to this "center" are also calculated. Based on these distances, the next run is selected and put in the optimal design. Repeat this process until the required number of runs in the optimal design is reached.

Reference for the Algorithms:

- Fedorov, V. V. (1972), "Theory of Optimal Experiments (Review)", *Biometrika*, cvol. 59, no. 3, 697-698. Translated and edited by W. J. Studden and E. M. Klimko.
- Dykstra, O. (1971), "The augmentation of experimental data to maximize |X'X|", *Technometrics*, vol. 13, no. 3, 682-688.
- Galil, Z. and Kiefer, J. (1980), "Time and Space Saving Computer Methods, Related to Mitchell's DETMAX, for Finding D-Optimal Designs", *Technometrics*, vol. 22, no. 3, 301-313.
- Cook, R. D. and Nachtsheim, C. J. (1980), "A Comparison of Algorithms for Constructing Exact D-Optimal Designs," *Technometrics*, vol. 22, no. 3, 315-324.
- Johnson, M. E. and Nachtsheim, C. J. (1983), "Some Guidelines for Constructing Exact D-Optimal Designs on Convex Design Spaces," *Technometrics* , vol. 25, no. 3, 271-277.

Robust Parameter Design

IN THIS CHAPTER

Taguchi's Philosophy **317**

Robust Parameter Design **319**
Control and Noise Factor Interaction 319
Inner and Outer Arrays 321

Signal to Noise Ratios **322**
Location Model 323
Dispersion Model 323
Analysis Strategy 325

Limitations of Taguchi's Approach **331**

In Response Surface Methods for Optimization, techniques used to optimize the response were discussed. Once an optimum value of the response has been achieved, the next goal of the experimenter should be to make the optimum response value insensitive to the noise factors so that a consistent performance is obtained at all times. For example, if the yield from a chemical process has been optimized at 95%, then this value of yield should be obtained regardless of the variation in factors such as the quality of reactants or fluctuations in humidity or other weather conditions. These factors are beyond the control of the operator. Therefore, the product or process should be such that it is not affected by minor fluctuations in these factors. The process of making a system insensitive to such factors is referred to as *robust design*. Robust design was pioneered by the Japanese industrialist Dr. Genichi Taguchi in the early 1980s. This chapter briefly discusses his approach.

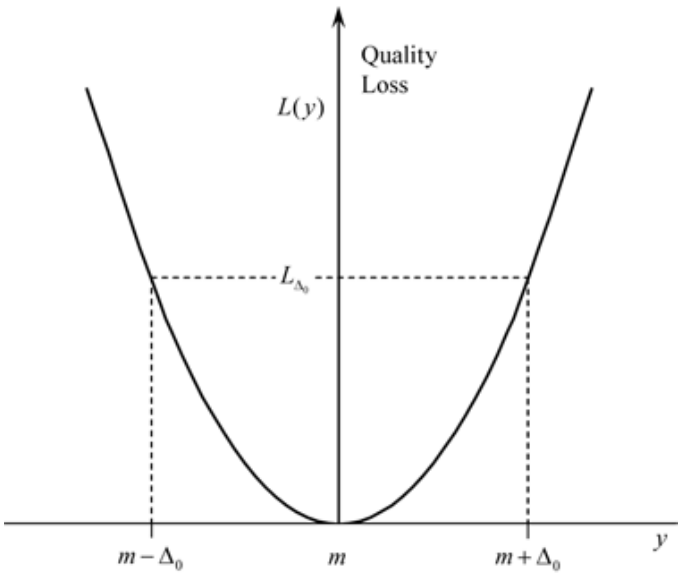
Taguchi's Philosophy

Taguchi's philosophy is based on the fact that any decrease in the quality of a system leads to customer dissatisfaction. This occurs even if the departure in quality lies within the specified limits of the system and is considered acceptable to the customer. For example, consider the case of a laptop that develops a defect on its screen within the warranty period. Although the

customer is able to get a warranty-replacement for the screen this might lead to a little dissatisfaction on the part of the customer. If the same laptop then develops a problem with its DVD drive, the customer might declare the laptop "useless," even if the problem occurs during the warranty period and the customer is able to get a free replacement. Therefore, to maintain a good reputation, the laptop manufacturer needs to produce laptops that offer the same quality to all customers consistently. This can only be done when the required quality is built into the laptops. Note how this approach differs from traditional quality control where it is considered sufficient to manufacture products within certain specifications and carry out pre-shipment quality control inspections (i.e., sampling inspections) to filter out products that fall out of specification.

Taguchi's philosophy requires that systems be designed in such a way that they are produced, not just within the specified limits, but right on target specifications or best values. Such a proactive approach is much more fruitful and efficient than the reactive approach of sampling inspections. The philosophy of Taguchi is summarized by his quality loss function (see the figure below). The function states that any deviation from the target value leads to a quadratic loss in quality or customer satisfaction. Mathematically, the function may be expressed as:

$$L = k(y - m)^2$$



where y represents the performance parameter of the system, m represents the target or the nominal value of y, L represents the quality loss and k is a constant.

Taguchi's approach to achieve a high quality system consists of three stages: *system design*, *parameter design* and *tolerance design*. System design refers to the stage when ideas for a new system are used to decide upon the combinations of factors to obtain a functional and

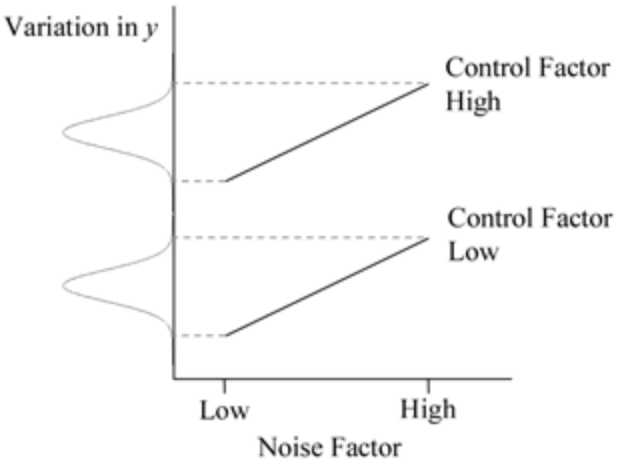
economical design. Parameter design refers to the stage when factor settings are selected that make the system less sensitive to variations in the uncontrollable factors affecting the system. Therefore, if this stage is carried out successfully, the resulting system will have little variation and the resulting tolerances will be tight. Tolerance design refers to the final stage when tolerances are tightened around the best value. This stage increases cost and is only needed if the required quality is not achieved during parameter design. Thus, using parameter design, it is possible to achieve the desired quality without much increase in the cost. The tolerance design stage is discussed in detail next.

Robust Parameter Design

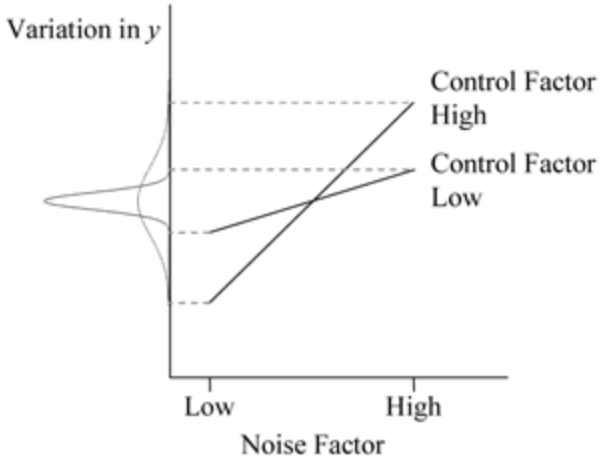
Taguchi divided the factors affecting any system into two categories: *control factors* and *noise factors*. Control factors are factors affecting a system that are easily set by the experimenter. For example, if in a chemical process the reaction time is found to be a factor affecting the yield, then this factor is a control factor since it can be easily manipulated and set by the experimenter. The experimenter will chose the setting of the reaction time that maximizes the yield. Noise factors are factors affecting a system that are difficult or impossible to control. For example, ambient temperature may also have an effect on the yield of a chemical process, but ambient temperature could be a noise factor if it is beyond the control of the experimenter. Thus, change in ambient temperature will lead to variations in the yield but such variations are undesirable.

Control and Noise Factor Interaction

In our example, since the experimenter does not have any control on the change in ambient temperature, he/she needs to find the setting of the reaction time at which there is minimal variation of yield due to change in ambient temperature. Note that this can only occur if there is an interaction between the reaction time (control factor) and ambient temperature (noise factor). If there is no such interaction, variation in yield due to changes in ambient temperature will always occur regardless of the setting of the reaction time. Therefore, to solve a robust parameter design problem, interaction between control and noise factors must exist. This fact is further explained by the figure shown next.



(a)



(b)

The figure shows the variation of the response (yield) for two levels of the noise factor (ambient temperature). The response values are plotted at two levels of the control factor (reaction time). Figure (a) shows the case where there is no interaction between the control and noise factors. It can be seen that, regardless of the settings of the control factor (low or high), the variation in the response remains the same (as is evident from the constant spread of the probability distribution of the response at the two levels of the control factor). Figure (b) shows the case where an interaction exists between the control and noise factors. The figure indicates that in the present case it is advantageous to have the control factor at the low setting, since at this setting there is not much variation in the response due to change in the noise factor (as is evident from the smaller spread of the probability distribution of the response at the low level of the control factor).

Inner and Outer Arrays

Taguchi studied the interaction between the control and noise factors using two experiment designs: the *inner array* and the *outer array*. The inner array is essentially any experiment design that is used to study the effect of the control factors on the response. Taguchi then used an outer array for the noise factors so that each run of the inner array was repeated for every treatment of the outer array. The resulting experiment design, that uses both inner and outer arrays, is referred to as a *cross array*.

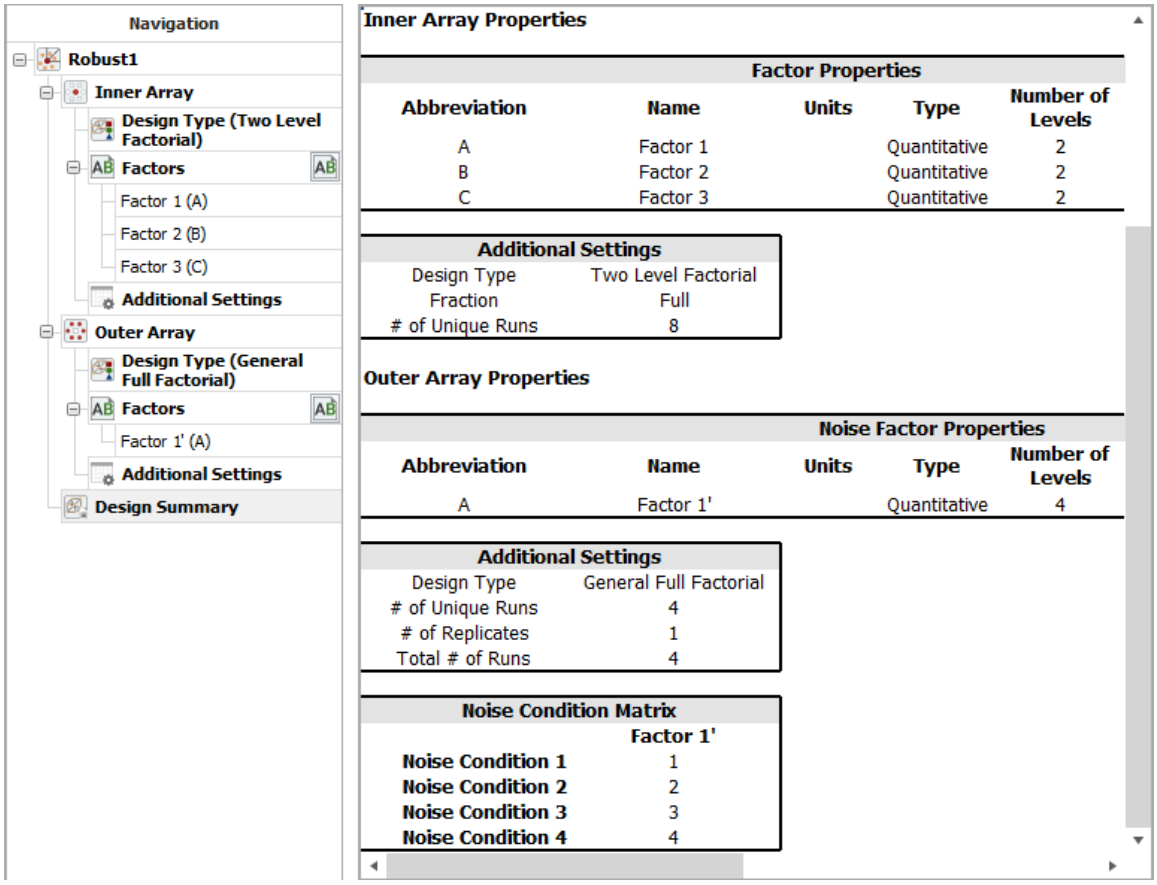
Example

To illustrate Taguchi's use of the inner and outer arrays consider the case of a chemical process where the experimenter wants the product to be neither acidic nor basic (i.e., the pH of the product needs to be as close to 7 as possible). It is thought that the pH of the product depends on the concentration of the three reactants, A, B and C, used to obtain the product. There are three control factors here, namely the concentration of each of the three reactants. It has also been found that the pH of the product depends on the ambient temperature which varies naturally and cannot be controlled. Thus, there is one noise factor in this case - the ambient temperature. The experimenter chooses Taguchi's robust parameter design approach to investigate the settings of the control factors to make the product insensitive to changes in ambient temperature. It is decided to carry out a 2^3 experiment to study the effect of the three control factors on the pH of the product. Therefore, the 2^3 design is the inner array here. It is also decided to carry out the experiment at four levels of the ambient temperature by using a special enclosure where the surrounding temperature of the chemical process can be controlled. Therefore, the outer array consists of a single factor experiment with the factor at four levels. Note that, in order to carry out the robust parameter design approach, the noise factor should be such that it can be controlled in an experimental setting. The resulting setup of the robust parameter design experiment is shown in the following table.

				Outer Array			
				Noise Factor 1:	Level 1	Level 2	Level 3	Level 4
	Inner Array							
Run	Factors							
	A	*B*	*C*					
(1)	-1	-1	-1		5.05	6.50	7.00	7.15
a	1	-1	-1		5.05	5.05	5.35	6.15
b	-1	1	-1		5.95	6.85	7.65	7.95
ab	1	1	-1		4.65	4.95	5.35	5.00
c	-1	-1	1		6.35	7.00	7.25	7.00
ac	1	-1	1		4.05	6.55	6.15	6.45
bc	-1	1	1		6.85	7.05	7.00	8.00
abc	1	1	1		4.65	5.55	6.00	6.15

The experiment requires $2^3 \times 4 = 32$ runs in all as each run of the inner array is repeated for every treatment of the outer array. The above table also shows the pH values obtained for the

experiment. In a Weibull++ DOE folio, this design is set up by specifying the properties for the inner and outer arrays as shown in the following figure.



The resulting design is shown in the next figure.

	Standard Order	A:Factor 1	B:Factor 2	C:Factor 3	Noise Condition 1	Noise Condition 2	Noise Condition 3	Noise Condition 4
1	1	-1	-1	-1	5.05	6.5	7	7.15
2	2	1	-1	-1	5.05	5.05	5.35	6.15
3	3	-1	1	-1	5.95	6.85	7.65	7.95
4	4	1	1	-1	4.65	4.95	5.35	5
5	5	-1	-1	1	6.32	7	7.25	7
6	6	1	-1	1	4.05	6.55	6.15	6.48
7	7	-1	1	1	6.85	7.05	7	8
8	8	1	1	1	4.65	5.55	6	6.15

Signal to Noise Ratios

Depending upon the objective of the robust parameter design, Taguchi defined three different statistics called *signal to noise ratios*. These ratios were defined as the means to measure the

variation of the response with respect to the noise factors. Taguchi's approach essentially consists of two models: a *location model* and a *dispersion model*.

Location Model

The location model is the regression model for the mean value of the response at each treatment combination of the inner array. If y_{ij} $(j = 1 \text{ to } m)$ represents the m response values obtained at the i th treatment combination of the inner array (corresponding to the m levels of the noise factors), then the mean response at the i th level is:

$$\bar{y}_i = \frac{1}{m}\sum_{j=1}^{m} y_{ij}$$

The location model is obtained by fitting a regression model to all $\bar{y}_i$ values, by treating these values as the response at each of the i th treatments of the inner array. As an example, the location model for an inner array with two factors can be written as:

$$\bar{y}_i = \beta_0^L + \beta_1^L x_{i1} + \beta_2^L x_{i2} + \beta_{12}^L x_{i1} x_{i2}$$

where:

- β_0^L is the intercept
- β_1^L is the coefficient for the first factor
- β_2^L is the coefficient for the second factor
- β_{12}^L is the coefficient for the interaction
- x_1 and x_2 are respectively the variables for the two factors

The objective of using the location model is to bring the response to its goal regardless of whether this is a target value, maximum value or minimum value. This is done by identifying significant effects and then using the least square estimates of the corresponding coefficients, β^L s, to fit the location model. The fitted model is used to decide the settings of the variables that bring the response to the goal.

Dispersion Model

The dispersion model measures the variation of the response due to the noise factors. The standard deviation of the response values at each treatment combination, s_i , is used. Usually, the standard deviation is used as a log function of s_i^2 because $\ln s_i^2$ are approximately normally distributed. These values can be calculated as follows:

$$\ln s_i^2 = \ln(\frac{\sum_{j=1}^{m}(y_{ij} - \bar{y}_i)^2}{m-1})$$

Thus, the dispersion model consists of using $\ln s_i^2$ as the response and investigating what treatment of the control factors results in the minimum variation of the response. Clearly, the objective of using the dispersion model is to minimize variation in the response. Instead of using standard deviations directly, Taguchi defined three signal to noise ratios (abbreviated SN) based on the objective function for the response. If the response is to be maximized, the SN ratio is defined as follows:

$$SN_L = -10\log_{10}(\frac{1}{m}\sum_{j=1}^{m}\frac{1}{y_{ij}^2})$$

The previous ratio is referred to as the larger-the-better ratio and is defined to decrease variability when maximizing the response. When the response is to be minimized, the SN ratio is defined as:

$$SN_S = -10\log_{10}(\frac{1}{m}\sum_{j=1}^{m}y_{ij}^2)$$

This ratio is referred to as the smaller-the-better ratio and is defined to decrease variability when minimizing the response. If the objective for the response is to achieve a target or nominal value, then the SN ratio is defined as follows:

$$SN_T = 10\log_{10}(\frac{\bar{y}_i^2}{s_i^2})$$

This ratio is referred to as the nominal-the-best ratio and is defined to decrease variability around a target response value. The dispersion model for any of the three signal to noise ratios can be written as follows for an inner array with two factors:

$$SN = \beta_0^D + \beta_1^D x_1 + \beta_2^D x_2 + \beta_{12}^D x_1 x_2$$

where:

- β_0^D is the intercept
- β_1^D is the coefficient for the first factor
- β_2^D is the coefficient for the second factor

- β_{12}^{D} is the coefficient for the interaction
- x_1 and x_2 are respectively the variables for the two factors

The dispersion model is fit by identifying significant effects and then using the least square estimates of the coefficients β^{D} s. Once the fitted dispersion model is known, settings for the control factors are found that result in the maximum value of SN, thereby minimizing the response variation.

Analysis Strategy

The location and dispersion regression models are usually obtained by using graphical techniques to identify significant effects. This is because the responses used in the two models are such that only one response value is obtained for each treatment of the inner array. Therefore, the experiment design in the case of the two models is an unreplicated design.

Once the location and dispersion models have been obtained by identification of the significant effects, the following analysis strategy [Wu, 2000] may be used:

- To obtain the best settings of the factors for larger-the-better and smaller-the-better cases:
  - The experimenter must first select the settings of the significant control factors in the location model to either maximize or minimize the response.
  - Then the experimenter must choose the settings of those significant control factors in the dispersion model, that are not significant in the location model, to maximize the SN ratio.
- For nominal-the-best cases:
  - The experimenter must first select the settings of the significant control factors in the dispersion model to maximize the SN ratio.
  - Then the experimenter must choose the levels of the significant control factors in the location model to bring the response on target. At times, the same control factor(s) may show up as significant in both the location and dispersion models. In these cases, the experimenter must use his judgement to obtain the best settings of the control factors based upon the two models.

Factors that do not show up as significant in both the models should be set at levels that result in the greatest economy. Generally, a follow-up experiment is usually carried out with the best settings to verify that the system functions as desired.

Example

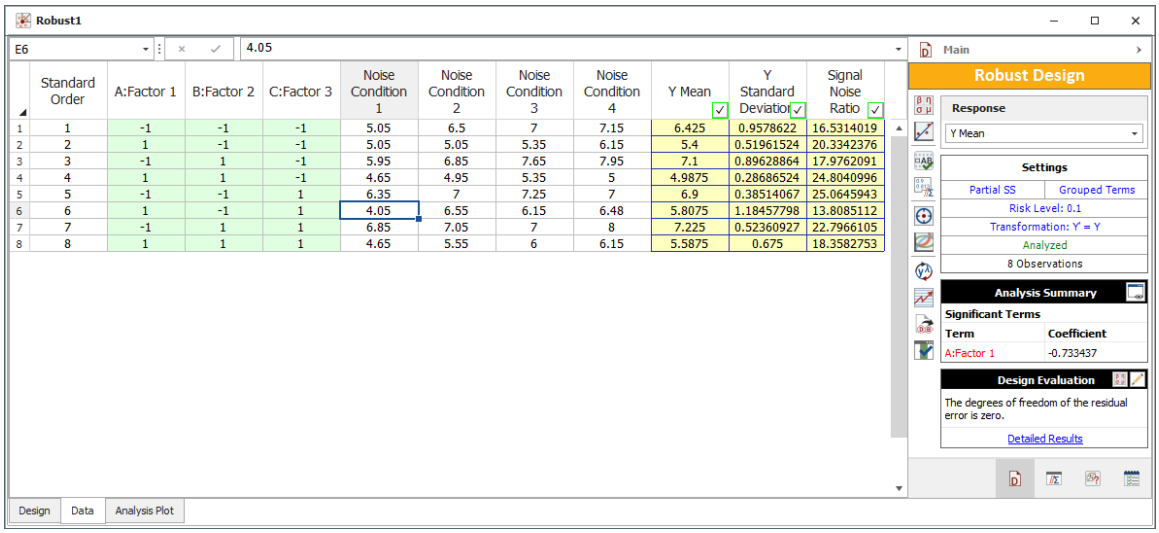
This example illustrates the procedure to obtain the location and dispersion models for the experiment in the previous example.

Location Model

The response values used in the location model can be calculated using the first equation given in Location Model. As an example, the response value for the third treatment is:

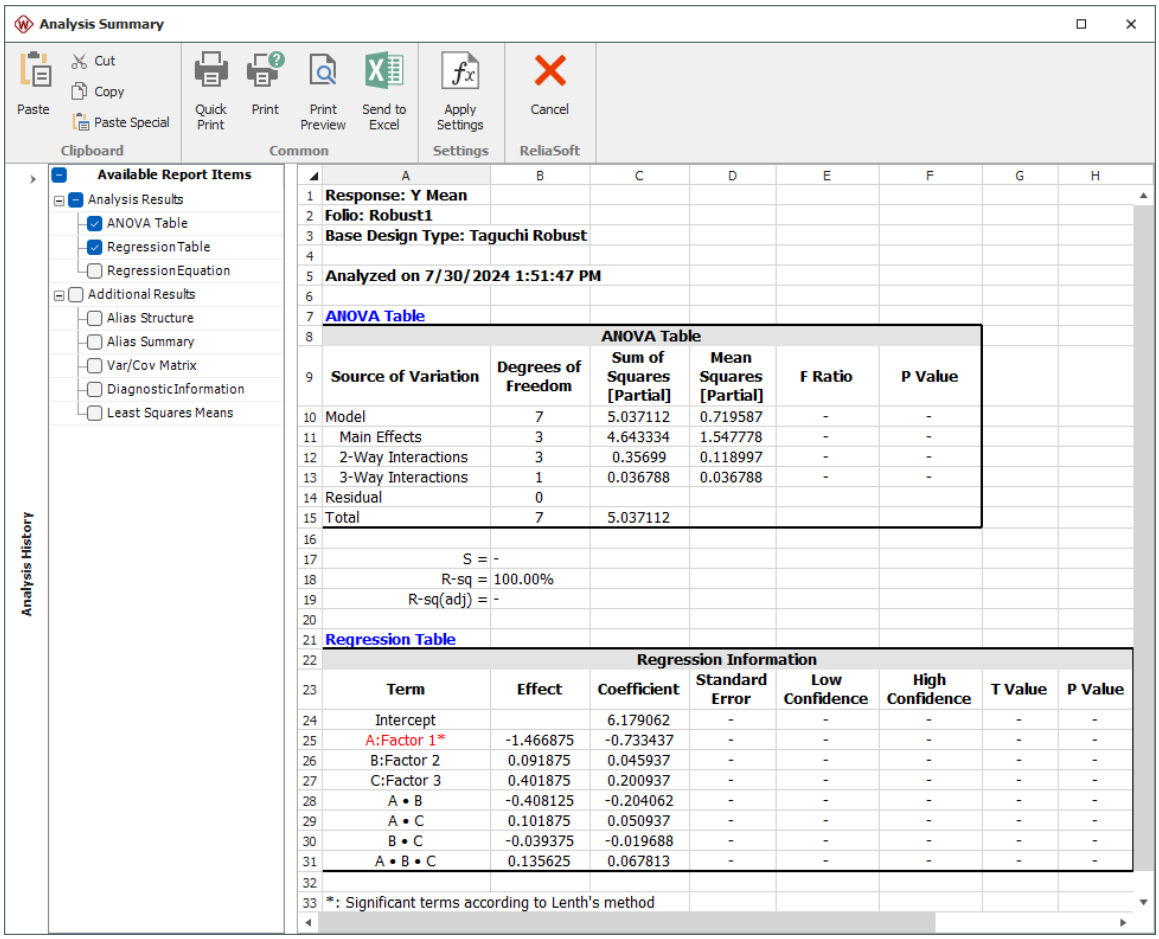
$$\begin{aligned}\bar{y}_3 &= \frac{1}{m}\sum_{j=1}^{m} y_{3j} \\ &= \frac{1}{4}(y_{31} + y_{32} + y_{33} + y_{34}) \\ &= \frac{1}{4}(5.95 + 6.85 + 7.65 + 7.95) \\ &= 7.1\end{aligned}$$

Response values for the remaining seven treatments can be calculated in a similar manner. These values are shown next under the Y Mean column.



	Standard Order	A:Factor 1	B:Factor 2	C:Factor 3	Noise Condition 1	Noise Condition 2	Noise Condition 3	Noise Condition 4	Y Mean	Y Standard Deviation	Signal Noise Ratio
1	1	-1	-1	-1	5.05	6.5	7	7.15	6.425	0.9578622	16.5314019
2	2	1	-1	-1	5.05	5.05	5.35	6.15	5.4	0.51961524	20.3342376
3	3	-1	1	-1	5.95	6.85	7.65	7.95	7.1	0.89628864	17.9762091
4	4	1	1	-1	4.65	4.95	5.35	5	4.9875	0.28686524	24.8040996
5	5	-1	-1	1	6.35	7	7.25	7	6.9	0.38514067	25.0645943
6	6	1	-1	1	4.05	6.55	6.15	6.48	5.8075	1.18457798	13.8085112
7	7	-1	1	1	6.85	7.05	7	8	7.225	0.52360927	22.7966105
8	8	1	1	1	4.65	5.55	6	6.15	5.5875	0.675	18.3582753

Once the response values for all the treatments are known, the analysis to fit the location model can be carried out by treating the experiment as a single replicate of the 2^3 design. The results obtained from the DOE folio are shown in the figure below.



Response: Y Mean
Folio: Robust1
Base Design Type: Taguchi Robust

Analyzed on 7/30/2024 1:51:47 PM

ANOVA Table

ANOVA Table

Source of Variation	Degrees of Freedom	Sum of Squares [Partial]	Mean Squares [Partial]	F Ratio	P Value
Model	7	5.037112	0.719587	-	-
Main Effects	3	4.643334	1.547778	-	-
2-Way Interactions	3	0.35699	0.118997	-	-
3-Way Interactions	1	0.036788	0.036788	-	-
Residual	0				
Total	7	5.037112			

S = -
R-sq = 100.00%
R-sq(adj) = -

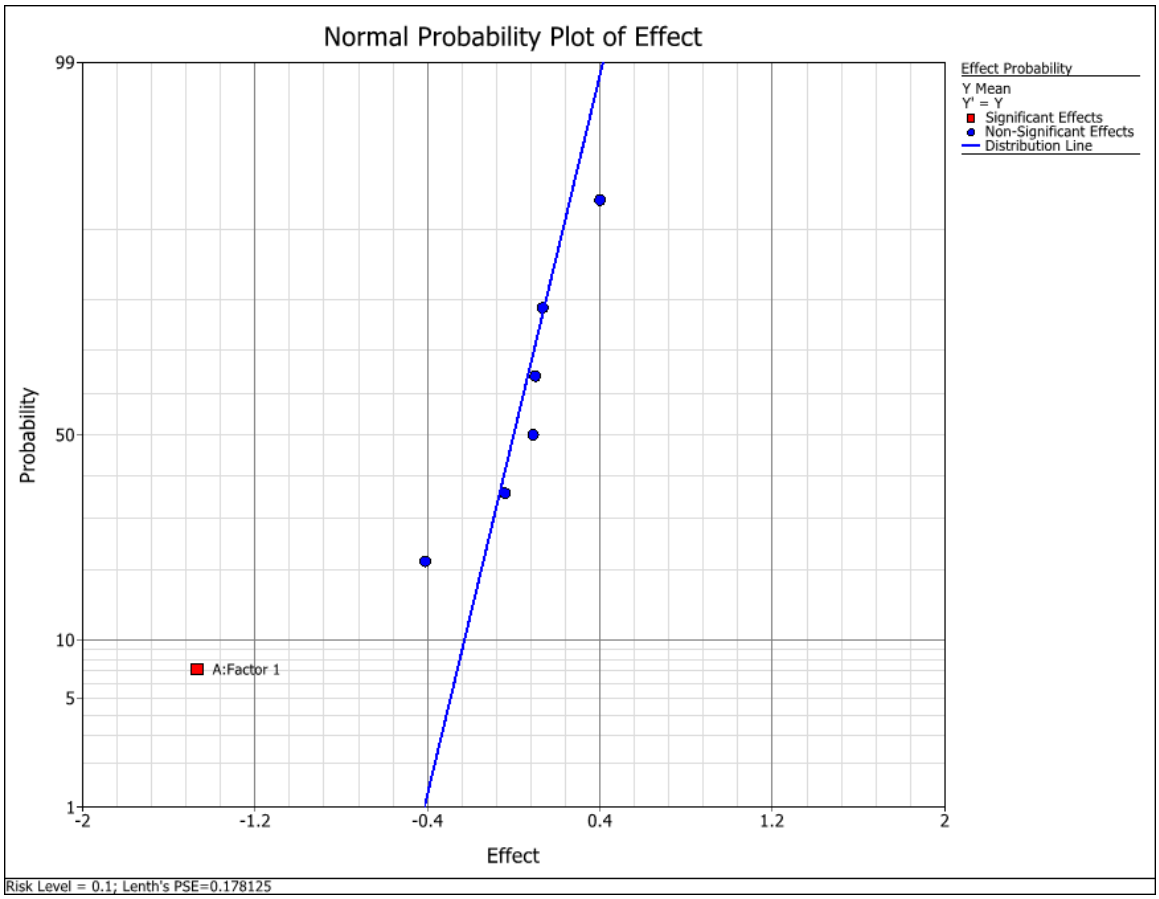
Regression Table

Regression Information

Term	Effect	Coefficient	Standard Error	Low Confidence	High Confidence	T Value	P Value
Intercept		6.179062	-	-	-	-	-
A:Factor 1*	-1.466875	-0.733437	-	-	-	-	-
B:Factor 2	0.091875	0.045937	-	-	-	-	-
C:Factor 3	0.401875	0.200937	-	-	-	-	-
A • B	-0.408125	-0.204062	-	-	-	-	-
A • C	0.101875	0.050937	-	-	-	-	-
B • C	-0.039375	-0.019688	-	-	-	-	-
A • B • C	0.135625	0.067813	-	-	-	-	-

*: Significant terms according to Lenth's method

The normal probability plot of effects for this model shows that only the main effect of factor ***A*** is significant for the location model (see the figure below).



Using the corresponding coefficient from the figure below, the location model can be written as:

$$\bar{y} = 6.1781 - 0.7344x_1$$

where x_1 is the variable representing factor A.

Dispersion Model

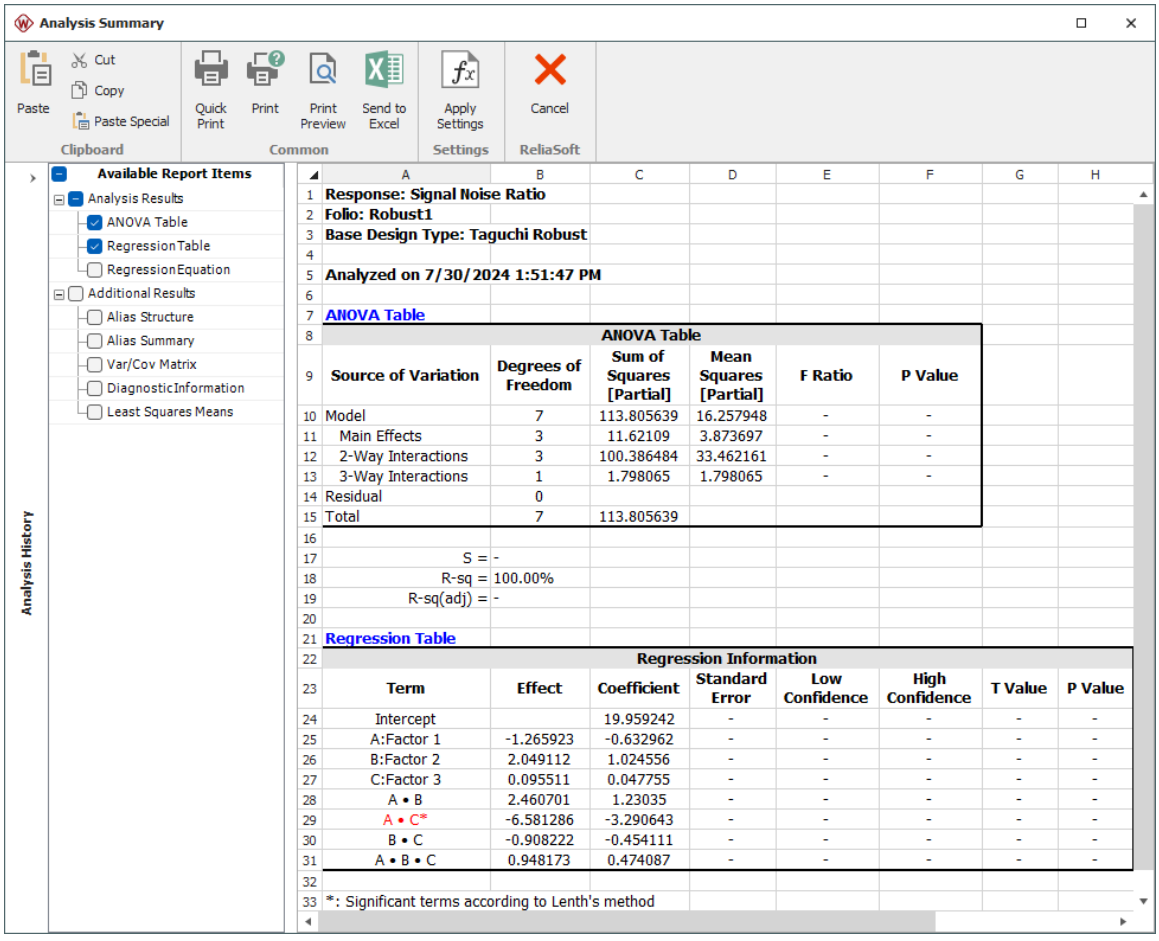
For the dispersion model, the applicable signal to noise ratio is given by the equation for the nominal-the-best ratio:

$$SN_{Ti} = -10\log_{10}(\frac{\bar{y}_i^2}{s_i^2})$$

The response values for the dispersion model can now be calculated. As an example, the response value for the third treatment is:

$$SN_{T3} = -10\log_{10}\left(\frac{\bar{y}_3^2}{s_1^2}\right)$$
$$= -10\log_{10}\left(\frac{\bar{y}_3^2}{\frac{1}{(4-1)}\sum_{j=1}^{4}(y_{3j}-\bar{y}_3)^2}\right)$$
$$= -10\log_{10}\left(\frac{7.1^2}{0.8963^2}\right)$$
$$= 17.9762$$

Other SN_{Ti} values can be obtained in a similar manner. The values are under the column Signal Noise Ratio in the data sheet shown above. As in the case of the location model, the analysis to fit the dispersion model can be carried out by treating the experiment as a single replicate of the 2^3 design. The results obtained from the DOE folio are shown in the figure below.



	A	B	C	D	E	F	G	H
1	**Response: Signal Noise Ratio**							
2	**Folio: Robust1**							
3	**Base Design Type: Taguchi Robust**							
4								
5	**Analyzed on 7/30/2024 1:51:47 PM**							
6								
7	**ANOVA Table**							

ANOVA Table

Source of Variation	Degrees of Freedom	Sum of Squares [Partial]	Mean Squares [Partial]	F Ratio	P Value
Model	7	113.805639	16.257948	-	-
Main Effects	3	11.62109	3.873697	-	-
2-Way Interactions	3	100.386484	33.462161	-	-
3-Way Interactions	1	1.798065	1.798065	-	-
Residual	0				
Total	7	113.805639			

S =	-
R-sq =	100.00%
R-sq(adj) =	-

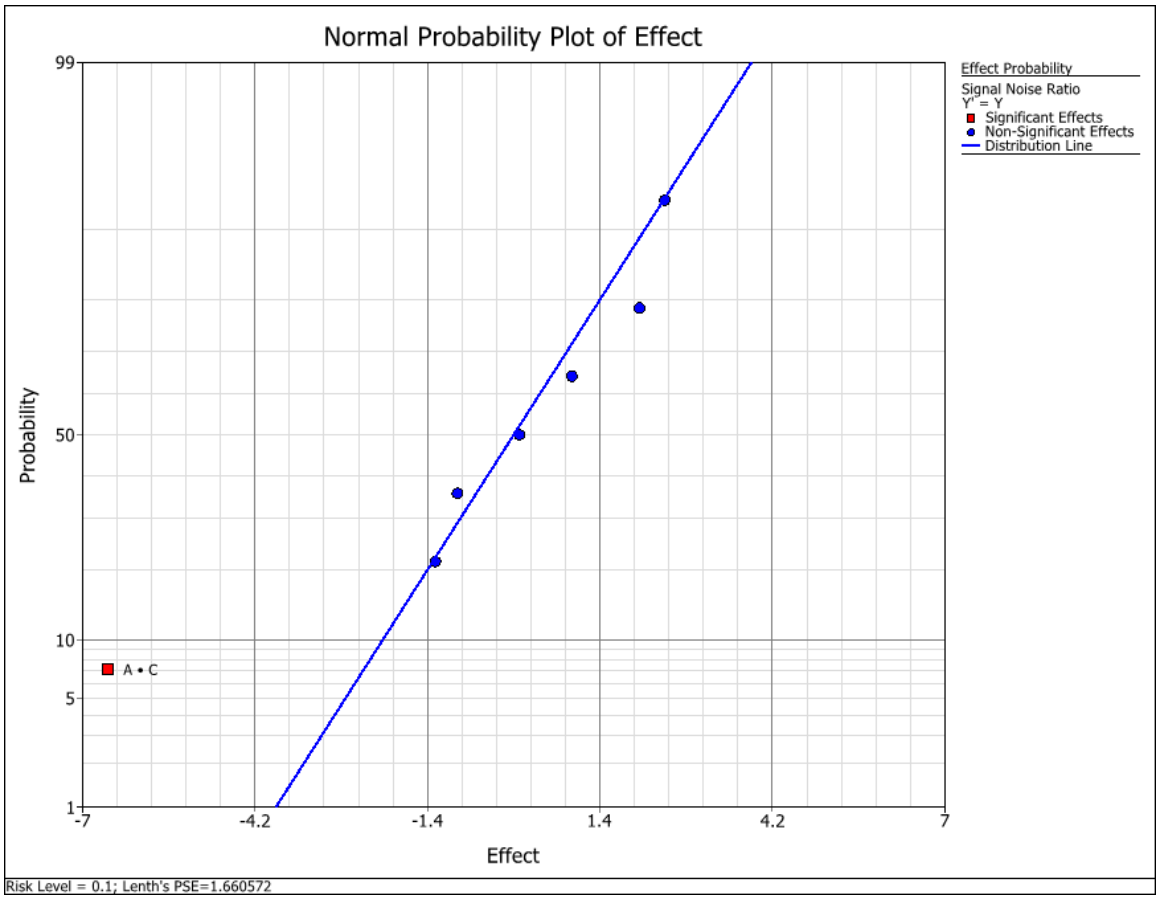
Regression Table

Regression Information

Term	Effect	Coefficient	Standard Error	Low Confidence	High Confidence	T Value	P Value
Intercept		19.959242	-	-	-	-	-
A:Factor 1	-1.265923	-0.632962	-	-	-	-	-
B:Factor 2	2.049112	1.024556	-	-	-	-	-
C:Factor 3	0.095511	0.047755	-	-	-	-	-
A • B	2.460701	1.23035	-	-	-	-	-
A • C*	-6.581286	-3.290643	-	-	-	-	-
B • C	-0.908222	-0.454111	-	-	-	-	-
A • B • C	0.948173	0.474087	-	-	-	-	-

*: Significant terms according to Lenth's method

The normal probability plot of effects for this model (displayed next) shows that the interaction AC is the only significant effect for this model.



Using the corresponding coefficient from the figure below, the dispersion model can be written as:

$$SN_T = 19.963 - 3.2869 x_1 x_3$$

where x_3 is the variable representing factor C.

Robust Parameter Design

Following the analysis strategy mentioned in Analysis Strategy, for the nominal-the-best case, the dispersion model should be considered first. The equation for the model shows that to maximize SN_T, either one of the following options can be used:

- $x_1 = -1$ and $x_3 = +1$ or
- $x_1 = +1$ and $x_3 = -1$

Then, considering the location model of this example, to achieve a target response value as close to 7 as possible, the only significant effect for this model, x_1, should be set at the level of -1. Therefore, the first should be used for the dispersion model's settings. The final settings for the three factors, as a result of the robust parameter design, are:

- Factor A is set at the low level
- Factor C is set at the high level and
- Factor B, which is not significant in any of the two models, can be set at the level that is most economical.

With these settings the predicted pH value for the product is:

$$\begin{aligned}\bar{y} &= 6.1781 - 0.7344x_1 \\ &= 6.1781 - 0.7344(-1) \\ &= 6.9125\end{aligned}$$

The predicted signal to noise ratio is:

$$\begin{aligned}SN_T &= 19.963 - 3.2869x_1x_3 \\ &= 19.963 - 3.2869(-1)(1) \\ &= 23.2499\end{aligned}$$

To make the signal to noise ratio model hierarchical, x_1 and x_3, have to be included in the model. Then, the predicted SN ratio is:

$$\begin{aligned}SN_T &= 19.963 - 0.6292x_1 + 0.0515x_3 - 3.2869x_1x_3 \\ &= 19.963 - 0.6292(-1) + 0.0515(1) - 3.2869(-1)(1) \\ &= 23.9306\end{aligned}$$

Limitations of Taguchi's Approach

Although Taguchi's approach towards robust parameter design introduced innovative techniques to improve quality, a few concerns regarding his philosophy have been raised. Some of these concerns relate to the signal to noise ratios defined to reduce variations in the response, and some others are related to the absence of the means to test for higher order control factor interactions when his orthogonal arrays are used as inner arrays for the design. For these reasons, other approaches to carry out robust parameter design have been suggested including response modeling and the use of $\ln s_i^2$ in the place of the signal to noise ratios in the dispersion model. In response modeling, the noise factors are included in the model as additional factors, along with the other control factors. Details of these methods can be found in [Wu, 2000] and other theory books published on the subject.

Mixture Design

IN THIS CHAPTER

Introduction **332**

Mixture Design Types **333**
Simplex Plot 333
Simplex Lattice Design 335
Simplex Centroid Design 336
Simplex Axial Design 338
Extreme Vertex Design 339

Mixture Design Data Analysis **341**
Models Used in Mixture Design 341
L-Pseudocomponent, Proportion, and Actual Values 343
Response Trace Plot 346
Example 347

Mixture Design with Process Variables **353**
Model with Process Variables 354
Example 355

References **362**

Introduction

When a product is formed by mixing together two or more ingredients, the product is called a mixture, and the ingredients are called mixture components. In a general mixture problem, the measured response is assumed to depend only on the proportions of the ingredients in the mixture, not the amount of the mixture. For example, the taste of a fruit punch recipe (i.e., the response) might depend on the proportions of watermelon, pineapple and orange juice in the mixture. The taste of a small cup of fruit punch will obviously be the same as a big cup.

Sometimes the responses of a mixture experiment depend not only on the proportions of ingredients, but also on the settings of variables in the process of making the mixture. For example, the

tensile strength of stainless steel is not only affected by the proportions of iron, copper, nickel and chromium in the alloy; it is also affected by process variables such as temperature, pressure and curing time used in the experiment.

One of the purposes of conducting a mixture experiment is to find the best proportion of each component and the best value of each process variable, in order to optimize a single response or multiple responses simultaneously. In this chapter, we will discuss how to design effective mixture designs and how to analyze data from mixture experiments with and without process variables.

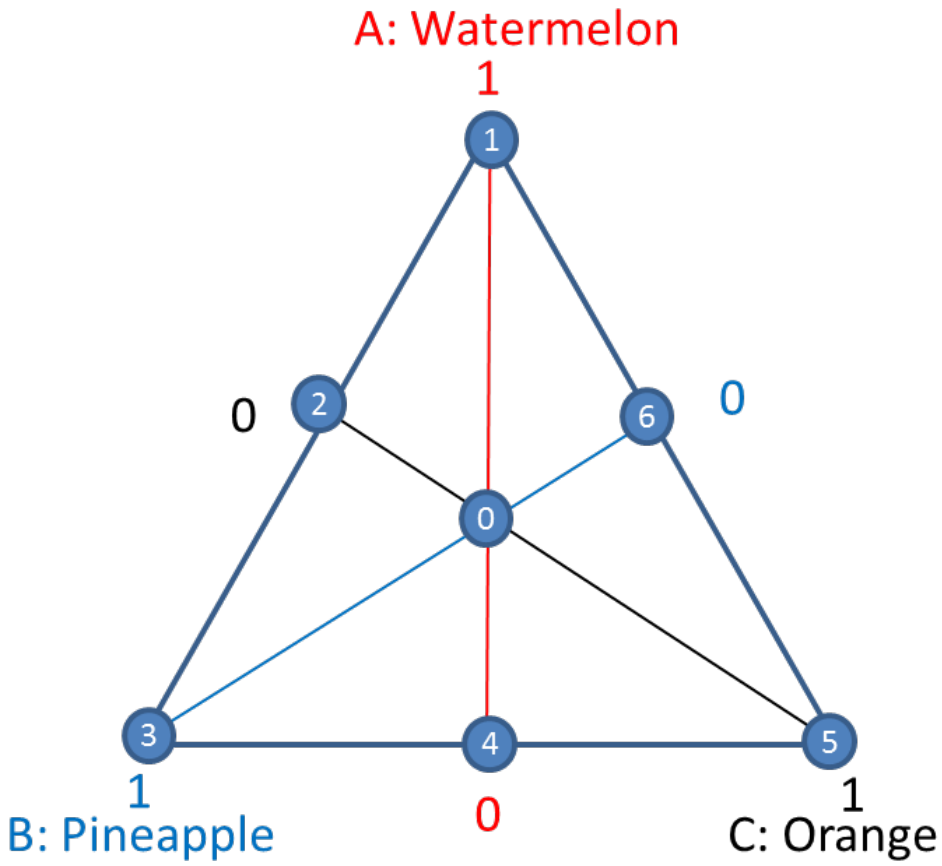
Mixture Design Types

There are several different types of mixture designs. The most common ones are simplex lattice, simplex centroid, simplex axial and extreme vertex designs, each of which is used for a different purpose.

- If there are many components in a mixture, the first choice is to screen out the most important ones. Simplex axial and Simplex centroid designs are used for this purpose.
- If the number of components is not large, but a high order polynomial equation is needed in order to accurately describe the response surface, then a simplex lattice design can be used.
- Extreme vertex designs are used for the cases when there are constraints on one or more components (e.g., if the proportion of watermelon juice in a fruit punch recipe is required to be less than 30%, and the combined proportion of watermelon and orange juice should always be between 40% and 70%).

Simplex Plot

Since the sum of all the mixture components is always 100%, the experiment space usually is given by a plot. The experiment space for the fruit punch experiment is given in the following triangle or simplex plot.



The triangle area in the above plot is defined by the fact that the sum of the three ingredients is 1 (100%). For the points that are on the vertices, the punch only has one ingredient. For instance, point 1 only has watermelon. The line opposite of point 1 represents a mixture with no watermelon.

The coordinate system used for the value of each ingredient x_i ,$i = 1, 2, \ldots, q$ is called a simplex coordinate system. q is the number of ingredients. The simplex plot can only visually display three ingredients. If there are more than three ingredients, the values for other ingredients must be provided. For the fruit punch example, the coordinate for point 1 is (1, 0, 0). The interior points of the triangle represent mixtures in which none of the three components is absent. It means all $x_i > 0$, $i = 1, 2, 3$. Point 0 in the middle of the triangle is called the center point. In this case, it is the centroid of a face/plane. The coordinate for point 0 is (1/3, 1/3, 1/3). Points 2, 4 and 6 are each called a centroid of edge. Their coordinates are (0.5, 0.5, 0), (0, 0.5, 0.5), and (0.5, 0, 0.5).

Simplex Lattice Design

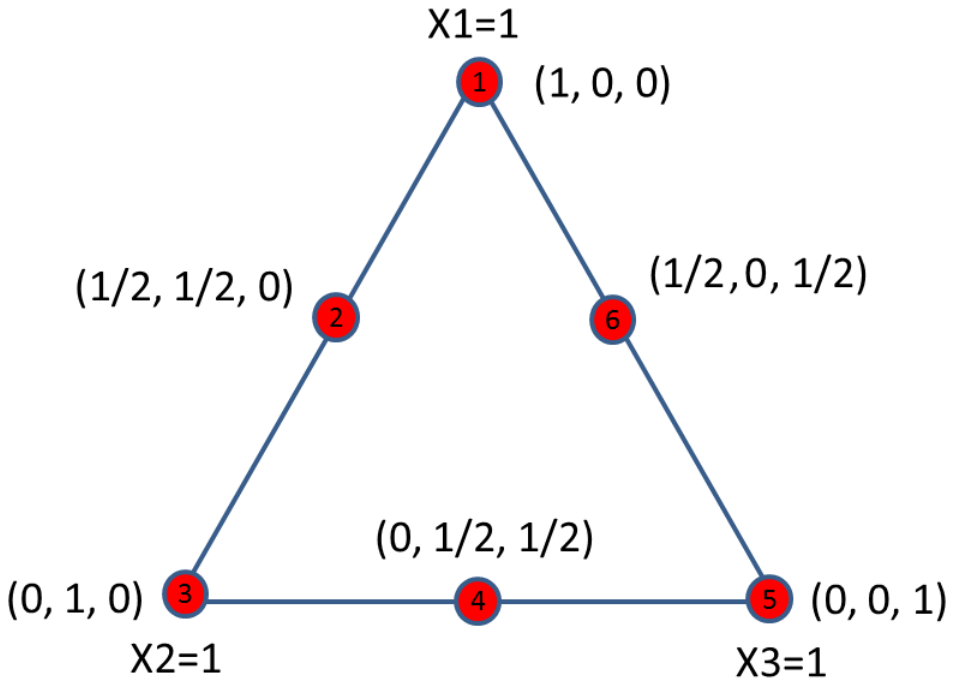
The response in a mixture experiment usually is described by a polynomial function. This function represents how the components affect the response. To better study the shape of the response surface, the natural choice for a design would be the one whose points are spread evenly over the whole simplex. An ordered arrangement consisting of a uniformly spaced distribution of points on a simplex is known as a lattice.

A {q, m} simplex lattice design for q components consists of points defined by the following coordinate settings: the proportions assumed by each component take the m+1 equally spaced values from 0 to 1,

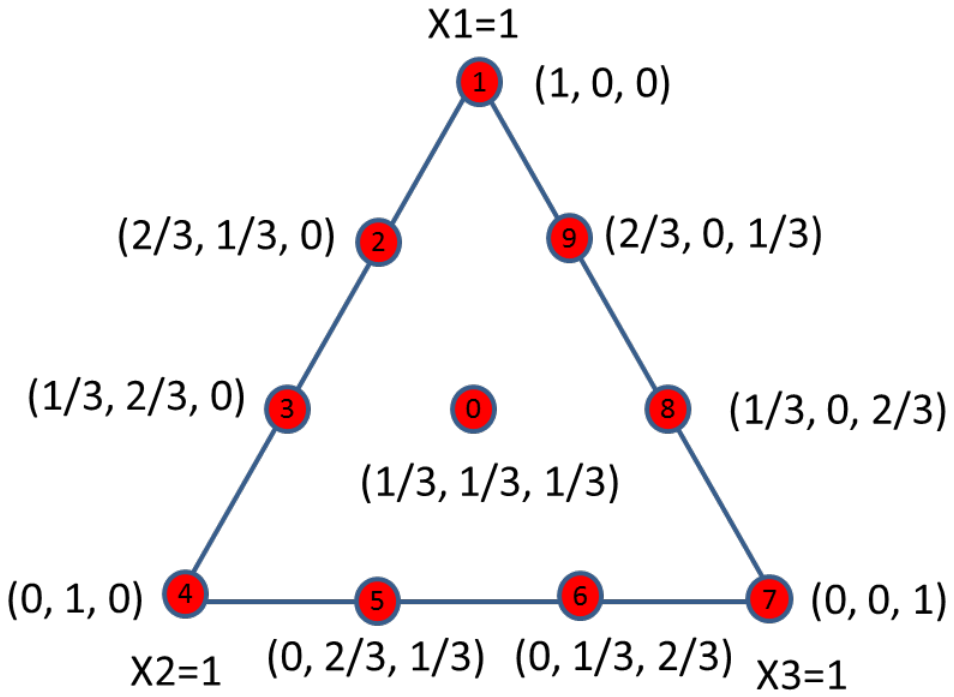
$$x_i = 0, \frac{1}{m}, \frac{2}{m}, \ldots, 1 \; i = 1, 2, \ldots, q$$

and the design space consists of all the reasonable combinations of all the values for each factor.

m is usually called the degree of the lattice. For example, for a {3, 2} design, $x_i = 0, \frac{1}{2}, 1$ and its design space has 6 points. They are:



For a {3, 3} design, $x_i = 0, \frac{1}{3}, \frac{2}{3}, 1$, and its design space has 10 points. They are:



For a simplex design with degree of m, each component has m + 1 different values, therefore, the experiment results can be used to fit a polynomial equation up to an order of m. A {3, 3} simplex lattice design can be used to fit the following model.

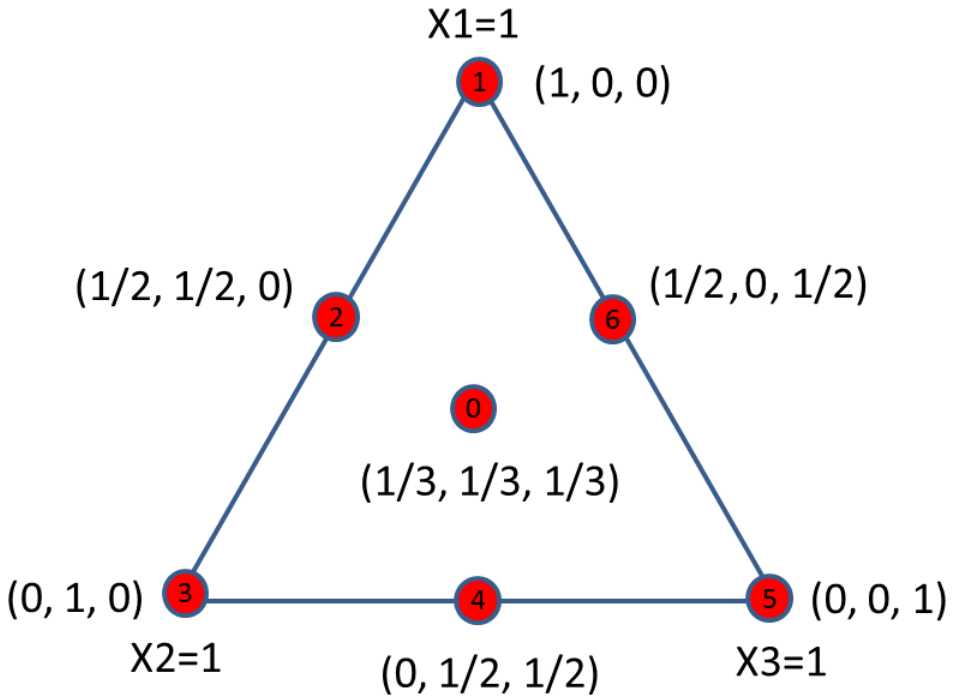
$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \delta_{12} x_1 x_2 (x_1 - x_2) + \delta_{13} x_1 x_3 (x_1 - x_3) + \delta_{23} x_2 x_3 (x_2 - x_3) + \beta_{123} x_1 x_2 x_3$$

The above model is called the full cubic model. Note that the intercept term is not included in the model due to the correlation between all the components (their sum is 100%).

Simplex lattice design includes all the component combinations. For a {q, m} design, the total number of runs is $\binom{q+m-1}{m}$. Therefore to reduce the number of runs and still be able to fit a high order polynomial model, sometimes we can use simplex centroid design which is explained next.

Simplex Centroid Design

A simplex centroid design only includes the centroid points. For the components that appear in a run in a simplex centroid design, they have the same values.



In the above simplex plot, points 2, 4 and 6 are 2nd degree centroids. Each of them has two non-zero components with equal values. Point 0 is a 3rd degree centroid and all three components have the same value. For a design with q components, the highest degree of centroid is q. It is called the overall centroid, or the center point of the design.

For a q component simplex centroid design with a degree of centroid of q, the total number of runs is $2^q - 1$. The runs correspond to the q permutations of (1, 0, 0,…, 0), $\binom{q}{2}$ permutations of (1/2, 1/2, 0, 0, 0, 0, ...,0), the $\binom{q}{3}$ permutations of (1/3, 1/3, 1/3, 0, 0, 0, 0,…, 0)…., and the overall centroid (1/q, 1/q, …, 1/q). If the degree of centroid is defined as m (m < q), then the total number of runs will be $\binom{q}{1} + \binom{q}{2} + \ldots + \binom{q}{m}$.

Since a simplex centroid design usually has fewer runs than a simplex lattice design with the same degree, a polynomial model with fewer terms should be used. A {3, 3} simplex centroid design can be used to fit the following model.

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3$$

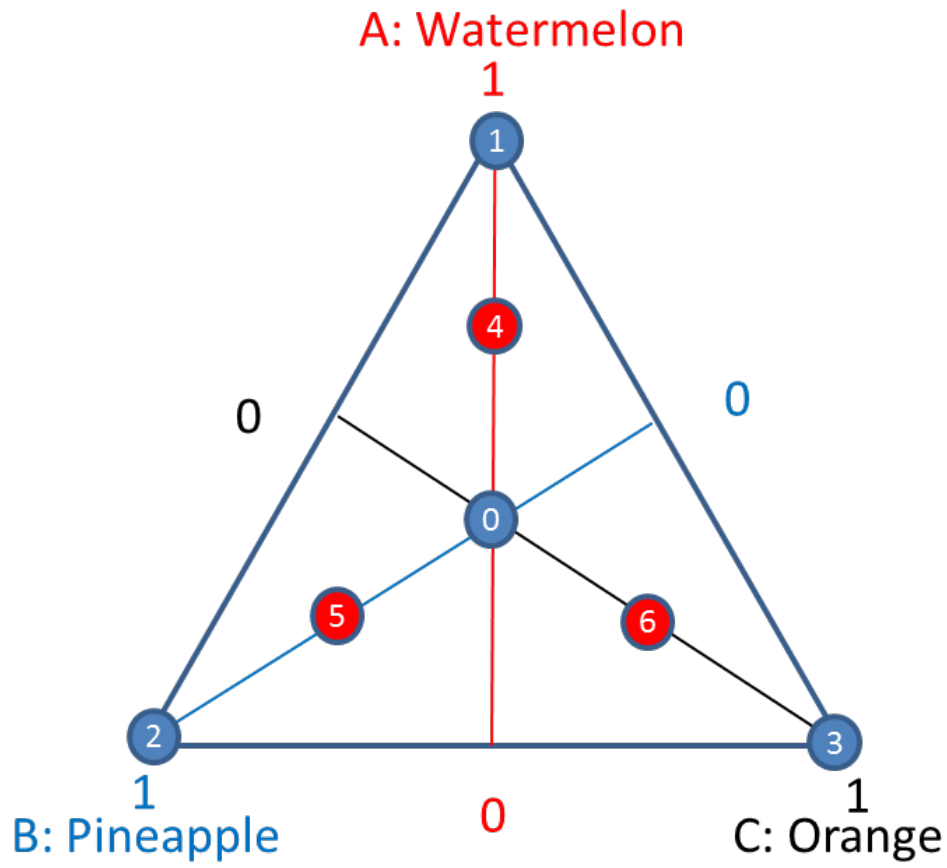
The above model is called the special cubic model. Note that the intercept term is not included due to the correlation between all the components (their sum is 100%).

Simplex Axial Design

The simplex lattice and simplex centroid designs are boundary designs since the points of these designs are positioned on boundaries (vertices, edges, faces, etc.) of the simplex factor space, with the exception of the overall centroid. Axial designs, on the other hand, are designs consisting mainly of the points positioned inside the simplex. Axial designs have been recommended for use when component effects are to be measured in a screening experiment, particularly when first degree models are to be fitted.

Definition of Axial: The axial of a component i is defined as the imaginary line extending from the base point $x_i = 0$, $x_j = 1/(q-1)$ for all $j \neq i$, to the vertex where $x_i = 1, x_j = 0$ all $j \neq i$. [John Cornell]

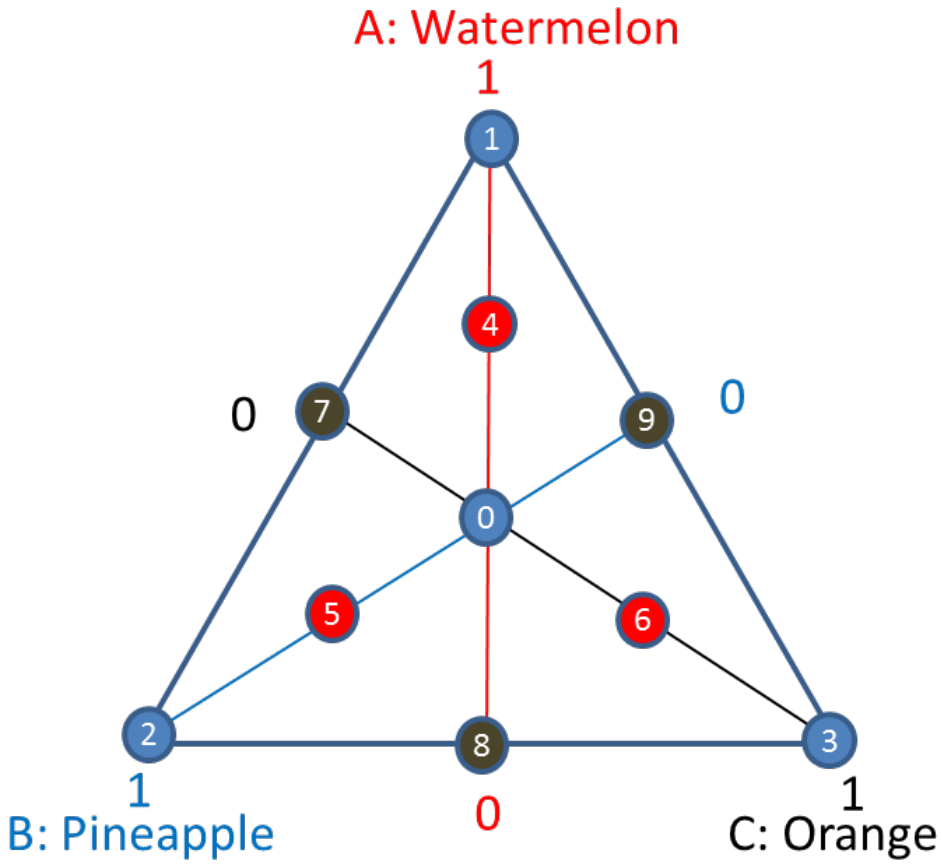
In a simplex axial design, all the points are on the axial. The simplest form of axial design is one whose points are positioned equidistant from the overall centroid $(1/q, 1/q, 1/q, \dots)$. Traditionally, points located at the half distance from the overall centroid to the vertex are called axial points/blends. This is illustrated in the following plot.



Points 4, 5 and 6 are the axial blends.

By default, a simple axial design in a DOE folio only has vertices, axial blends, centroid of the constraint planes and the overall centroid. For a design with q components, constraint plane centroids are the center points of dimension of q-1 space. One component is 0, and the remaining components have the same values for the center points of constraint planes. The number of the constraint plane centroids is the number of components q. The total number of runs in a simple axial design will be 3q+1. They are q vertex runs, q centroids of constraint planes, q axial blends and 1 overall centroid.

A simplex axial design for 3 components has 10 points as given below.



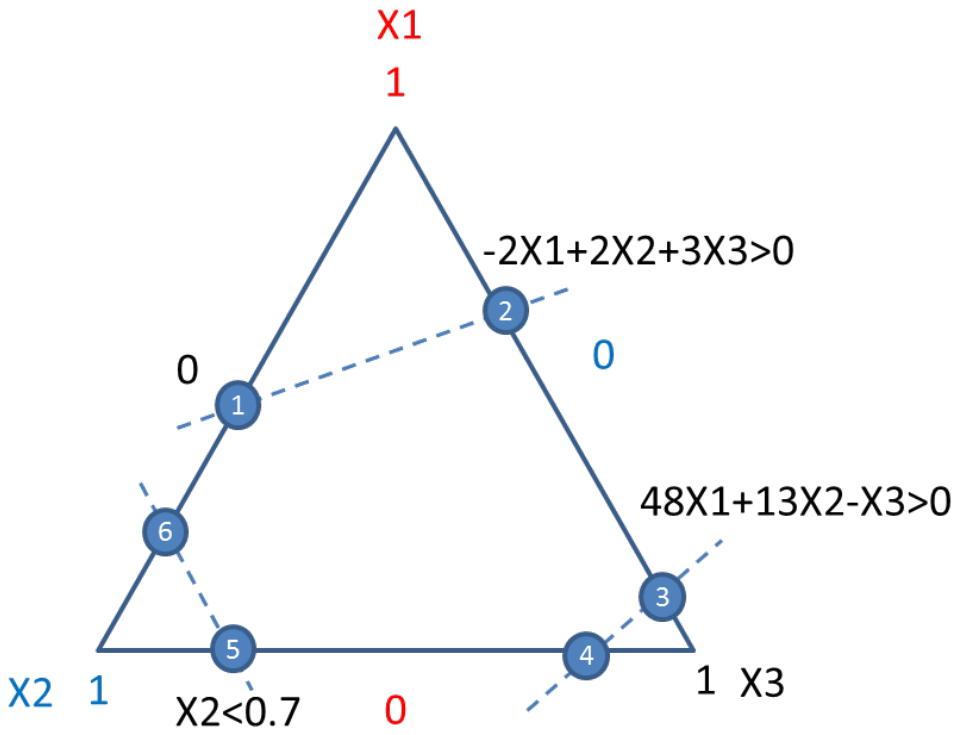
Points 1, 2 and 3 are the three vertices; points 4, 5, 6 are the axial blends; points 7, 8 and 9 are the centroids of constraint planes, and point 0 is the overall center point.

Extreme Vertex Design

Extreme vertex designs are used when both lower and upper bound constraints on the components are presented, or when linear constraints are added to several components. For example, if a mixture design with 3 components has the following constraints:

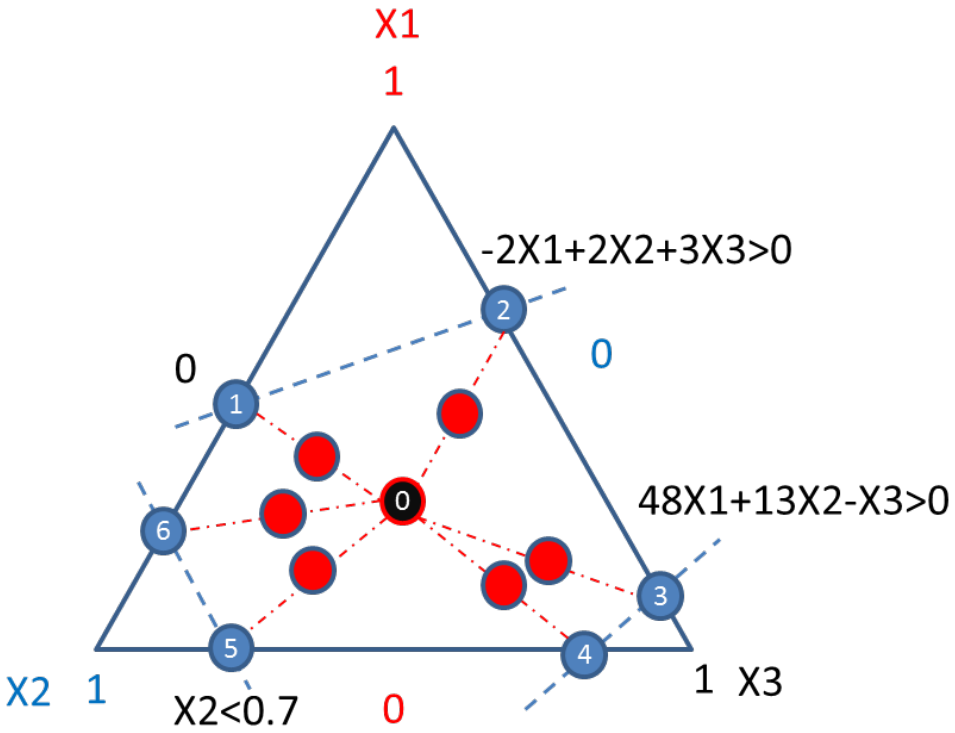
- $x_2 \leq 0.7$
- $-2x_1 + 2x_2 + 3x_3 \geq 0$
- $48x_1 + 13x_2 - x_3 \geq 0$

Then the feasible region is defined by the six points in the following simplex plot. To meet the above constraints, all the runs conducted in the experiment should be in the feasible region or on its boundary.



The CONSIM method described in [Snee 1979] is used in a Weibull++ DOE folio to check the consistency of all the constraints and to get the vertices defined by them.

Extreme vertex designs by default use the vertices at the boundary. Additional points such as the centroid of spaces of different dimensions, axial points and the overall center point can be added. In extreme vertex designs, axial points are between the overall center point and the vertices. For the above example, if the axial points and the overall center point are added, then all the runs in the experiment will be:



Point 0 in the center of the feasible region is the overall centroid. The other red points are the axial points. They are at the middle of the lines connecting the center point with the vertices.

Mixture Design Data Analysis

In the following section, we will discuss the most popular regression models in mixture design data analysis. Due to the correlation between all the components in mixture designs, the intercept term usually is not included in the regression model.

Models Used in Mixture Design

For a design with three components, the following models are commonly used.

- Linear model:

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

If the intercept were included in the model, then the linear model would be

$$y = \beta_0' + \beta_1' x_1 + \beta_2' x_2 + \beta_3' x_3$$

However, since $x_1 + x_2 + x_3 = 1$ (can be other constants as well), the above equation can be written as

$$\begin{aligned} y &= \beta_0' (x_1 + x_2 + x_3) + \beta_1' x_1 + \beta_2' x_2 + \beta_3' x_3 \\ &= (\beta_0' + \beta_1') x_1 + (\beta_0' + \beta_2') x_2 + (\beta_0' + \beta_3') x_3 \\ &= \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \end{aligned}$$

The equation has thus been reformatted to omit the intercept.

- Quadratic model:

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3$$

There are no classic quadratic terms such as x_1^2 . This is because

$$x_1^2 = x_1 (1 - x_2 - x_3) = x_1 - x_1 x_2 - x_1 x_3$$

- Full cubic model:

$$\begin{aligned} y &= \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 \\ &+ \delta_{12} x_1 x_2 (x_1 - x_2) + \delta_{13} x_1 x_3 (x_1 - x_3) + \delta_{23} x_2 x_3 (x_2 - x_3) \\ &+ \beta_{123} x_1 x_2 x_3 \end{aligned}$$

- Special cubic model: $\delta_{ij} x_i x_j (x_i - x_j)$ are removed from the full cubic model.

$$\begin{aligned} y &= \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 \\ &+ \beta_{123} x_1 x_2 x_3 \end{aligned}$$

The above types of models are called Scheffe type models. They can be extended to designs with more than three components.

In regular regression analysis, the effect of an exploratory variable or factor is represented by the value of the coefficient. The ratio of the estimated coefficient and its standard error is used for the t-test. The t-test can tell us if a coefficient is 0 or not. If a coefficient is statistically 0, then the corresponding factor has no significant effect on the response. However, for Scheffe type models, since the intercept term is not included in the model, we cannot use the regular t-test to test each individual main effect. In other words, we cannot test if the coefficient for each component is 0 or not.

Similarly, in the ANOVA analysis, the linear effects of all the components are tested together as a single group. The main effect test for each individual component is not conducted. To perform ANOVA analysis, the Scheffe type model needs to be reformatted to include the hidden intercept. For example, the linear model

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$

can be rewritten as

$$\begin{aligned} y &= \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \\ &= \beta_1 x_1 + \beta_2 x_2 + \beta_3 (1 - x_1 - x_2) \\ &= \beta_3 + (\beta_1 - \beta_3) x_1 + (\beta_2 - \beta_3) x_2 \\ &= \beta_0 + \beta_1' x_1 + \beta_2' x_2 \end{aligned}$$

where $\beta_0 = \beta_3$, $\beta_1' = \beta_1 - \beta_3$, $\beta_2' = \beta_2 - \beta_3$. All other models such as the quadratic, cubic and special cubic model can be reformatted using the same procedure. By including the intercept in the model, the correct sum of squares can be calculated in the ANOVA table. If ANOVA analysis is conducted directly using the Scheffe type models, the result will be incorrect.

L-Pseudocomponent, Proportion, and Actual Values

In mixture designs, the total amount of the mixture is usually given. For example, we can make either a one-pound or a two-pound cake. Regardless of whether the cake is one or two pounds, the proportion of each ingredient is the same. When the total amount is given, the upper and lower limits for each ingredient are usually given in amounts, which is easier for the experimenter to understand. Of course, if the limits or other constraints are given in terms of proportions, these proportions need be converted to the real amount values when conducting the experiment. To keep everything consistent, all the constraints in a DOE folio are treated as amounts.

In regular factorial design and response surface methods, the regression model is calculated using coded values. Coded values scale all the factors to the same magnitude, which makes the analysis much easier and reduces convergence error. Similarly, the analysis in mixture design is conducted using the so-called L-pseudocomponent value. L-pseudocomponent values scale all the components' values within 0 and 1. In a DOE folio all the designs and calculations for mixture factors are based on L-pseudocomponent values. The relationship between L-pseudocomponent values, proportions and actual amounts are explained next.

Example for L-Pseudocomponent Value

We are going to make one gallon (about 3.8 liters) of fruit punch. Three ingredients will be in the punch with the following constraints.

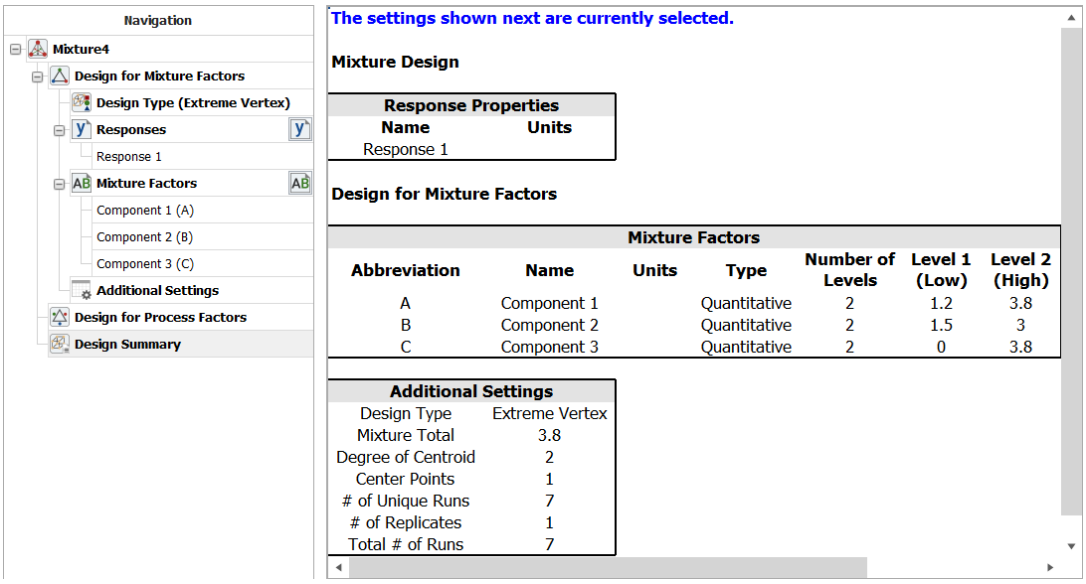
$$1.2 \le A \le 3.8 , 1.5 \le B \le 3 , 0 \le C \le 3.8$$

Let x_i^A (i = 1, 2, 3) be the actual amount value, x_i be the L-pseudocomponent value and x_i^R be the proportion value. Then the equations for the conversion between them are:

$$x_i = \frac{x_i^A - l_i}{\left(T - \sum_{j=1}^{p} l_j\right)}, \quad x_i^A = l_i + \left(T - \sum_{j=1}^{p} l_i\right) x_i, \quad x_i^R = \frac{x_i^A}{T}$$

where x_1, x_1^A and x_1^R are for component A, x_2, x_2^A and x_2^R are for component B, and x_3, x_3^A and x_3^R are for component C.

Since components in this example have both lower and upper limit constraints, an extreme vertex design is used. The design settings are given below.



The created design in terms of L-pseudocomponent values is:

	Standard Order	A:Component 1	B:Component 2	C:Component 3
1	1	0	1	0
2	2	0	0	1
3	3	1	0	0
4	4	0.333333333	0.333333333	0.333333333
5	5	0.5	0.5	0
6	6	0	0.5	0.5
7	7	0.5	0	0.5

Displayed in amount values, it is:

	Standard Order	A:Component 1	B:Component 2	C:Component 3
1	1	1.2	2.6	0
2	2	1.2	1.5	1.1
3	3	2.3	1.5	0
4	4	1.566666667	1.866666667	0.366666667
5	5	1.75	2.05	0
6	6	1.2	2.05	0.55
7	7	1.75	1.5	0.55

Displayed in proportion values, it is:

	Standard Order	A:Component 1	B:Component 2	C:Component 3
1	1	0.315789474	0.684210526	0
2	2	0.315789474	0.394736842	0.289473684
3	3	0.605263158	0.394736842	0
4	4	0.412280702	0.49122807	0.096491228
5	5	0.460526316	0.539473684	0
6	6	0.315789474	0.539473684	0.144736842
7	7	0.460526316	0.394736842	0.144736842

Check Constraint Consistency

In the above example, all the constraints are consistent. However, if we set the constraints to

$$1.2 \le A \le 3.8,\ 1.5 \le B \le 3,\ 2 \le C \le 3.8,$$

then they are not consistent. This is because the total is only 3.8, but the sum of all the lower limits is 4.7. Therefore, not all the lower limits can be satisfied at the same time. If only lower limits and upper limits are presented for all the components, then we can adjust the lower bounds to make the constraints consistent. The method given by [Pieple 1983] is used and summarized below.

Defined the range of a component to be $R_i = U_i - L_i$. U_i and L_i are the upper and lower limit for component i. The implied range of component i is $R_i^* = U_i^* - L_i^*$, where $L_i^* = T - \sum_{j \neq i}^{q} U_i$, and $U_i^* = T - \sum_{j \neq i}^{q} L_i$. T is the total amount. The steps for checking and adjusting bounds are given below.

Step 1: Check if L_i^* and U_i^* are greater than 0, if they are, then these constraints meet the basic requirement to be consistent. We can move forward to step 2. If not, these constraints cannot be adjusted to be consistent. We should stop.

Step 2: For each component, check if $L_i \geq L_i^*$ and $U_i \leq U_i^*$. If they are, then this component's constraints are consistent. Otherwise, if $L_i < L_i^*$, then set $L_i = L_i^*$, and if $U_i > U_i^*$, then set $U_i = U_i^*$.

Step 3: Whenever a bound is changed, restart from Step 1 to use the new bound to check if all the constraints are consistent. Repeat this until all the limits are consistent.

For extreme vertex design where linear constraints are allowed, the DOE folio will give a warning and stop creating the design if inconsistent linear combination constraints are found. No adjustment will be conducted for linear constraints.

Response Trace Plot

Due to the correlation between all the components, the regular t-test is not used to test the significance of each component. A special plot called the Response Trace Plot can be used to see how the response changes when each component changes from its reference point [John Cornell].

A reference point can be any point inside the experiment space. An imaginary line can be drawn from this reference point to each vertex $x_i = 1$, and $x_j = 0$ $(i \neq j)$. This line is the direction for component i to change. Component i can either increase or decrease its value along this line, while the ratio of other components x_j/x_k $(j, k \neq i)$ will keep constant. If the simplex plot is defined in terms of proportion, then the direction is called Cox's direction, and x_j/x_k is the ratio of proportion. If the simplex plot is defined in terms of pseudocomponent value, then the direction is called Pieple's direction, and x_j/x_k will be the ratio of pseudocomponent values.

Assume the reference point in terms of proportion is $s = (s_1, s_2, \ldots, s_q)$ where $s_1 + s_2 + \ldots + s_q = 1$. Suppose the proportion of component i at s_i is now changed by Δ_i (Δ_i could be greater than or less than 0) in Cox's direction, so that the new proportion becomes $x_i = s_i + \Delta_i$

Then the proportions of the remaining $q - 1$ components resulting from the change from s_i will be

$$x_j = s_j - \frac{\Delta_i s_j}{1 - s_i}$$

After the change, the ratio of component j and k is unchanged. This is because

$$\frac{x_j}{x_k} = \frac{s_j - \frac{\Delta_i s_j}{1-s_i}}{s_k - \frac{\Delta_i s_k}{1-s_i}} = \frac{s_j}{s_k}\frac{\frac{\Delta_i}{1-s_i}}{\frac{\Delta_i}{1-s_i}} = \frac{s_j}{s_k}$$

While x_i is changed along Cox's direction, we can use a fitted regression model to get the response value y. A response trace plot for a mixture design with three components will look like

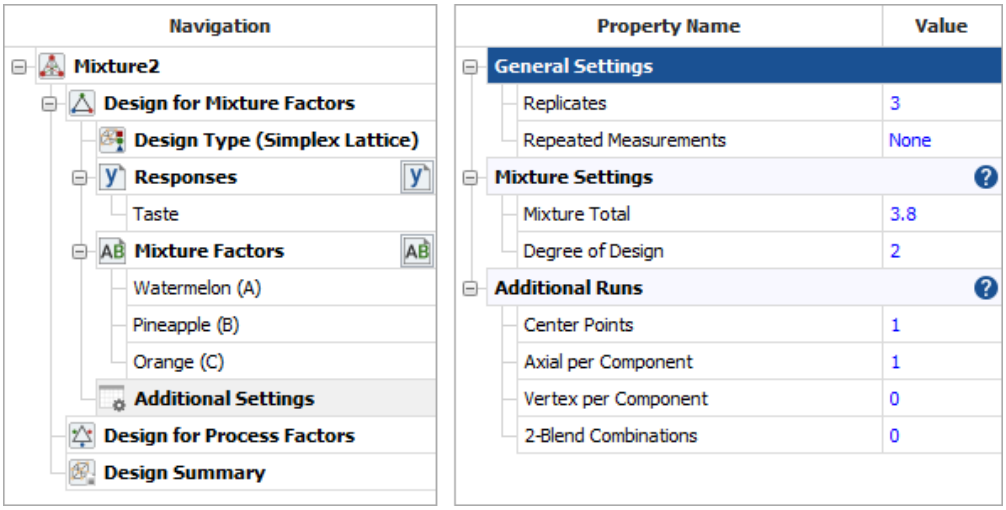
The x-axis is the deviation amount from the reference point, and the y-value is the fitted response. Each component has one curve. Since the red curve for component A changes significantly, this means it has a significant effect along its axial. The blue curve for component C is almost flat; this means when C changes along Cox's direction and other components keep the same ratio, the response Y does not change very much. The effect of component B is between component A and C.

Example

Watermelon (A), pineapple (B) and orange juice (C) are used for making 3.8 liters of fruit punch. At least 30% of the fruit punch must be watermelon. Therefore the constraints are

$$1.14 \le A \le 3.8\,,\ 0 \le B \le 3.8\,,\ 0 \le C \le 3.8,$$

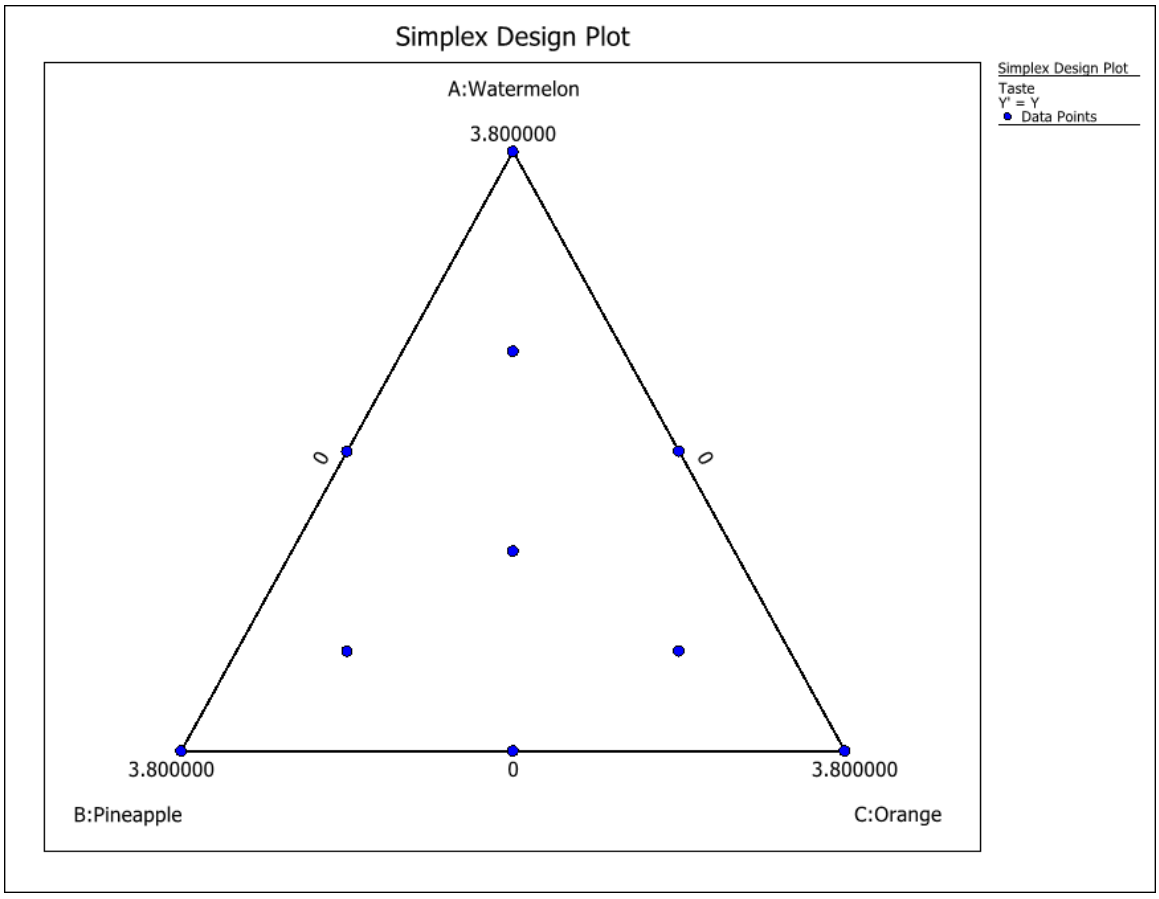
Different blends of the three-juice recipe were evaluated by a panel. A value from 1 (extremely poor) to 9 (very good) is used for the response [John Cornell, page 74]. A {3, 2} simplex lattice design is used with one center point and three axial points. Three replicates were conducted for each ingredient combination. The settings for creating this design in a DOE folio is



The generated design in L-pseudocomponent values and the response values from the experiment are

	Standard Order	Point Type	A:Watermelon	B:Pineapple	C:Orange	Taste
1	1	1	1	0	0	4.3
2	2	2	0.5	0.5	0	6.3
3	3	2	0.5	0	0.5	6.5
4	4	1	0	1	0	6.2
5	5	2	0	0.5	0.5	6.9
6	6	1	0	0	1	6.1
7	7	0	0.333333333	0.333333333	0.333333333	6
8	8	-1	0.666666667	0.166666667	0.166666667	5.4
9	9	-1	0.166666667	0.666666667	0.166666667	5.7
10	10	-1	0.166666667	0.166666667	0.666666667	5.2
11	11	1	1	0	0	4.7
12	12	2	0.5	0.5	0	5.8
13	13	2	0.5	0	0.5	6.2
14	14	1	0	1	0	6.2
15	15	2	0	0.5	0.5	7
16	16	1	0	0	1	6.5
17	17	0	0.333333333	0.333333333	0.333333333	5.8
18	18	-1	0.666666667	0.166666667	0.166666667	5.8
19	19	-1	0.166666667	0.666666667	0.166666667	5
20	20	-1	0.166666667	0.166666667	0.666666667	6.4
21	21	1	1	0	0	4.8
22	22	2	0.5	0.5	0	6.1
23	23	2	0.5	0	0.5	6.3
24	24	1	0	1	0	6.1
25	25	2	0	0.5	0.5	7.4
26	26	1	0	0	1	5.9
27	27	0	0.333333333	0.333333333	0.333333333	6.4
28	28	-1	0.666666667	0.166666667	0.166666667	6.6
29	29	-1	0.166666667	0.666666667	0.166666667	5.6
30	30	-1	0.166666667	0.166666667	0.666666667	6.4

The simplex design point plot is



Main effect and 2-way interactions are included in the regression model. The result for the regression model in terms of L-pseudocomponents is

$$y = 4.81x_1 + 6.03x_2 + 6.16x_3 + 1.13x_1x_2 + 2.45x_1x_3 + 1.69x_2x_3$$

The regression information table is

Regression Information							
Term	**Coef-ficient**	**Stand-ard Error**	**Low Con-fidence**	**High Con-fidence**	**T Value**	**P Value**	**Vari-ance Infla-tion Factor**
A: Water-melon	4.8093	0.3067	4.2845	5.3340			1.9636
B: Pine-apple	6.0274	0.3067	5.5027	6.5522			1.9636

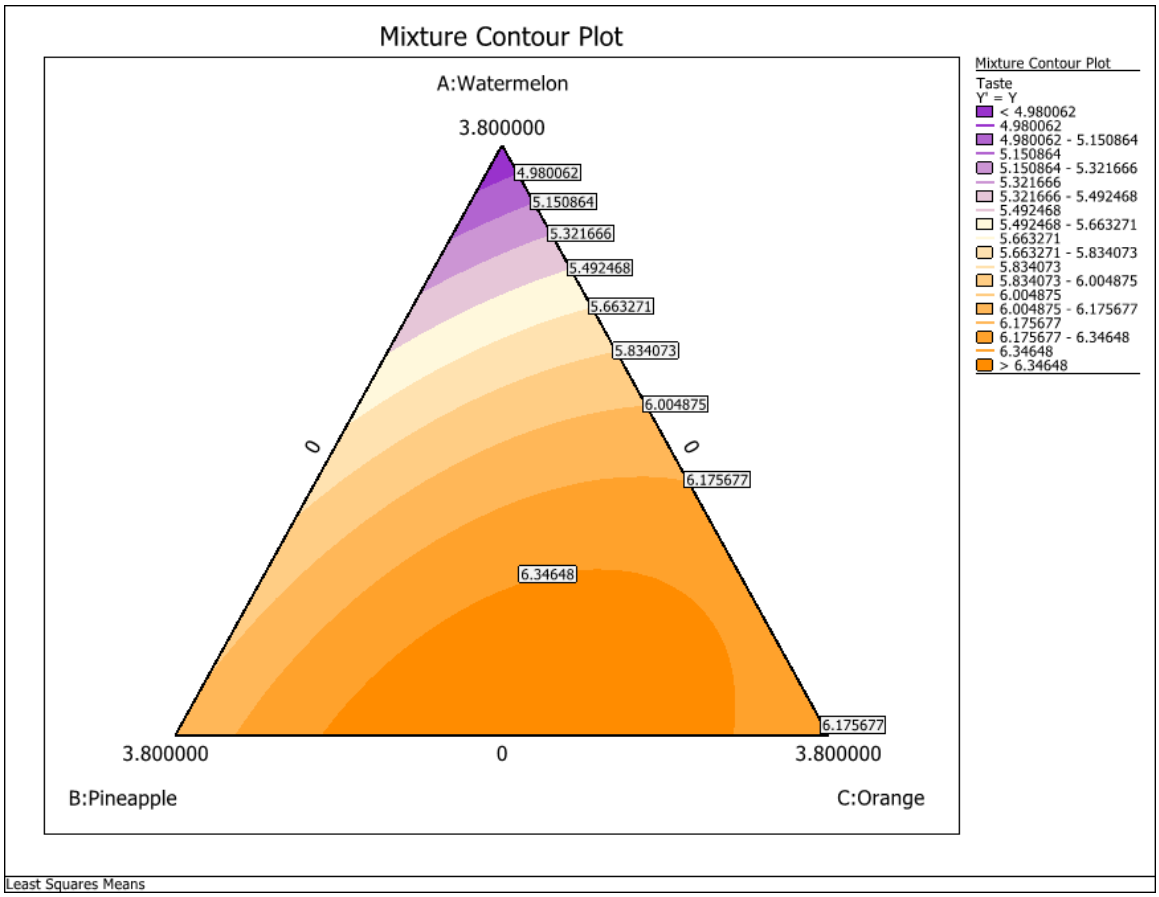
C: Orange	6.1577	0.3067	5.6330	6.6825			1.9636
A • B	1.1253	1.4137	-1.2934	3.5439	0.796-0	0.433-9	1.9819
A • C	2.4525	1.4137	0.0339	4.8712	1.734-8	0.095-6	1.9819
B • C	1.6889	1.4137	-0.7298	4.1075	1.194-7	0.243-9	1.9819

The result shows that the taste of the fruit punch is significantly affected by the interaction between watermelon and orange.

The ANOVA table is

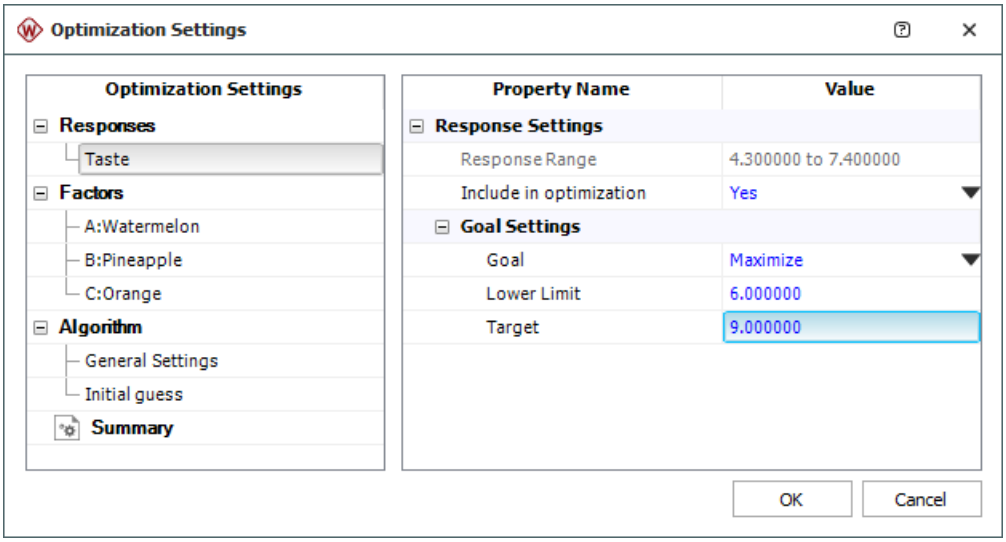
Anova Table					
Source of Variation	**Degrees of Freedom**	**Standard ErrorSum of Squares [Partial]**	**Mean Squares [Partial]**	**F Ratio**	**P Value**
Model	5	6.5517	1.3103	4.3181	0.0061
Linear	2	3.6513	1.8256	6.0162	0.0076
A • B	1	0.1923	0.1923	0.6336	0.4339
A • C	1	0.9133	0.9133	3.0097	0.0956
B • C	1	0.4331	0.4331	1.4272	0.2439
Residual	24	7.2829	0.3035		
Lack of Fit	4	4.4563	1.1141	7.8825	0.0006
Pure Error	20	2.8267	0.1413		
Total	29	13.8347			

The simplex contour plot in L-pseudocomponent values is

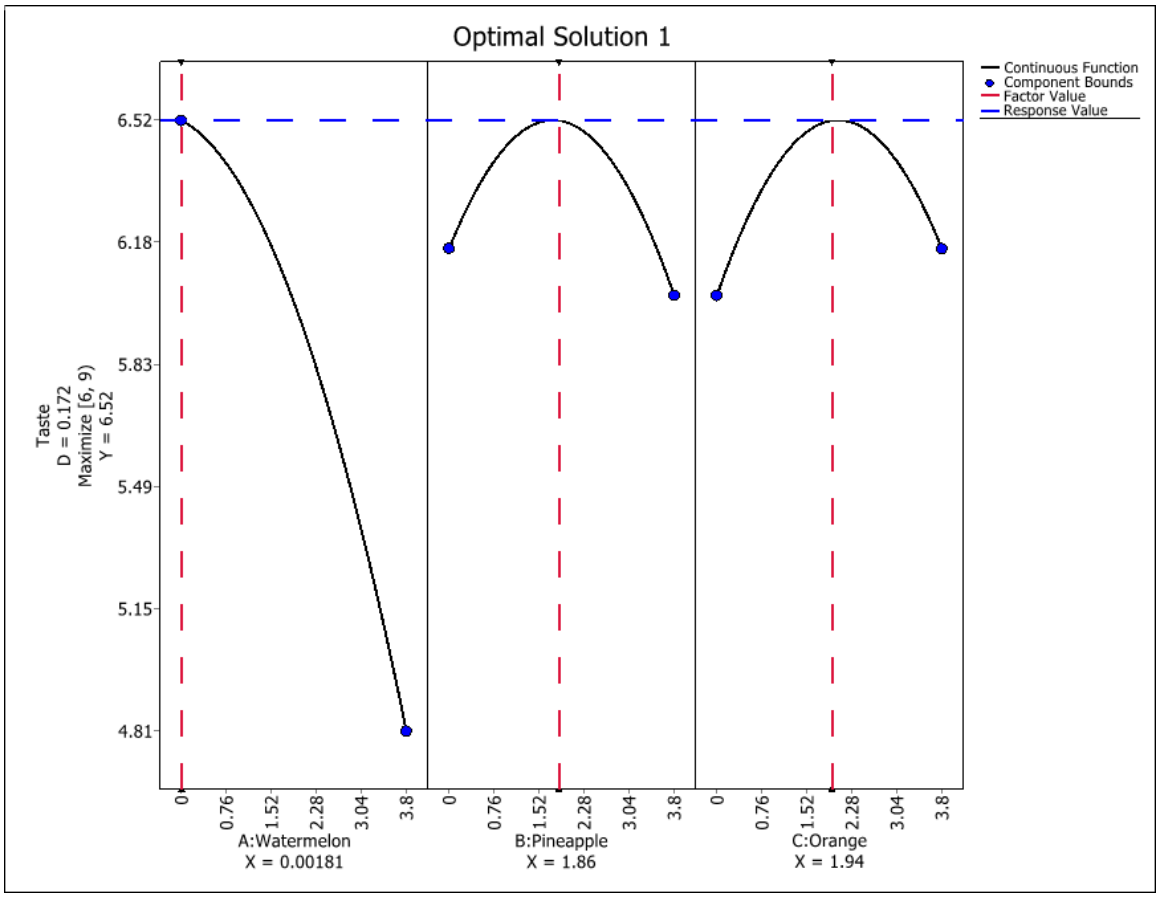


From this plot we can see that as the amount of watermelon is reduced, the taste of the fruit punch becomes better.

In order to find the best proportion of each ingredient, the optimization tool in the Weibull++ DOE folio can be utilized. Set the settings as



The resulting optimal plot is



This plot shows that when the amounts for watermelon, pineapple and orange juice are 1.141, 1.299 and 1.359, respectively, the rated taste of the fruit punch is highest.

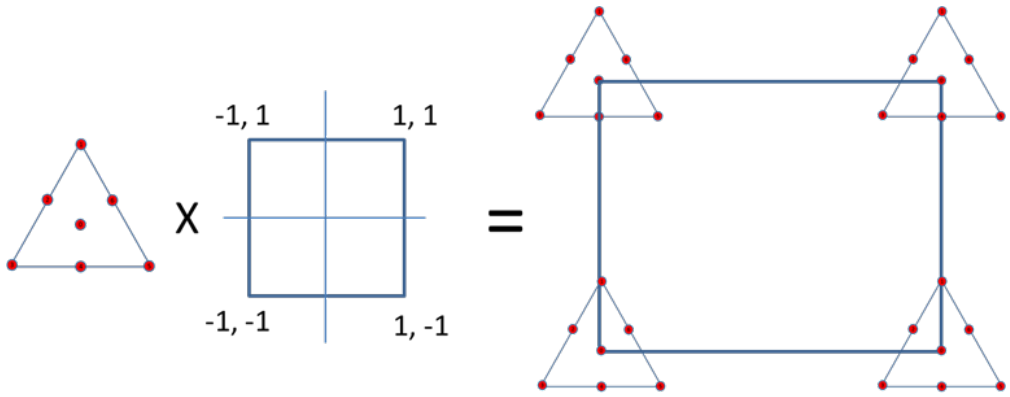
Mixture Design with Process Variables

Process variables often play very important roles in mixture experiments. A simple example is baking a cake. Even with the same ingredients, different baking temperatures and baking times can produce completely different results. In order to study the effect of process variables and find their best settings, we need to consider them when conducting a mixture experiment.

An easy way to do this is to make mixtures with the same ingredients in different combinations of process variables. If all the process variables are independent, then we can plan a regular factorial design for these process variables. By combining these designs with a separated mixture design, the effect of mixture components and effect of process variables can be studied.

For example, a {3, 2} simplex lattice design is used for a mixture with 3 components. Together with the center point, it has total of 7 runs or 7 different ingredient combinations. Assume 2 process variables are potentially important and a two level factorial design is used for them. It has a

total of 4 combinations for these 2 process variables. If the 7 different mixtures are made under each of the 4 process variable combinations, then the experiment has a total of 28 runs. This is illustrated in the figure below.



Of course, if it is possible, all the 28 experiments should be conducted in a random order.

Model with Process Variables

In a DOE folio, regression models including both mixture components and process variables are available. For mixture components, we use L-pseudocomponent values, and for process variables coded values are used.

Assume a design has 3 mixture components and 2 process variables, as illustrated in the above figure. We can use the following models for them.

- For the 3 mixture components, the following special cubic model is used.

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{123} x_1 x_2 x_3$$

- For the 2 process variables the following model is used.

$$y = \alpha_0 + \alpha_1 z_1 + \alpha_2 z_2 + \alpha_{12} z_1 z_2$$

- The combined model with both mixture components and process variables is

$$
\begin{aligned}
y = & \sum_{i=1}^{3} \gamma_i^0 x_i + \sum \sum_{i<j}^{3} \gamma_{ij}^0 x_i x_j + \gamma_{123}^0 x_1 x_2 x_3 \\
& + \left(\sum_{i=1}^{3} \gamma_i^1 x_i + \sum \sum_{i<j}^{3} \gamma_{ij}^1 x_i x_j + \gamma_{123}^1 x_1 x_2 x_3 \right) z_1 \\
& + \left(\sum_{i=1}^{3} \gamma_i^2 x_i + \sum \sum_{i<j}^{3} \gamma_{ij}^2 x_i x_j + \gamma_{123}^2 x_1 x_2 x_3 \right) z_2 \\
& + \left(\sum_{i=1}^{3} \gamma_i^{12} x_i + \sum \sum_{i<j}^{3} \gamma_{ij}^{12} x_i x_j + \gamma_{123}^{12} x_1 x_2 x_3 \right) z_1 z_2
\end{aligned}
$$

The above combined model has total of 7x4=28 terms. By expanding it, we get the following model:

$$
\begin{aligned}
y = & \gamma_1^0 x_1 + \gamma_2^0 x_2 + \gamma_3^0 x_3 + \gamma_{12}^0 x_1 x_2 + \gamma_{13}^0 x_1 x_3 + \gamma_{23}^0 x_2 x_3 + \gamma_{123}^0 x_1 x_2 x_3 \\
& + \gamma_1^1 x_1 z_1 + \gamma_2^1 x_2 z_1 + \gamma_3^1 x_3 z_1 + \gamma_{12}^1 x_1 x_2 z_1 + \gamma_{13}^1 x_1 x_3 z_1 + \gamma_{23}^1 x_2 x_3 z_1 + \gamma_{123}^1 x_1 x_2 x_3 z_1 \\
& + \gamma_1^2 x_1 z_2 + \gamma_2^2 x_2 z_2 + \gamma_3^2 x_3 z_2 + \gamma_{12}^2 x_1 x_2 z_2 + \gamma_{13}^2 x_1 x_3 z_2 + \gamma_{23}^2 x_2 x_3 z_2 + \gamma_{123}^2 x_1 x_2 x_3 z_2 \\
& + \gamma_1^{12} x_1 z_1 z_2 + \gamma_2^{12} x_2 z_1 z_2 + \gamma_3^{12} x_3 z_1 z_2 + \gamma_{12}^{12} x_1 x_2 z_1 z_2 + \gamma_{13}^{12} x_1 x_3 z_1 z_2 + \gamma_{23}^{12} x_2 x_3 z_1 z_2 + \gamma_{123}^{12} x_1 x_2 x_3 z_1 z_2
\end{aligned}
$$

The combined model basically crosses every term in the mixture components model with every term in the process variables model. From a mathematical point of view, this model is just a regular regression model. Therefore, the traditional regression analysis method can still be used for obtaining the model coefficients and calculating the ANOVA table.

Example

Three kinds of meats (beef, pork and lamb) are mixed together to form burger patties. The meat comprises 90% of the total mixture, with the remaining 10% reserved for flavoring ingredients. A {3, 2} simplex design with the center point is used for the experiment. The design has 7 meat combinations, which are given below using L-pseudocomponent values.

A: Beef	B: Pork	C: Lamb
1	0	0
0.5	0.5	0
0.5	0	0.5
0	1	0
0	0.5	0.5
0	0	1

0.333333	0.333333	0.333333

Two process variables on making the patties are also studied: cooking temperature and cooking time. The low and high temperature values are 375'F and 425'F, and the low and high time values are 25 and 40 minutes. A two level full factorial design is used and displayed below with coded values.

Temperature	Time
-1	-1
-1	1
1	-1
1	1

One of the properties of the burger patties is texture. The texture is measured by a compression test that measures the grams of force required to puncture the surface of the patty.

Combining the simplex design and the factorial design together, we get the following 28 runs. The corresponding texture reading for each blend is also provided.

Standard Order	A: Beef	B: Pork	C: Lamb	Z1: Temperature	Z2: Time	Texture (10^3 gram)
1	1	0	0	-1	-1	1.84
2	0.5	0.5	0	-1	-1	0.67
3	0.5	0	0.5	-1	-1	1.51
4	0	1	0	-1	-1	1.29
5	0	0.5	0.5	-1	-1	1.42
6	0	0	1	-1	-1	1.16
7	0.333	0.333	0.333	-1	-1	1.59
8	1	0	0	1	-1	2.86
9	0.5	0.5	0	1	-1	1.1

10	0.5	0	0.5	1	-1	1.6
11	0	1	0	1	-1	1.53
12	0	0.5	0.5	1	-1	1.81
13	0	0	1	1	-1	1.5
14	0.333	0.333	0.333	1	-1	1.68
15	1	0	0	-1	1	3.01
16	0.5	0.5	0	-1	1	1.21
17	0.5	0	0.5	-1	1	2.32
18	0	1	0	-1	1	1.93
19	0	0.5	0.5	-1	1	2.57
20	0	0	1	-1	1	1.83
21	0.333	0.3333	0.333	-1	1	1.94
22	1	0	0	1	1	4.13
23	0.5	0.5	0	1	1	1.67
24	0.5	0	0.5	1	1	2.57
25	0	1	0	1	1	2.26
26	0	0.5	0.5	1	1	3.15
27	0	0	1	1	1	2.22
28	0.333	0.333	0.333	1	1	2.6

Using a quadratic model for the mixture component and a 2-way interaction model for the process variables, we get the following results.

Term	Coefficient	Standard Error	T Value	P Value	Variance Inflation Factor
A:Beef	2.9421	0.1236	*	*	1.5989

B:Pork	1.7346	0.1236	*	*	1.5989
C:Lamb	1.6596	0.1236	*	*	1.5989
A • B	-4.4170	0.5680	-7.7766	0.0015	1.5695
A • C	-0.9170	0.5680	-1.6146	0.1817	1.5695
B • C	2.4480	0.5680	4.3099	0.0125	1.5695
Z1 • A	0.5324	0.1236	4.3084	0.0126	1.5989
Z1 • B	0.1399	0.1236	1.1319	0.3209	1.5989
Z1 • C	0.1799	0.1236	1.4557	0.2192	1.5989
Z1 • A • B	-0.4123	0.5680	-0.7260	0.5081	1.5695
Z1 • A • C	-1.0423	0.5680	-1.8352	0.1404	1.5695
Z1 • B • C	0.3727	0.5680	0.6561	0.5476	1.5695
Z2 • A	0.6193	0.1236	5.0117	0.0074	1.5989
Z2 • B	0.3518	0.1236	2.8468	0.0465	1.5989
Z2 • C	0.3568	0.1236	2.8873	0.0447	1.5989
Z2 • A • B	-0.9802	0.5680	-1.7258	0.1595	1.5695
Z2 • A • C	-0.3202	0.5680	-0.5638	0.6030	1.5695
Z2 • B • C	0.9248	0.5680	1.6282	0.1788	1.5695
Z1 • Z2 • A	0.0177	0.1236	0.1433	0.8930	1.5989
Z1 • Z2 • B	0.0152	0.1236	0.1231	0.9080	1.5989
Z1 • Z2 • C	0.0052	0.1236	0.0422	0.9684	1.5989
Z1 • Z2 • A • B	0.0808	0.5680	0.1423	0.8937	1.5695
Z1 • Z2 • A • C	0.2308	0.5680	0.4064	0.7052	1.5695
Z1 • Z2 • B • C	0.2658	0.5680	0.4680	0.6641	1.5695

The above table shows that all the terms with $z_1 \times z_2$ have very large P values, therefore, we can remove these terms from the model. We can also remove other terms with P values larger than 0.5. After recalculating with the desired terms, the final results are

Term	Coefficient	Standard Error	T Value	P Value	Variance Inflation Factor
A:Beef	2.9421	0.0875	*	*	1.5989
B:Pork	1.7346	0.0875	*	*	1.5989
C:Lamb	1.6596	0.0875	*	*	1.5989
A • B	-4.4170	0.4023	-10.9782	6.0305E-08	1.5695
A • C	-0.9170	0.4023	-2.2792	0.0402	1.5695
B • C	2.4480	0.4023	6.0842	3.8782E-05	1.5695
Z1 • A	0.4916	0.0799	6.1531	3.4705E-05	1.3321
Z1 • B	0.1365	0.0725	1.8830	0.0823	1.0971
Z1 • C	0.2176	0.0799	2.7235	0.0174	1.3321
Z1 • A • C	-1.0406	0.4015	-2.5916	0.0224	1.5631
Z2 • A	0.5910	0.0800	7.3859	5.3010E-06	1.3364
Z2 • B	0.3541	0.0875	4.0475	0.0014	1.5971
Z2 • C	0.3285	0.0800	4.1056	0.0012	1.3364
Z2 • A • B	-0.9654	0.4019	-2.4020	0.0320	1.5661
Z2 • B • C	0.9396	0.4019	2.3378	0.0360	1.5661

The regression model is

$$y = 2.9421x_1 + 1.7346x_2 + 1.6596x_3 - 4.4170x_1x_2 - 0.9170x_1x_3 + 2.4480x_2x_3 + 0.4916x_1z_1 + 0.1365x_2z_1 + 0.2176x_3z_1 - 1.0406x_1x_3z_1 + 0.5910x_1z_2 + 0.3541x_2z_2 + 0.3285x_3z_2 - 0.9654x_1x_2z_2 + 0.9396x_2x_3z_2$$

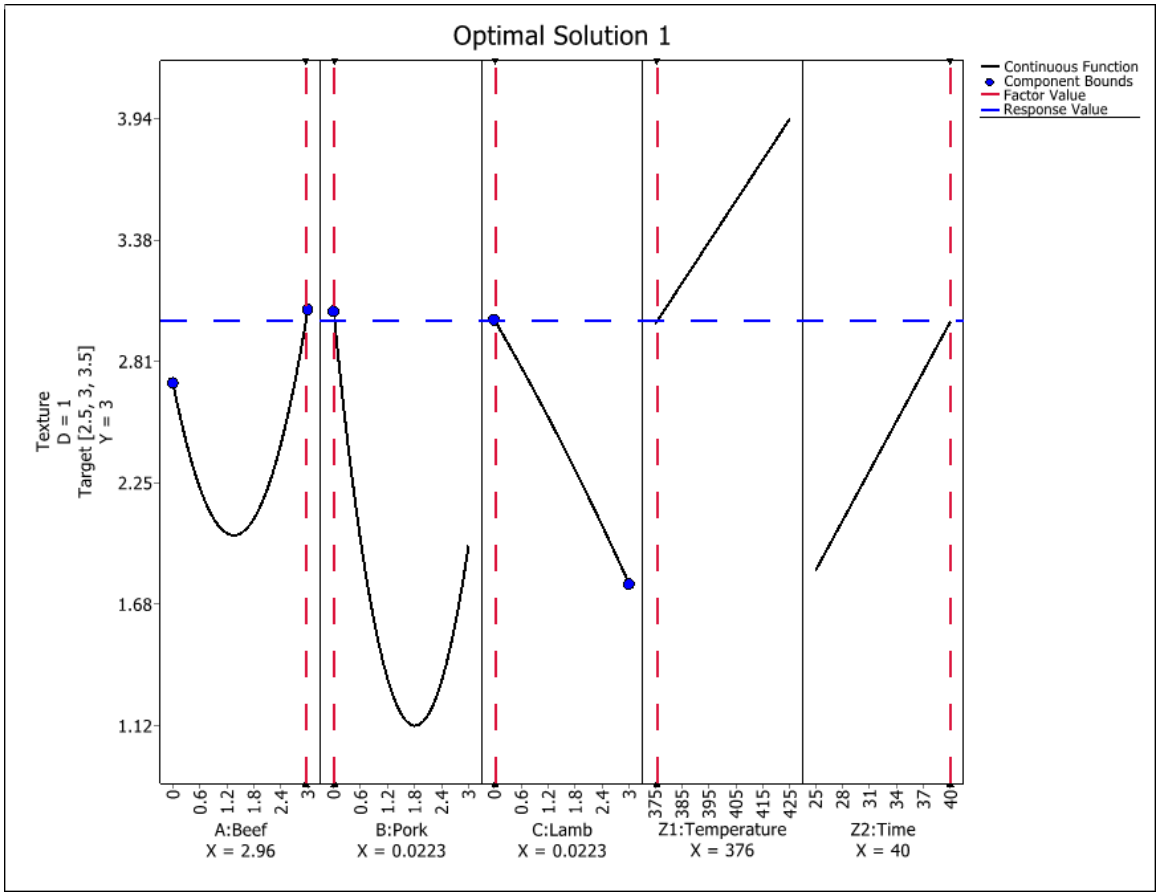
The ANOVA table for this model is

ANOVA Table					
Source of Variation	Degrees of Freedom	Sum of Squares [Partial]	Mean Squares [Partial]	F Ratio	P Value
Model	14	14.5066	1.0362	33.5558	6.8938E-08
Component Only					
Linear	2	4.1446	2.0723	67.1102	1.4088E-07
A • B	1	3.7216	3.7216	120.5208	6.0305E-08
A • C	1	0.1604	0.1604	5.1949	0.0402
B • C	1	1.1431	1.1431	37.0173	3.8782E-05
Component • Z1					
Z1 • A	1	1.1691	1.1691	37.8604	3.4705E-05
Z1 • B	1	0.1095	0.1095	3.5456	0.0823
Z1 • C	1	0.2290	0.2290	7.4172	0.0174
Z1 • A • C	1	0.2074	0.2074	6.7165	0.0224
Component • Z2					
Z2 • A	1	1.6845	1.6845	54.5517	5.3010E-06
Z2 • B	1	0.5059	0.5059	16.3819	0.0014
Z2 • C	1	0.5205	0.5205	16.8556	0.0012
Z2 • A • B	1	0.1782	0.1782	5.7698	0.0320

Z2 • B • C	1	0.1688	0.1688	5.4651	0.0360
Residual	13	0.4014	0.0309		
Lack of Fit	13	0.4014	0.0309		
Total	27	14.9080			

The above table shows both process factors have significant effects on the texture of the patties. Since the model is pretty complicate, the best settings for the process variables and for components cannot be easily identified.

The optimization tool in the DOE folio is used for the above model. The target texture value is 3×10^3 grams with an acceptable range of $2.5 - 3.5 \times 10^3$ grams.



The optimal solution is Beef = 2.95547, Pork = .022265, Lamb = 0.22265, Temperature = 375.7217, and Time = 40.

References

1. Cornell, John (2002), *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data*, John Wiley & Sons, Inc. New York.

2. Piepel, G. F. (1983), "Defining consistent constraint regions in mixture experiments," *Technometrics*, Vol. 25, pp. 97-101.

3. Snee, R. D. (1979), "Experimental designs for mixture systems with multiple component constraints," *Communications in Statistics, Theory and Methods*, Bol. A8, pp. 303-326.

Reliability DOE for Life Tests

IN THIS CHAPTER

Reliability DOE Analysis **364**

R-DOE Analysis of Lognormally Distributed Data **364**
Maximum Likelihood Estimation for the Lognormal Distribution 366
Hypothesis Tests 368
MLE Estimates 371
Likelihood Ratio Test 373
Fisher Matrix Bounds on Parameters 375

R-DOE Analysis of Data Following the Weibull Distribution **379**
Maximum Likelihood Estimation for the Weibull Distribution 380

R-DOE Analysis of Data Following the Exponential Distribution **382**

Model Diagnostics **382**

Application Examples **383**
Using R-DOE to Determine the Best Factor Settings 383
Using R-DOE and Weibull++ to Estimate B10 Life 385

Single Factor R-DOE Analyses **388**
Example 388

Reliability analysis is commonly thought of as an approach to model failures of existing products. The usual reliability analysis involves characterization of failures of the products using distributions such as exponential, Weibull and lognormal. Based on the fitted distribution, failures are mitigated, or warranty returns are predicted, or maintenance actions are planned. However, by adopting the methodology of Design for Reliability (DFR), reliability analysis can also be used as a powerful tool to design robust products that operate with minimal failures. In DFR, reliability analysis is carried out in conjunction with physics of failure and experiment design techniques. Under this approach, Design of Experiments (DOE) uses life data to "build" reliability into the products, not just quantify the existing reliability. Such an approach, if properly implemented, can result in significant cost savings, especially in terms of fewer warranty

returns or repair and maintenance actions. Although DOE techniques can be used to improve product reliability and also make this reliability robust to noise factors, the discussion in this chapter is focused on reliability improvement. The robust parameter design method discussed in Robust Parameter Design can be used to produce robust and reliable product.

Reliability DOE Analysis

Reliability DOE (R-DOE) analysis is fairly similar to the analysis of other designed experiments except that the response is the life of the product in the respective units (e.g., for an automobile component the units of life may be miles, for a mechanical component this may be cycles, and for a pharmaceutical product this may be months or years). However, two important differences exist that make R-DOE analysis unique. The first is that life data of most products are typically well modeled by either the lognormal, Weibull or exponential distribution, but usually do not follow the normal distribution. Traditional DOE techniques follow the assumption that response values at any treatment level follow the normal distribution and therefore, the error terms, ϵ, can be assumed to be normally and independently distributed. This assumption may not be valid for the response data used in most of the R-DOE analyses. Further, the life data obtained may either be complete or censored, and in this case standard regression techniques applicable to the response data in traditional DOEs can no longer be used.

Design parameters, manufacturing process settings, and use stresses affecting the life of the product can be investigated using R-DOE analysis. In this case, the primary purpose of any R-DOE analysis is to identify which of the inputs affect the life of the product (by investigating if change in the level of any input factors leads to a significant change in the life of the product). For example, once the important stresses affecting the life of the product have been identified, detailed analyses can be carried out using ReliaSoft's Weibull++ software. Weibull++ includes a number of life-stress relationships (LSRs) to model the relation between life and the stress affecting the life of the product.

R-DOE Analysis of Lognormally Distributed Data

Assume that the life, T, for a certain product has been found to be lognormally distributed. The probability density function for the lognormal distribution is:

$$f(T)=\frac{1}{T\sigma'\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\ln(T)-\mu'}{\sigma'}\right)^2}$$

where μ' represents the mean of the natural logarithm of the times-to-failure and σ' represents the standard deviation of the natural logarithms of the times-to-failure [Meeker and Escobar

1998, Wu 2000, ReliaSoft 2007b]. If the analyst wants to investigate a single two level factor that may affect the life, T, then the following model may be used:

$$T_i = \mu_i + \xi_i$$

where:

- T_i represents the times-to-failure at the i th treatment level of the factor
- μ_i represents the mean value of T_i for the i th treatment
- ξ_i is the random error term
- The subscript i represents the treatment level of the factor with $i = 1, 2$ for a two level factor

The model of the equation shown above is analogous to the ANOVA model, $Y_i = \mu_i + \epsilon_i$, used in the One Factor Designs and General Full Factorial Designs chapters for traditional DOE analyses. Note, however, that the random error term, ξ_i , is not normally distributed here because the response, T, is lognormally distributed. It is known that the logarithmic value of a lognormally distributed random variable follows the normal distribution. Therefore, if the logarithmic transformation of T, $ln(T)$, is used in the above equation, then the model will be identical to the ANOVA model, $Y_i = \mu_i + \epsilon_i$, used in the other chapters. Thus, using the logarithmic failure times, the model can be written as:

$$\ln(T_i) = \mu_i' + \epsilon_i$$

where:

- $\ln(T_i)$ represents the logarithmic times-to-failure at the i th treatment
- μ_i' represents the mean of the natural logarithm of the times-to-failure at the i th treatment
- σ' represents the standard deviation of the natural logarithms of the times-to-failure

The random error term, ϵ_i , is normally distributed because the response, $\ln(T_i)$, is normally distributed. Since the model of the equation given above is identical to the ANOVA model used in traditional DOE analysis, regression techniques can be applied here and the R-DOE analysis can be carried out similar to the traditional DOE analyses. Recall from Two Level Factorial Experiments that if the factor(s) affecting the response has only two levels, then the notation of the regression model can be applied to the ANOVA model. Therefore, the model of the above equation can be written using a single indicator variable, x_1 , to represent the two level factor as:

$$\ln(T_i) = \beta_0 + \beta_1 x_{i1} + \epsilon_i$$

where β_0 is the intercept term and β_1 is the effect coefficient for the investigated factor. Setting the two equations above equal to each other returns:

$$\mu_i' = \beta_0 + \beta_1 x_{i1}$$

The natural logarithm of the times-to-failure at any factor level, μ_i', is referred to as the *life characteristic* because it represents a characteristic point of the underlying life distribution. The life characteristic used in the R-DOE analysis will change based on the underlying distribution assumed for the life data. If the analyst wants to investigate the effect of two factors (each at two levels) on the life of the product, then the life characteristic equation can be easily expanded as follows:

$$\mu_i' = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$$

where β_2 is the effect coefficient for the second factor and x_2 is the indicator variable representing the second factor. If the interaction effect is also to be investigated, then the following equation can be used:

$$\mu_i' = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{12} x_{i1} x_{i2}$$

In general the model to investigate a given number of factors can be expressed as:

$$\mu_i' = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{12} x_{i1} x_{i2} + \ldots$$

Based on the model equations mentioned thus far, the analyst can easily conduct an R-DOE analysis for the lognormally distributed life data using standard regression techniques. However this is no longer true once the data also includes censored observations. In the case of censored data, the analysis has to be carried out using maximum likelihood estimation (MLE) techniques.

Maximum Likelihood Estimation for the Lognormal Distribution

The maximum likelihood estimation method can be used to estimate parameters in R-DOE analyses when censored data are present. The likelihood function is calculated for each observed time to failure, t_i, and the parameters of the model are obtained by maximizing the log-likelihood function. The likelihood function for complete data following the lognormal distribution is given as:

$$L_{failures} = \prod_{i=1}^{F_e} f(t_i, \mu_i')$$
$$= \prod_{i=1}^{F_e} \left[\frac{1}{t_i \sigma' \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(t_i)-\mu_i'}{\sigma'}\right)^2} \right]$$
$$= \prod_{i=1}^{F_e} \left[\frac{1}{t_i \sigma' \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(t_i)-(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\ldots)}{\sigma'}\right)^2} \right]$$

where:

- F_e is the total number of observed times-to-failure
- μ_i' is the life characteristic
- t_i is the time of the i th failure

For right censored data the likelihood function [Meeker and Escobar 1998, Wu 2000, ReliaSoft 2007b] is:

$$L_{suspensions} = \prod_{i=1}^{S_e} \left[1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\left(\frac{\ln(t_i)-\mu_i'}{\sigma'}\right)} e^{-\frac{g^2}{2}} dg \right]$$

where:

- S_e is the total number of observed suspensions
- t_i is the time of i th suspension

For interval data the likelihood function [Meeker and Escobar 1998, Wu 2000, ReliaSoft 2007b] is:

$$L_{interval} = \prod_{i=11}^{FI} \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\left(\frac{\ln(t_i^2)-\mu_i'}{\sigma'}\right)} e^{-\frac{g^2}{2}} dg - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\left(\frac{\ln(t_i^1)-\mu_i'}{\sigma'}\right)} e^{-\frac{g^2}{2}} dg \right]$$

where:

- FI is the total number of interval data
- t_i^1 is the beginning time of the i th interval
- t_i^2 is the end time of the i th interval

The complete likelihood function when all types of data (complete, right censored and interval) are present is:

$$L(\sigma', \beta_0, \beta_1 \ldots) = L_{failures} \cdot L_{suspensions} \cdot L_{interval}$$

Then the log-likelihood function is:

$$\Lambda(\sigma', \beta_0, \beta_1 \ldots) = \ln(L)$$

The MLE estimates are obtained by solving for parameters $(\sigma', \beta_0, \beta_1 \ldots)$ so that:

$$\begin{aligned} \frac{\partial \Lambda}{\partial \sigma'} &= 0 \\ \frac{\partial \Lambda}{\partial \beta_0} &= 0 \\ \frac{\partial \Lambda}{\partial \beta_1} &= 0 \\ &\ldots \end{aligned}$$

Once the estimates are obtained, the significance of any parameter, θ_i, can be assessed using the likelihood ratio test.

Hypothesis Tests

Hypothesis testing in R-DOE analyses is carried out using the likelihood ratio test. To test the significance of a factor, the corresponding effect coefficient(s), θ_i, is tested. The following statements are used:

$$\begin{aligned} H_0 &: \quad \theta_i = 0 \\ H_1 &: \quad \theta_i \neq 0 \end{aligned}$$

The statistic used for the test is the likelihood ratio, LR. The likelihood ratio for the parameter θ_i is calculated as follows:

$$LR = -2 \ln \frac{L(\hat{\theta}_{(-i)})}{L(\hat{\theta})}$$

where:

- $\hat{\theta}$ is the vector of all parameter estimates obtained using MLE (i.e., $\hat{\theta} = [\hat{\sigma}' \; \hat{\beta}_0 \; \hat{\beta}_1 \; \ldots \;]'$)
- $\hat{\theta}_{(-i)}$ is the vector of all parameter estimates excluding the estimate of θ_i
- $L(\hat{\theta})$ is the value of the likelihood function when all parameters are included in the model

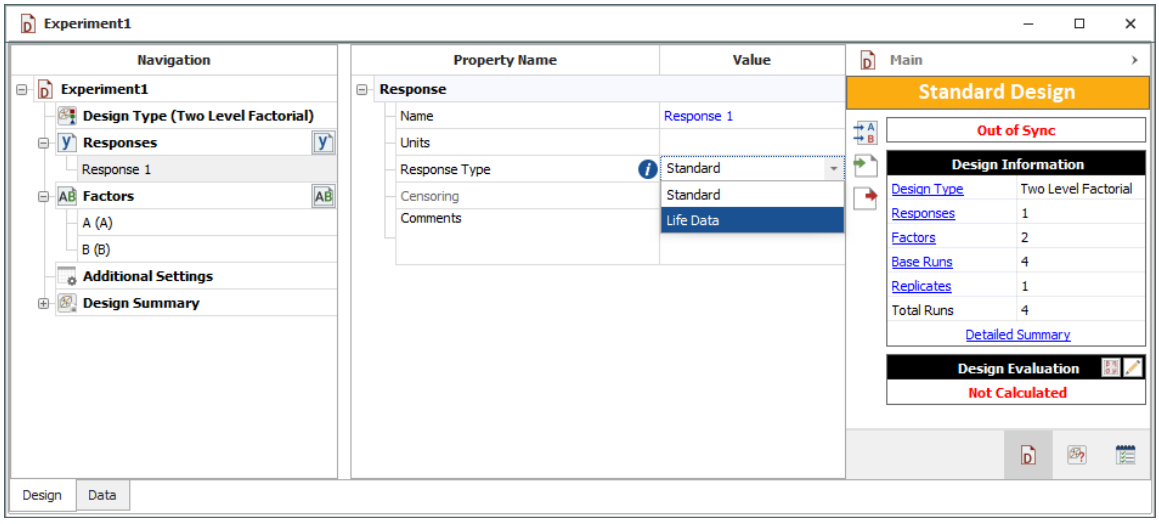
- $L(\hat{\theta}_{(-i)})$ is the value of the likelihood function when all parameters except θ_i are included in the model

If the null hypothesis, H_0, is true then the ratio, $-2\ln L(\hat{\theta}_{(-i)})/L(\hat{\theta})$, follows the chi-squared distribution with one degree of freedom. Therefore, H_0 is rejected at a significance level, α, if LR is greater than the critical value $\chi^2_{1,\alpha}$.

The likelihood ratio test can also be used to test the significance of a number of parameters, r, at the same time. In this case, $L(\hat{\theta}_{(-i)})$ represents the likelihood value when all r parameters to be tested are not included in the model. In other words, $L(\hat{\theta}_{(-i)})$ would represent the likelihood value for the reduced model that does not contain the r parameters under test. Here, the ratio $-2\ln L(\hat{\theta}_{(-i)})/L(\hat{\theta})$ will follow the chi-squared distribution with $k-r$ degrees of freedom if all r parameters are insignificant (with k representing the number of parameters in the full model). Thus, if $LR > \chi^2_{k-r,\alpha}$, the null hypothesis, H_0, is rejected and it can be concluded that at least one of the r parameters is significant.

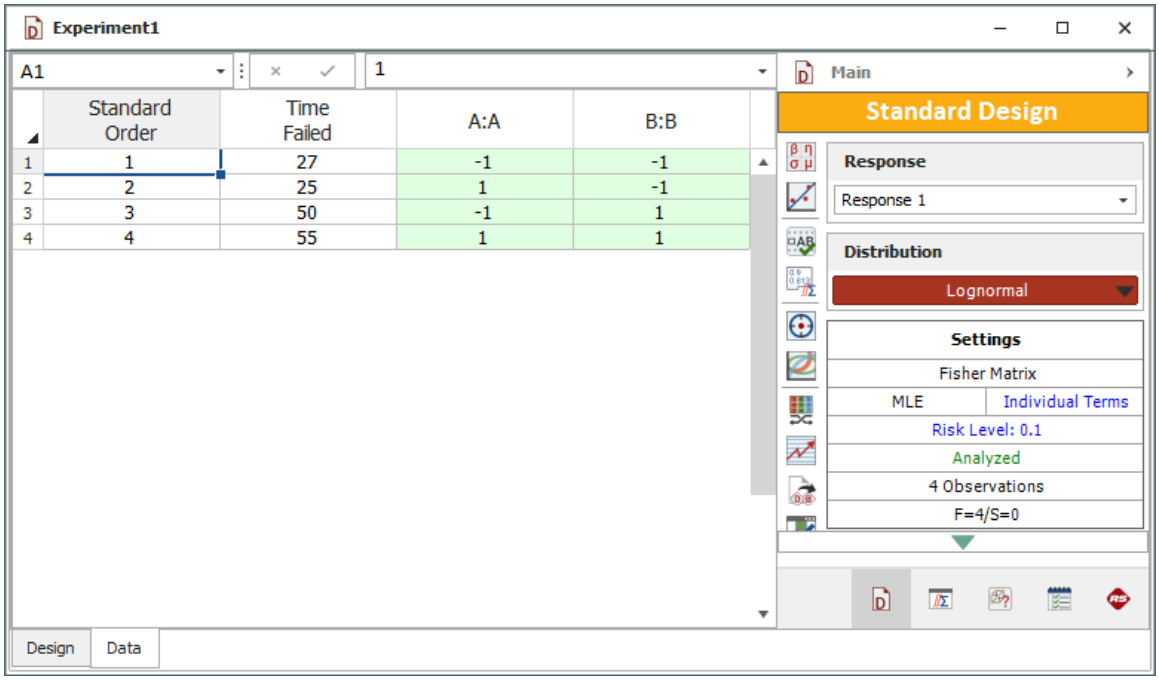
Example

To illustrate the use of MLE in R-DOE analysis, consider the case where the life of a product is thought to be affected by two factors, A and B. The failure of the product has been found to follow the lognormal distribution. The analyst decides to run an R-DOE analysis using a single replicate of the 2^2 design. Previous studies indicate that the interaction between A and B does not affect the life of the product. The design for this experiment can be set up in a Weibull++ DOE folio as shown in the following figure.



The resulting experiment design and the corresponding times-to-failure data obtained are shown next. Note that, although the life data set contains *complete data* and regression techniques are

applicable, calculations are shown using MLE. Weibull++ DOE folios use MLE for all R-DOE analysis calculations.



Because the purpose of the experiment is to study two factors without considering their interaction, the applicable model for the lognormally distributed response data is:

$$\mu_i' = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$$

where μ_i' is the mean of the natural logarithm of the times-to-failure at the i th treatment combination ($i = 1, 2, 3, 4$), β_1 is the effect coefficient for factor A and β_2 is the effect coefficient for factor B. The analysis for this case is carried out in a DOE folio by excluding the interaction AB from the analysis.

The following hypotheses need to be tested in this example:

1) $H_0 : \beta_1 = 0$ $H_1 : \beta_1 \neq 0$

This test investigates the main effect of factor A. The statistic for this test is:

$$LR_A = -2 \ln \frac{L_{-A}}{L}$$

where L represents the value of the likelihood function when all coefficients are included in the model and L_{-A} represents the value of the likelihood function when all coefficients except β_1 are included in the model.

2) $H_0 : \beta_2 = 0 \; H_1 : \beta_2 \neq 0$

This test investigates the main effect of factor B. The statistic for this test is:

$$LR_B = -2 \ln \frac{L_{-B}}{L}$$

where L represents the value of the likelihood function when all coefficients are included in the model and L_{-B} represents the value of the likelihood function when all coefficients except β_2 are included in the model.

To calculate the test statistics, the maximum likelihood estimates of the parameters must be known. The estimates are obtained next.

MLE Estimates

Since the life data for the present experiment are complete and follow the lognormal distribution, the likelihood function can be written as:

$$L = \prod_{i=1}^{4} \left[\frac{1}{t_i \sigma' \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(t_i)-\mu_i'}{\sigma'}\right)^2} \right]$$

Substituting μ_i' from the applicable model for the lognormally distributed response data, the likelihood function is:

$$L = \prod_{i=1}^{4} \left[\frac{1}{t_i \sigma' \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(t_i)-(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2})}{\sigma'}\right)^2} \right]$$

Then the log-likelihood function is:

$$\begin{aligned}
\Lambda(\sigma', \beta_0, \beta_1, \beta_2) &= \ln(L) \\
&= \sum_{i=1}^{4} \ln \left[\frac{1}{t_i \sigma' \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(t_i)-(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2})}{\sigma'}\right)^2} \right] \\
&= \ln \left[\frac{1}{t_1 t_2 t_3 t_4 (\sigma' \sqrt{2\pi})^4} \right] + \\
&\quad \left[-\frac{1}{2} \sum_{i=1}^{4} \left(\frac{\ln(t_i) - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})}{\sigma'} \right)^2 \right] \\
&= -[\ln(t_1 t_2 t_3 t_4) + 4 \ln(\sigma') + 2 \ln(2\pi)] + \\
&\quad \left[-\frac{1}{2} \sum_{i=1}^{4} \left(\frac{\ln(t_i) - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})}{\sigma'} \right)^2 \right]
\end{aligned}$$

To obtain the MLE estimates of the parameters, $\sigma', \beta_0, \beta_1$ and β_2, the log-likelihood function must be differentiated with respect to these parameters:

$$\begin{aligned}
\frac{\partial \Lambda}{\partial \sigma'} &= -\frac{4}{\sigma'} + \frac{1}{(\sigma')^3}\sum_{i=1}^{4}[\ln(t_i) - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})]^2 \\
\frac{\partial \Lambda}{\partial \beta_0} &= \frac{1}{(\sigma')^2}\sum_{i=1}^{4}[\ln(t_i) - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})] \\
\frac{\partial \Lambda}{\partial \beta_1} &= \frac{1}{(\sigma')^2}\sum_{i=1}^{4} x_{i1}[\ln(t_i) - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})] \\
\frac{\partial \Lambda}{\partial \beta_2} &= \frac{1}{(\sigma')^2}\sum_{i=1}^{4} x_{i2}[\ln(t_i) - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})]
\end{aligned}$$

Equating the $\partial\Lambda/\partial\theta_i$ terms to zero returns the required estimates. The coefficients $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ are obtained first as these are required to estimate $\hat{\sigma}'$. Setting $\partial\Lambda/\partial\beta_0 = 0$:

$$\sum_{i=1}^{4}[\ln(t_i) - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})] = 0$$

Substituting the values of t_i, x_{i1} and x_{i2} from the example's experiment design and corresponding data and simplifying:

$$\ln t_1 + \ln t_2 + \ln t_3 + \ln t_4 - 4\beta_0 = 0$$

Thus:

$$\begin{aligned}
\hat{\beta}_0 &= \frac{1}{4}(\ln t_1 + \ln t_2 + \ln t_3 + \ln t_4) \\
&= \frac{1}{4}(3.2958 + 3.2189 + 3.912 + 4.0073) \\
&= 3.6085
\end{aligned}$$

Setting $\partial\Lambda/\partial\beta_1 = 0$:

$$x_{i1}\ln t_1 + x_{i1}\ln t_2 + x_{i1}\ln t_3 + x_{i1}\ln t_4 - 4\beta_1 = 0$$

Thus:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{1}{4}(-\ln t_1 + \ln t_2 - \ln t_3 + \ln t_4) \\
&= \frac{1}{4}(-3.2958 + 3.2189 - 3.912 + 4.0073) \\
&= 0.0046
\end{aligned}$$

Setting $\partial\Lambda/\partial\beta_2 = 0$:

$$x_{i2}\ln t_1 + x_{i2}\ln t_2 + x_{i3}\ln t_3 + x_{i4}\ln t_4 - 4\beta_2 = 0$$

Thus:

$$\begin{aligned}\hat{\beta}_2 &= \frac{1}{4}(-\ln t_1 - \ln t_2 + \ln t_3 + \ln t_4)\\ &= \frac{1}{4}(-3.2958 - 3.2189 + 3.912 + 4.0073)\\ &= 0.3512\end{aligned}$$

Knowing $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$, $\hat{\sigma}'$ can now be obtained. Setting $\partial\Lambda/\partial\sigma' = 0$:

$$-\frac{4}{\sigma'} + \frac{1}{(\sigma')^3}\sum_{i=1}^{4}[\ln(t_i) - (3.6085 + 0.0046x_{i1} + 0.3512x_{i2})]^2 = 0$$

Thus:

$$\begin{aligned}\hat{\sigma}' &= \frac{1}{2}\sqrt{\sum_{i=1}^{4}[\ln(t_i) - (3.6085 + 0.0046x_{i1} + 0.3512x_{i2})]^2}\\ &= 0.043\end{aligned}$$

Once the estimates have been calculated, the likelihood ratio test can be carried out for the two factors.

Likelihood Ratio Test

The likelihood ratio test for factor A is conducted by using the likelihood value corresponding to the full model and the likelihood value when A is not included in the model. The likelihood value corresponding to the full model (in this case $\mu_i' = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$) is:

$$\begin{aligned}L &= \prod_{i=1}^{4}\left[\frac{1}{t_i\hat{\sigma}'\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\ln(t_i)-(\hat{\beta}_0+\hat{\beta}_1x_{i1}+\hat{\beta}_2x_{i2})}{\hat{\sigma}'}\right)^2}\right]\\ &= 0.000537311\end{aligned}$$

The corresponding logarithmic value is $\ln(L) = \ln(0.000537311) = -7.529$. The likelihood value for the reduced model that does not contain factor A (in this case $\mu_i' = \beta_0 + \beta_2 x_{i2}$) is:

$$\begin{aligned}L_{\sim A} &= \prod_{i=1}^{4}\left[\frac{1}{t_i\hat{\sigma}'\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\ln(t_i)-(\hat{\beta}_0+\hat{\beta}_2x_{i2})}{\hat{\sigma}'}\right)^2}\right]\\ &= 0.000525337\end{aligned}$$

The corresponding logarithmic value is $\ln(L_{\sim A}) = \ln(0.000525337) = -7.552$. Therefore, the likelihood ratio to test the significance of factor A is:

$$\begin{aligned} LR_A &= -2\ln\frac{L_{\sim A}}{L} \\ &= -2\ln\frac{0.000525337}{0.000537311} \\ &= 0.0451 \end{aligned}$$

The p value corresponding to LR_A is:

$$\begin{aligned} p\,value &= 1 - P(\chi_1^2 < LR_A) \\ &= 1 - 0.1682 \\ &= 0.8318 \end{aligned}$$

Assuming that the desired significance level for the present experiment is 0.1, since $p\,value > 0.1$, $H_0 : \beta_1 = 0$ cannot be rejected and it can be concluded that factor A does not affect the life of the product.

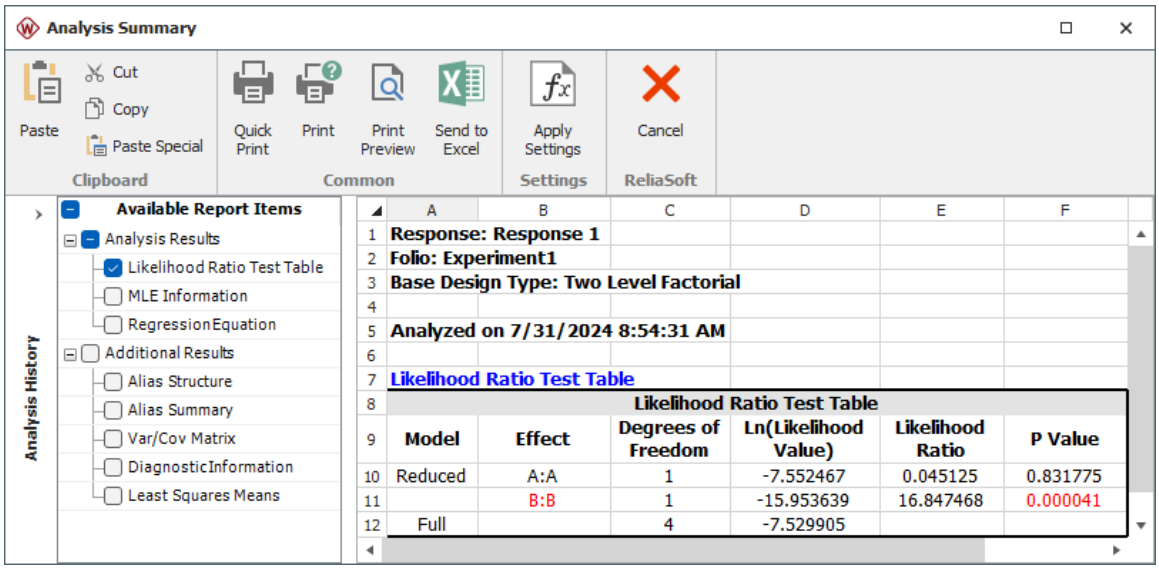
The likelihood ratio to test factor B can be calculated in a similar way as shown next:

$$\begin{aligned} LR_B &= -2\ln\frac{L_{\sim B}}{L} \\ &= -2\ln\frac{1.17995E-07}{0.000537311} \\ &= 16.8475 \end{aligned}$$

The p value corresponding to LR_B is:

$$\begin{aligned} p\,value &= 1 - P(\chi_1^2 < LR_B) \\ &= 1 - 0.99996 \\ &= 0.00004 \end{aligned}$$

Since $p\,value < 0.1$, $H_0 : \beta_2 = 0$ is rejected and it is concluded that factor B affects the life of the product. The previous calculation results are displayed as the Likelihood Ratio Test Table in the results obtained from the DOE folio as shown next.



Fisher Matrix Bounds on Parameters

In general, the MLE estimates of the parameters are asymptotically normal. This means that for large sample sizes the distribution of the estimates from the same population would be very close to the normal distribution[Meeker and Escobar 1998]. If $\hat{\theta}$ is the MLE estimate of any parameter, θ, then the $(1-\alpha)$% two-sided confidence bounds on the parameter are:

$$\hat{\theta} - z_{\alpha/2} \cdot \sqrt{Var(\hat{\theta})} < \theta < \hat{\theta} + z_{\alpha/2} \cdot \sqrt{Var(\hat{\theta})}$$

where $Var(\hat{\theta})$ represents the variance of $\hat{\theta}$ and $z_{\alpha/2}$ is the critical value corresponding to a significance level of $\alpha/2$ on the standard normal distribution. The variance of the parameter, $Var(\hat{\theta})$, is obtained using the Fisher information matrix. For k parameters, the Fisher information matrix is obtained from the log-likelihood function Λ as follows:

$$F = \begin{bmatrix} -\frac{\partial^2 \Lambda}{\partial \theta_1^2} & -\frac{\partial^2 \Lambda}{\partial \theta_1 \partial \theta_2} & \cdots & -\frac{\partial^2 \Lambda}{\partial \theta_1 \partial \theta_k} \\ -\frac{\partial^2 \Lambda}{\partial \theta_1 \partial \theta_2} & -\frac{\partial^2 \Lambda}{\partial \theta_2^2} & \cdots & -\frac{\partial^2 \Lambda}{\partial \theta_2 \partial \theta_k} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ -\frac{\partial^2 \Lambda}{\partial \theta_1 \partial \theta_k} & . & \cdots & -\frac{\partial^2 \Lambda}{\partial \theta_k^2} \end{bmatrix}$$

The variance-covariance matrix is obtained by inverting the Fisher matrix F:

$$\begin{bmatrix} Var(\hat{\theta}_1) & Cov(\hat{\theta}_1,\hat{\theta}_2) & \cdots & \\ Cov(\hat{\theta}_1,\hat{\theta}_2) & Var(\hat{\theta}_2) & \cdots & \\ \cdot & \cdot & \cdots & \\ \cdot & \cdot & \cdots & \\ Cov(\hat{\theta}_1,\hat{\theta}_k) & \cdot & \cdots & Var(\hat{\theta}_k) \end{bmatrix} = \begin{bmatrix} -\frac{\partial^2\Lambda}{\partial\theta_1^2} & -\frac{\partial^2\Lambda}{\partial\theta_1\partial\theta_2} & \cdots & \\ -\frac{\partial^2\Lambda}{\partial\theta_1\partial\theta_2} & -\frac{\partial^2\Lambda}{\partial\theta_2^2} & \cdots & \\ \cdot & \cdot & \cdots & \\ \cdot & \cdot & \cdots & \\ -\frac{\partial^2\Lambda}{\partial\theta_1\partial\theta_k} & \cdot & \cdots & -\frac{\partial^2\Lambda}{\partial\theta_k^2} \end{bmatrix}^{-1}$$

Once the variance-covariance matrix is known the variance of any parameter can be obtained from the diagonal elements of the matrix. Note that if a parameter, θ, can take only positive values, it is assumed that the $\ln(\hat{\theta})$ follows the normal distribution [Meeker and Escobar 1998]. The bounds on the parameter in this case are:

$$CI\ on\ \ln(\hat{\theta}) = \ln(\hat{\theta}) \pm z_{\alpha/2}\sqrt{Var(\ln(\hat{\theta}))}$$

Using $Var[f(\hat{\theta})] = (\partial f/\partial\theta)^2 \cdot Var(\hat{\theta})$ we get $Var(\ln(\hat{\theta})) = (1/\hat{\theta})^2 Var(\hat{\theta})$. Substituting this value we have:

$$\begin{aligned} CI\ on\ \ln(\hat{\theta}) &= \ln(\hat{\theta}) \pm z_{\alpha/2}\sqrt{(1/\hat{\theta})^2 Var(\hat{\theta})} \\ &= \ln(\hat{\theta}) \pm (z_{\alpha/2}/\hat{\theta})\sqrt{Var(\hat{\theta})} \\ or\ CI\ on\ \hat{\theta} &= \exp[\ln(\hat{\theta}) \pm (z_{\alpha/2}/\hat{\theta})\sqrt{Var(\hat{\theta})}] \\ &= \hat{\theta} \cdot \exp[\pm(z_{\alpha/2}/\hat{\theta})\sqrt{Var(\hat{\theta})}] \end{aligned}$$

Knowing $Var(\hat{\theta})$ from the variance-covariance matrix, the confidence bounds on $\hat{\theta}$ can then be determined.

Continuing with the present example, the confidence bounds on the MLE estimates of the parameters β_0, β_1, β_2 and σ' can now be obtained. The Fisher information matrix for the example is:

$$F = \begin{bmatrix} -\frac{\partial^2 \Lambda}{\partial \beta_0^2} & -\frac{\partial^2 \Lambda}{\partial \beta_0 \partial \beta_1} & -\frac{\partial^2 \Lambda}{\partial \beta_0 \partial \beta_2} & -\frac{\partial^2 \Lambda}{\partial \beta_0 \partial \sigma'} \\ & -\frac{\partial^2 \Lambda}{\partial \beta_1^2} & -\frac{\partial^2 \Lambda}{\partial \beta_1 \partial \beta_2} & -\frac{\partial^2 \Lambda}{\partial \beta_1 \partial \sigma'} \\ & & -\frac{\partial^2 \Lambda}{\partial \beta_2^2} & -\frac{\partial^2 \Lambda}{\partial \beta_2 \partial \sigma'} \\ sym. & & & -\frac{\partial^2 \Lambda}{\partial \sigma'^2} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{4}{\sigma'^2} & \frac{1}{\sigma'^2}\sum_{i=1}^{4} x_{i1} & \frac{1}{\sigma'^2}\sum_{i=1}^{4} x_{i2} & \frac{2}{\sigma'^3}[\sum_{i=1}^{4} (\ln t_i - \mu_i')] \\ & \frac{1}{\sigma'^2}\sum_{i=1}^{4} x_{i1}^2 & \frac{1}{\sigma'^2}\sum_{i=1}^{4} x_{i1}x_{i2} & \frac{2}{\sigma'^3}[\sum_{i=1}^{4} x_{i1} \cdot (\ln t_i - \mu_i')] \\ & & \frac{1}{\sigma'^2}\sum_{i=1}^{4} x_{i2}^2 & \frac{2}{\sigma'^3}[\sum_{i=1}^{4} x_{i2} \cdot (\ln t_i - \mu_i')] \\ sym. & & & \frac{4}{\sigma'^2} + \frac{(-3)}{\sigma'^4}\sum_{i=1}^{4} (\ln t_i - \mu_i')^2] \end{bmatrix}$$

$$= \begin{bmatrix} 2165.6741 & 0 & 0 & -1.1195E-11 \\ & 2165.6741 & 0 & -1.1195E-11 \\ & & 2165.6741 & -3.358E-11 \\ sym. & & & 4330.8227 \end{bmatrix}$$

The variance-covariance matrix can be obtained by taking the inverse of the Fisher matrix F:

$$\begin{bmatrix} Var(\hat{\beta}_0) & Cov(\hat{\beta}_0, \hat{\beta}_1) & Cov(\hat{\beta}_0, \hat{\beta}_2) & Cov(\hat{\beta}_0, \hat{\sigma}') \\ & Var(\hat{\beta}_1) & Cov(\hat{\beta}_1, \hat{\beta}_2) & Cov(\hat{\beta}_0, \hat{\sigma}') \\ & & Var(\hat{\beta}_2) & Cov(\hat{\beta}_0, \hat{\sigma}') \\ sym. & & & Var(\hat{\sigma}') \end{bmatrix} = F^{-1}$$

Inverting F returns the following matrix:

$$F^{-1} = \begin{bmatrix} 4.617E-4 & 0 & 0 & 0 \\ & 4.617E-4 & 0 & 0 \\ & & 4.617E-4 & 0 \\ sym. & & & 2.309E-4 \end{bmatrix}$$

Therefore, the variance of the parameter estimates are:

$$\begin{aligned} Var(\hat{\beta}_0) &= 4.617E-4 \\ Var(\hat{\beta}_1) &= 4.617E-4 \\ Var(\hat{\beta}_2) &= 4.617E-4 \\ Var(\hat{\sigma}') &= 2.309E-4 \end{aligned}$$

Knowing the variance, the confidence bounds on the parameters can be calculated. For example, the 90% bounds ($\alpha = 0.1$) on $\hat{\beta}_2$ can be calculated as shown next:

$$\begin{aligned} CI &= \hat{\beta}_2 \pm z_{\alpha/2} \cdot \sqrt{Var(\hat{\beta}_2)} \\ &= \hat{\beta}_2 \pm z_{0.05} \cdot \sqrt{Var(\hat{\beta}_2)} \\ &= 0.3512 \pm 1.645 \cdot \sqrt{4.617E-4} \\ &= 0.3512 \pm 0.0354 \\ &= 0.3158 \; and \; 0.3866 \end{aligned}$$

The 90% bounds on σ' are (considering that σ' can only take positive values):

$$\begin{aligned} CI &= \hat{\sigma}' \cdot \exp[\pm(z_{0.05}/\hat{\sigma}')\sqrt{Var(\hat{\sigma}')}] \\ &= 0.043 \cdot \exp[\pm(1.645/0.043)\sqrt{2.309E-4}] \\ &= 0.024 \; and \; 0.077 \end{aligned}$$

The standard error for the parameters can be obtained by taking the positive square root of the variance. For example, the standard error for $\hat{\beta}_1$ is:

$$\begin{aligned} se(\hat{\beta}_1) &= \sqrt{Var(\hat{\beta}_1)} \\ &= \sqrt{4.617E-4} \\ &= 0.0215 \end{aligned}$$

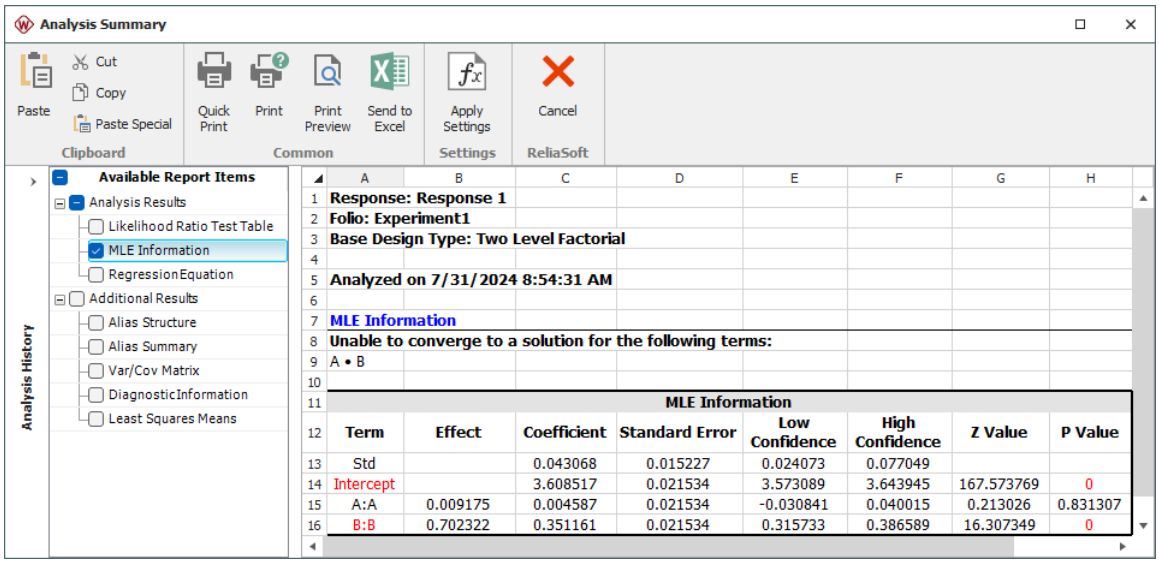
The z statistic for $\hat{\beta}_1$ is:

$$\begin{aligned} z_0 &= \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} \\ &= \frac{0.0046}{0.0215} \\ &= 0.21 \end{aligned}$$

The p value corresponding to this statistic based on the standard normal distribution is:

$$\begin{aligned} p\,value &= 2 \cdot (1 - P(Z \leq |z_0|) \\ &= 2 \cdot (1 - 0.58435) \\ &= 0.8313 \end{aligned}$$

The previous calculation results are displayed as MLE Information in the results obtained from the DOE folio as shown next.



In the figure, the Effect corresponding to each factor is simply twice the MLE estimate of the coefficient for that factor. Generally, the p value corresponding to any coefficient in the MLE Information table should match the value obtained from the likelihood ratio test (displayed in the Likelihood Ratio Test table of the results). If the sample size is not large enough, as in the case of the present example, a difference may be seen in the two values. In such cases, the p value from the likelihood ratio test should be given preference. For the present example, the p value of 0.8318 for $\hat{\beta}_1$, obtained from the likelihood ratio test, would be preferred to the p value of 0.8313 displayed under MLE information. For details see [Meeker and Escobar 1998].

R-DOE Analysis of Data Following the Weibull Distribution

The probability density function for the 2-parameter Weibull distribution is:

$$f(T) = \frac{\beta}{\eta}\left(\frac{T}{\eta}\right)^{\beta-1} \exp\left[-\left(\frac{T}{\eta}\right)^{\beta}\right]$$

where η is the scale parameter of the Weibull distribution and β is the shape parameter [Meeker and Escobar 1998, ReliaSoft 2007b]. To distinguish the Weibull shape parameter from the effect coefficients, the shape parameter is represented as $Beta$ instead of β in the remaining chapter. For data following the 2-parameter Weibull distribution, the life characteristic used in R-DOE analysis is the scale parameter, η [ReliaSoft 2007a, Wu 2000]. Since η represents life data that cannot take negative values, a logarithmic transformation is applied to it. The resulting model used in the R-DOE analysis for a two factor experiment with each factor at two levels can be written as follows:

$$\ln(\eta_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{12} x_{i1} x_{i2}$$

where:

- η_i is the value of the scale parameter at the i th treatment combination of the two factors
- x_1 is the indicator variable representing the level of the first factor
- x_2 is the indicator variable representing the level of the second factor
- β_0 is the intercept term
- β_1 and β_2 are the effect coefficients for the two factors
- and β_{12} is the effect coefficient for the interaction of the two factors

The model can be easily expanded to include other factors and their interactions. Note that when any data follows the Weibull distribution, the logarithmic transformation of the data follows the extreme-value distribution, whose probability density function is given as follows:

$$f(\ln(T)) = \frac{1}{\sigma''} \exp\left[\frac{\ln(T) - \mu''}{\sigma''} - \exp\left(\frac{\ln(T) - \mu''}{\sigma''}\right)\right]$$

where the T s follows the Weibull distribution, μ'' is the location parameter of the extreme-value distribution and σ'' is the scale parameter of the extreme-value distribution. The two equations given above show that for R-DOE analysis of life data that follows the Weibull distribution, the random error terms, ϵ_i , will follow the extreme-value distribution (and not the normal distribution). Hence, regression techniques are not applicable even if the data is complete. Therefore, maximum likelihood estimation has to be used.

Maximum Likelihood Estimation for the Weibull Distribution

The likelihood function for complete data in R-DOE analysis of Weibull distributed life data is:

$$L_{failures} = \prod_{i=1}^{F_e} f(t_i, \eta_i)$$
$$= \prod_{i=1}^{F_e} \left[\frac{Beta}{\eta_i}\left(\frac{t_i}{\eta_i}\right)^{Beta-1} \exp\left[-\left(\frac{t_i}{\eta_i}\right)^{Beta}\right]\right]$$

where:

- F_e is the total number of observed times-to-failure
- η_i is the life characteristic at the i th treatment
- t_i is the time of the i th failure

For right censored data, the likelihood function is:

$$L_{suspensions} = \prod_{i=1}^{S_e} \left[\exp\left[-\left(\frac{t_i}{\eta_i} \right)^{Beta} \right] \right]$$

where:

- S_e is the total number of observed suspensions
- t_i is the time of i th suspension

For interval data, the likelihood function is:

$$L_{interval} = \prod_{i=1}^{FI} \left[\exp\left[-\left(\frac{t_i^2}{\eta_i} \right)^{Beta} \right] - \exp\left[-\left(\frac{t_i^1}{\eta_i} \right)^{Beta} \right] \right]$$

where:

- FI is the total number of interval data
- t_i^1 is the beginning time of the i th interval
- t_i^2 is the end time of the i th interval

In each of the likelihood functions, η_i is substituted based on the equation for $ln(\eta_i)$ as:

$$\eta_i = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots)$$

The complete likelihood function when all types of data (complete, right and left censored) are present is:

$$L(Beta, \beta_0, \beta_1 \ldots) = L_{failures} \cdot L_{suspensions} \cdot L_{interval}$$

Then the log-likelihood function is:

$$\Lambda(Beta, \beta_0, \beta_1 \ldots) = \ln(L)$$

The MLE estimates are obtained by solving for parameters $(Beta, \beta_0, \beta_1 \ldots)$ so that:

$$\begin{aligned} \frac{\partial \Lambda}{\partial Beta} &= 0 \\ \frac{\partial \Lambda}{\partial \beta_0} &= 0 \\ \frac{\partial \Lambda}{\partial \beta_1} &= 0 \\ &\ldots \end{aligned}$$

Once the estimates are obtained, the significance of any parameter, θ_i, can be assessed using the likelihood ratio test. Other results can also be obtained as discussed in Maximum Likelihood Estimation for the Lognormal Distribution and Fisher Matrix Bounds on Parameters.

R-DOE Analysis of Data Following the Exponential Distribution

The exponential distribution is a special case of the Weibull distribution when the shape parameter $Beta$ is equal to 1. Substituting $Beta = 1$ in the probability density function for the 2-parameter Weibull distribution gives:

$$\begin{aligned} f(T) = & \quad \frac{1}{\eta}\exp\left(-\frac{T}{\eta}\right) \\ = & \quad \lambda\exp(-\lambda T) \end{aligned}$$

where $1/\eta$ of the *pdf* has been replaced by λ. Parameter λ is called the failure rate.[ReliaSoft 2007a] Hence, R-DOE analysis for exponentially distributed data can be carried out by substituting $Beta = 1$ and replacing $1/\eta$ by λ in the Weibull distribution.

Model Diagnostics

Residual plots can be used to check if the model obtained, based on the MLE estimates, is a good fit to the data. Weibull++ DOE folios use standardized residuals for R-DOE analyses. If the data follows the lognormal distribution, then standardized residuals are calculated using the following equation:

$$\begin{aligned} \hat{e}_i = & \quad \frac{\ln(t_i) - \hat{\mu}_i}{\hat{\sigma}'} \\ = & \quad \frac{\ln(t_i) - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \ldots)}{\hat{\sigma}'} \end{aligned}$$

For the probability plot, the standardized residuals are displayed on a normal probability plot. This is because under the assumed model for the lognormal distribution, the standardized residuals should follow a normal distribution with a mean of 0 and a standard deviation of 1.

For data that follows the Weibull distribution, the standardized residuals are calculated as shown next:

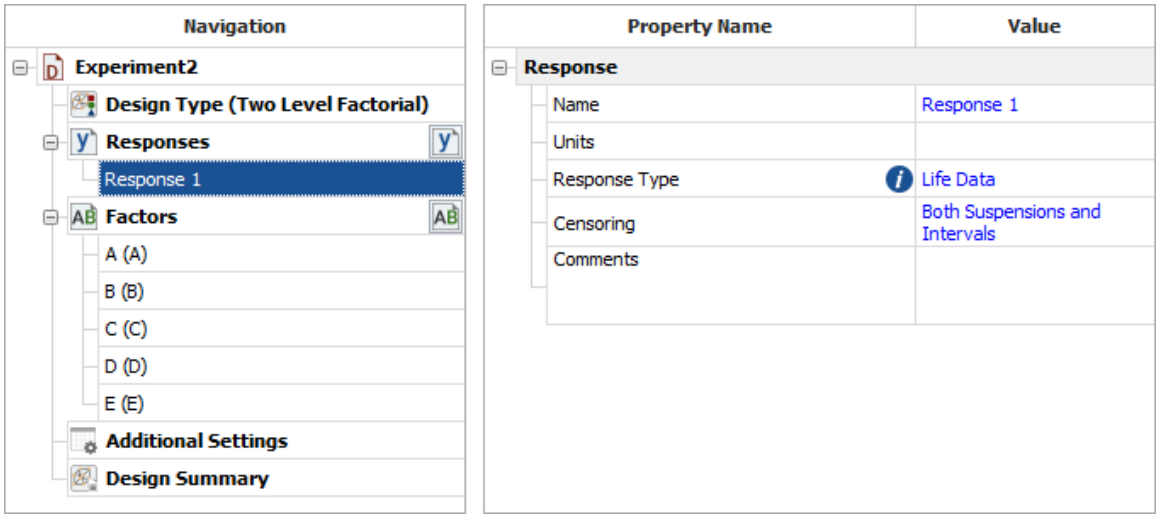
$$\begin{aligned} \hat{e}_i = & \quad \hat{Beta}[\ln(t_i) - \ln(\hat{\eta}_i)] \\ = & \quad \hat{Beta}[\ln(t_i) - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \ldots)] \end{aligned}$$

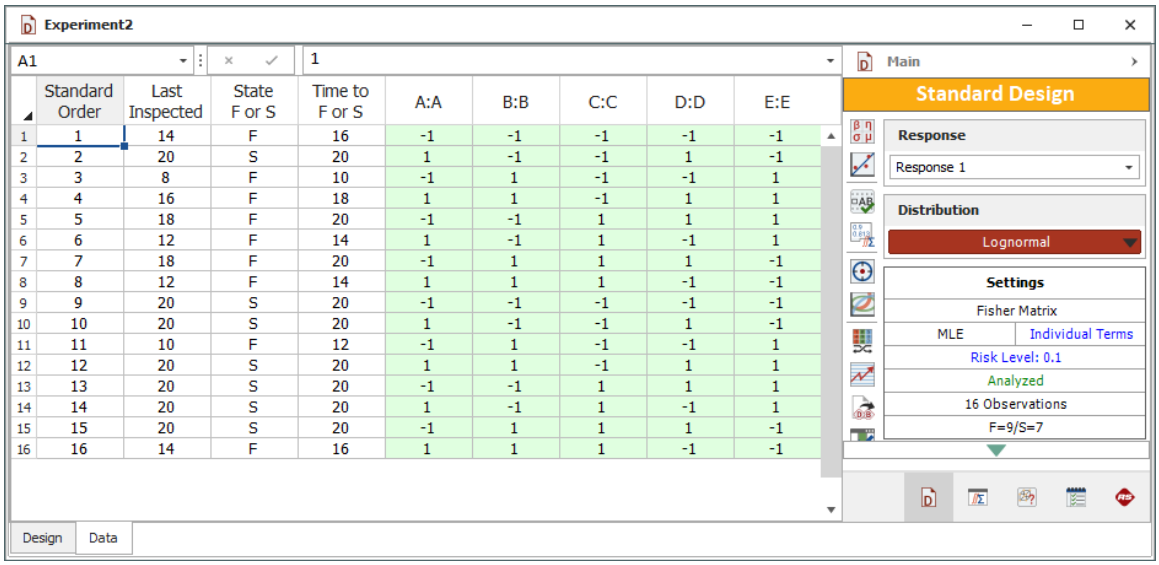
The probability plot, in this case, is used to check if the residuals follow the extreme-value distribution with a mean of 0. Note that in all residual plots, when an observation, t_i, is censored the corresponding residual is also censored.

Application Examples

Using R-DOE to Determine the Best Factor Settings

This example illustrates the use of R-DOE analysis to design reliability into a product by determining the optimal factor settings. An experiment was carried out to investigate the effect of five factors (each at two levels) on the reliability of fluorescent lights (Taguchi, 1987, p. 930). The factors, A through E, were studied using a 2^{5-2} design (with the defining relations $D = -AC$ and $E = -BC$) under the assumption that all interaction effects, except $AB\,(= DE)$, can be assumed to be inactive. For each treatment, two lights were tested (two replicates) with the readings taken every two days. The experiment was run for 20 days and, if a light had not failed by the 20th day, it was assumed to be a suspension. The experimental design and the corresponding failure times are shown next.

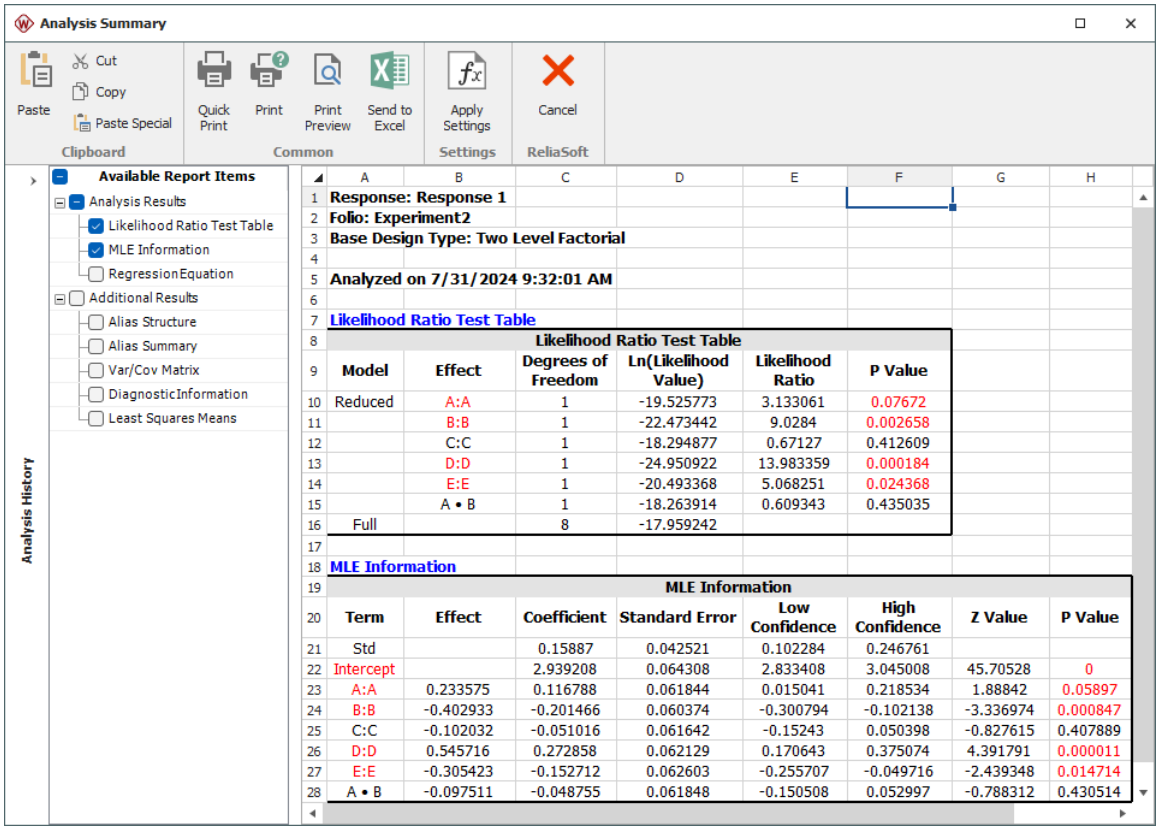




	Standard Order	Last Inspected	State F or S	Time to F or S	A:A	B:B	C:C	D:D	E:E
1	1	14	F	16	-1	-1	-1	-1	-1
2	2	20	S	20	1	-1	-1	1	-1
3	3	8	F	10	-1	1	-1	-1	1
4	4	16	F	18	1	1	-1	1	1
5	5	18	F	20	-1	-1	1	1	1
6	6	12	F	14	1	-1	1	-1	1
7	7	18	F	20	-1	1	1	1	-1
8	8	12	F	14	1	1	1	-1	-1
9	9	20	S	20	-1	-1	-1	-1	-1
10	10	20	S	20	1	-1	-1	1	-1
11	11	10	F	12	-1	1	-1	-1	1
12	12	20	S	20	1	1	-1	1	1
13	13	20	S	20	-1	-1	1	1	1
14	14	20	S	20	1	-1	1	-1	1
15	15	20	S	20	-1	1	1	1	-1
16	16	14	F	16	1	1	1	-1	-1

The short duration of the experiment and failure times were probably because the lights were tested under conditions which resulted in stress higher than normal conditions. The failure of the lights was assumed to follow the lognormal distribution.

The analysis results from the Weibull++ DOE folio for this experiment are shown next.



Response: Response 1
Folio: Experiment2
Base Design Type: Two Level Factorial

Analyzed on 7/31/2024 9:32:01 AM

Likelihood Ratio Test Table

Model	Effect	Degrees of Freedom	Ln(Likelihood Value)	Likelihood Ratio	P Value
Reduced	A:A	1	-19.525773	3.133061	0.07672
	B:B	1	-22.473442	9.0284	0.002658
	C:C	1	-18.294877	0.67127	0.412609
	D:D	1	-24.950922	13.983359	0.000184
	E:E	1	-20.493368	5.068251	0.024368
	A • B	1	-18.263914	0.609343	0.435035
Full		8	-17.959242		

MLE Information

Term	Effect	Coefficient	Standard Error	Low Confidence	High Confidence	Z Value	P Value
Std		0.15887	0.042521	0.102284	0.246761		
Intercept		2.939208	0.064308	2.833408	3.045008	45.70528	0
A:A	0.233575	0.116788	0.061844	0.015041	0.218534	1.88842	0.05897
B:B	-0.402933	-0.201466	0.060374	-0.300794	-0.102138	-3.336974	0.000847
C:C	-0.102032	-0.051016	0.061642	-0.15243	0.050398	-0.827615	0.407889
D:D	0.545716	0.272858	0.062129	0.170643	0.375074	4.391791	0.000011
E:E	-0.305423	-0.152712	0.062603	-0.255707	-0.049716	-2.439348	0.014714
A • B	-0.097511	-0.048755	0.061848	-0.150508	0.052997	-0.788312	0.430514

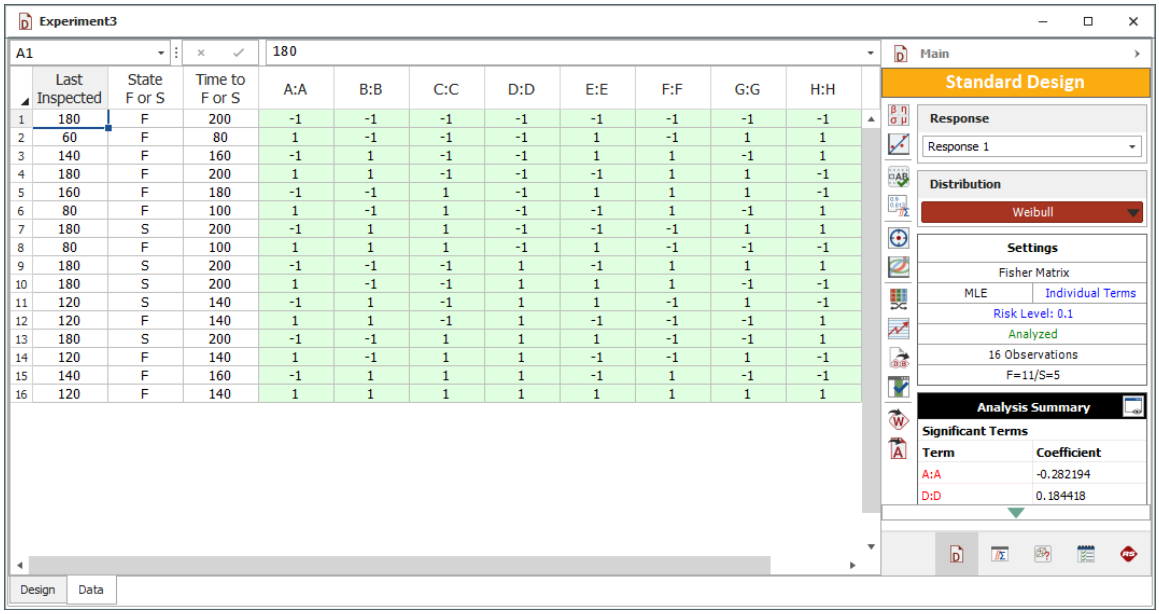
The results are obtained by selecting the main effects of the five factors and the interaction AB. The results show that factors A, B, D and E are active at a significance level of 0.1. The MLE estimates of the effect coefficients corresponding to these factors are 0.1168, -0.2015, 0.2729 and -0.1527, respectively. Based on these coefficients, the best settings for these effects to improve the reliability of the fluorescent lights (by maximizing the response, which in this case is the failure time) are:

- Factor A should be set at the higher level of 1 since its coefficient is positive
- Factor B should be set at the lower level of -1 since its coefficient is negative
- Factor D should be set at the higher level of 1 since its coefficient is positive
- Factor E should be set at the lower level of -1 since its coefficient is negative

Note that, since actual factor levels are not disclosed (presumably for proprietary reasons), predictions beyond the test conditions cannot be carried out in this case.

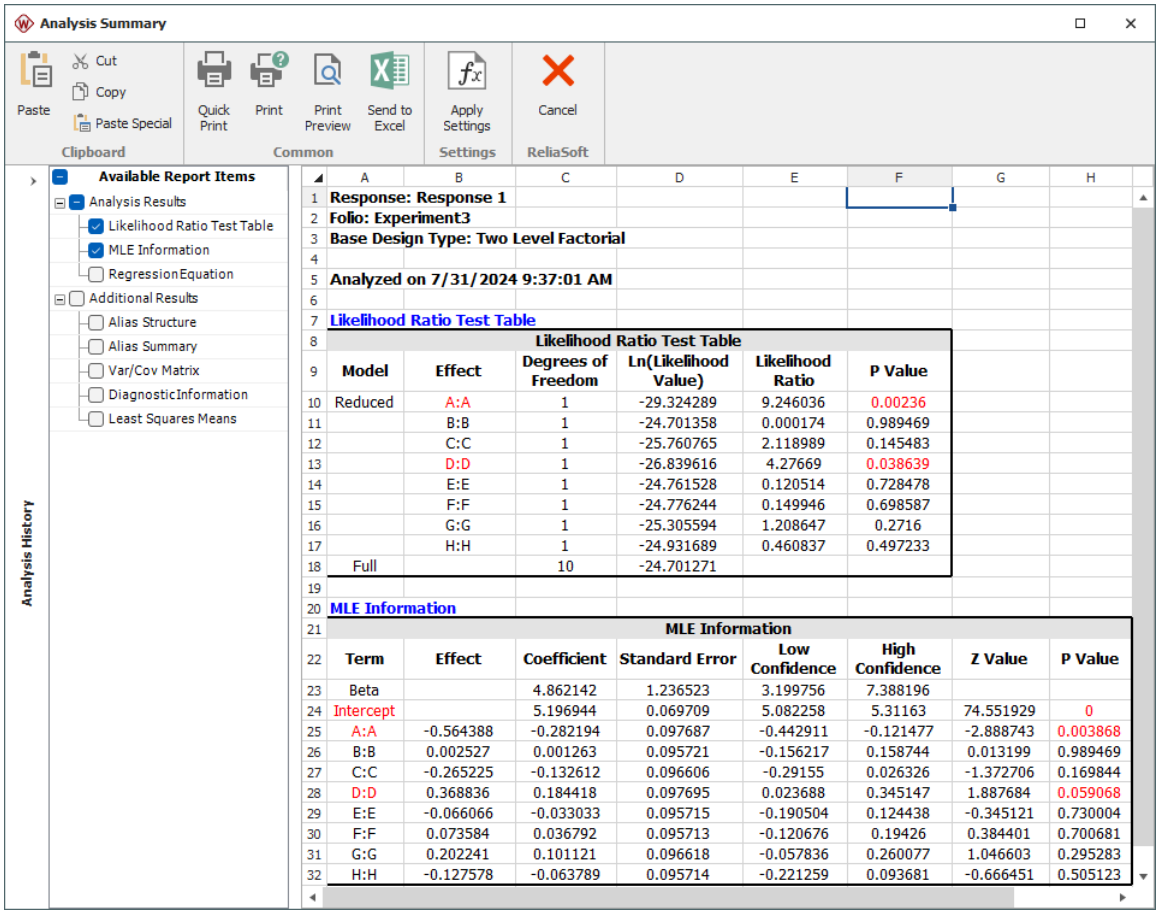
Using R-DOE and Weibull++ to Estimate B10 Life

Consider a product whose reliability is thought to be affected by eight potential factors: A (temperature), B (humidity), C (load), D (fan-speed), E (voltage), F (material), G (vibration) and H (current). Assuming that all interaction effects are absent, a 2^{8-4} design is used to investigate the eight factors at two levels. The generators used to obtain the design are $E = ABC$, $F = BCD$, $G = ACD$ and $H = ABD$. The design and the corresponding life data obtained are shown next.



Readings for the experiment are taken every 20 hours and the test is terminated at 200 hours. The life of the product is assumed to follow the Weibull distribution.

The results from Weibull++ for this experiment are shown next.



	A	B	C	D	E	F	G	H
1	**Response: Response 1**							
2	**Folio: Experiment3**							
3	**Base Design Type: Two Level Factorial**							
4								
5	**Analyzed on 7/31/2024 9:37:01 AM**							
6								
7	**Likelihood Ratio Test Table**							

Likelihood Ratio Test Table

Model	Effect	Degrees of Freedom	Ln(Likelihood Value)	Likelihood Ratio	P Value
Reduced	A:A	1	-29.324289	9.246036	0.00236
	B:B	1	-24.701358	0.000174	0.989469
	C:C	1	-25.760765	2.118989	0.145483
	D:D	1	-26.839616	4.27669	0.038639
	E:E	1	-24.761528	0.120514	0.728478
	F:F	1	-24.776244	0.149946	0.698587
	G:G	1	-25.305594	1.208647	0.2716
	H:H	1	-24.931689	0.460837	0.497233
Full		10	-24.701271		

MLE Information

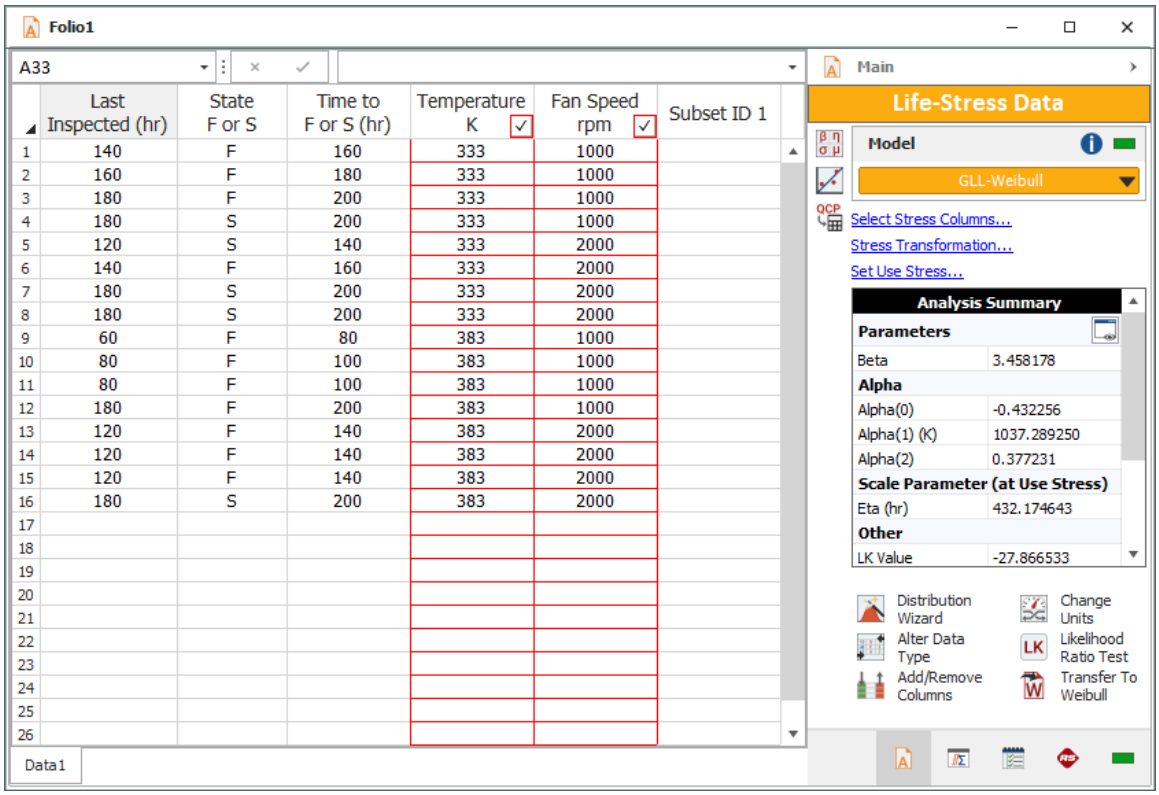
Term	Effect	Coefficient	Standard Error	Low Confidence	High Confidence	Z Value	P Value
Beta		4.862142	1.236523	3.199756	7.388196		
Intercept		5.196944	0.069709	5.082258	5.31163	74.551929	0
A:A	-0.564388	-0.282194	0.097687	-0.442911	-0.121477	-2.888743	0.003868
B:B	0.002527	0.001263	0.095721	-0.156217	0.158744	0.013199	0.989469
C:C	-0.265225	-0.132612	0.096606	-0.29155	0.026326	-1.372706	0.169844
D:D	0.368836	0.184418	0.097695	0.023688	0.345147	1.887684	0.059068
E:E	-0.066066	-0.033033	0.095715	-0.190504	0.124438	-0.345121	0.730004
F:F	0.073584	0.036792	0.095713	-0.120676	0.19426	0.384401	0.700681
G:G	0.202241	0.101121	0.096618	-0.057836	0.260077	1.046603	0.295283
H:H	-0.127578	-0.063789	0.095714	-0.221259	0.093681	-0.666451	0.505123

The results show that only factors A and D are active at a significance level of 0.1.

Assume that, in terms of the actual units, the -1 level of factor A corresponds to a temperature of 333 K and the $+1$ level corresponds to a temperature of 383 K. Similarly, assume that the two levels of factor D are 1000 rpm and 2000 rpm respectively. From the MLE estimates of the effect coefficients it can be noted that to improve reliability (by maximizing the response) factors A and D should be set as follows:

- Factor A should be set at the lower level of 333 K since its coefficient is negative
- Factor D should be set at the higher level of 2000 rpm since its coefficient is positive

Now assume that the use conditions for the product for the significant factors, A and D, are a temperature of 298 K and a fan-speed of 3000 rpm, respectively. The analysis can be taken a step further to obtain an estimate of the reliability of the product at the use conditions using ReliaSoft's Weibull++ software. The data is entered into Weibull++ as shown next.



Weibull++ allows for modeling of the nature of relationship between life and stress. It is assumed that the relation between life of the product and temperature follows the Arrhenius relation while the relation between life and fan-speed follows the inverse power law relation.[ReliaSoft 2007a] Using these relations, Webull++ fits the following model for the data:

$$\eta = \exp[-0.4322 + 1037.2886\frac{1}{\text{Temp}} + 0.3772 \cdot \ln(\text{Fan-Speed})]$$

Based on this model, the B10 life of the product at the use conditions is obtained as shown next. The Weibull reliability equation is:

$$R(t) = \exp\left[-(\frac{t}{\eta})^{Beta}\right]$$

Substituting the value of η from the Weibull++ model and the value of $Beta(= 3.4582)$ as obtained from Weibull++, the reliability equation becomes:

$$R(t) = \exp - \left[\frac{t}{\exp[-0.4322 + 1037.2886\frac{1}{\text{Temp}} + 0.3772 \cdot \ln(\text{Fan-Speed})]}\right]^{3.4582}$$

Finally, substituting the use conditions (Temp $= 298\ K$, Fan-Speed $= 3000\ rpm$) and the desired reliability value of 90%, the B10 life is obtained:

$$0.90 = \exp - \left[\frac{t}{\exp[-0.4322 + 1037.2886 \frac{1}{298} + 0.3772 \cdot \ln(3000)]} \right]^{3.4582}$$

$$t = 225.4482$$

Therefore, at the use conditions, the B10 life of the product is 225 hours. This result and other reliability metrics can be directly obtained from Weibull++.

Single Factor R-DOE Analyses

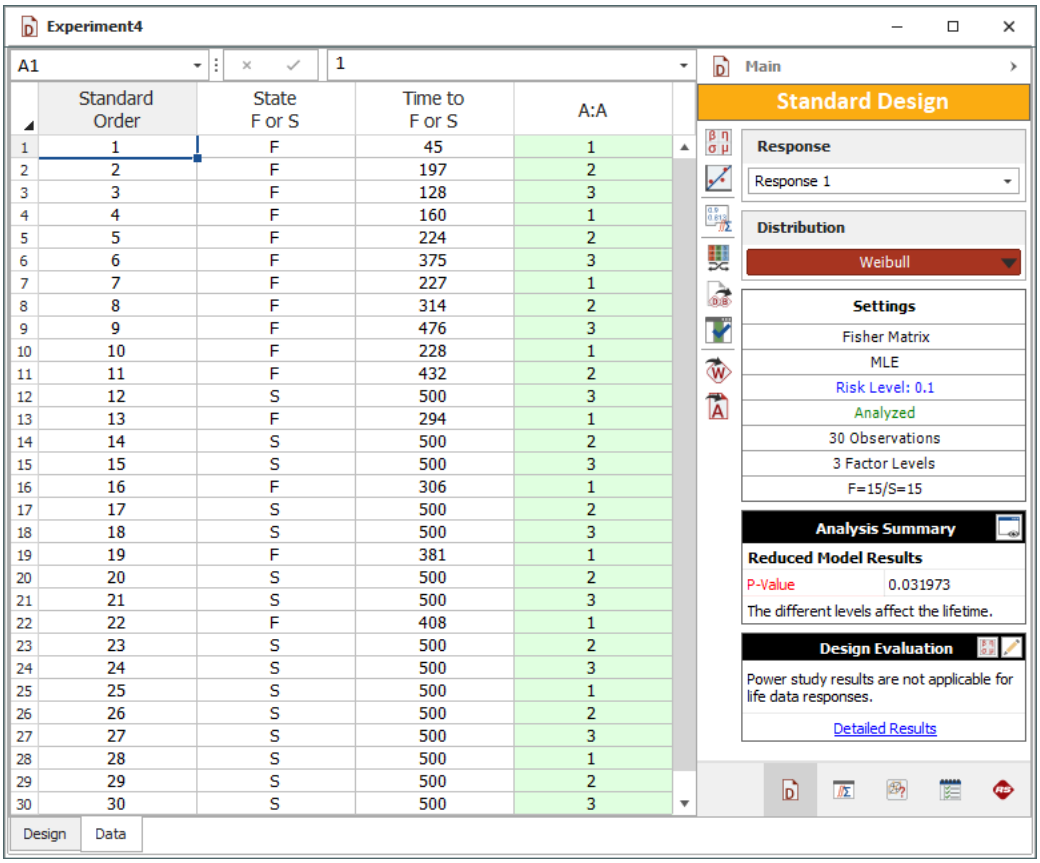
Webibull++ DOE folios also allow for the analysis of single factor R-DOE experiments. This analysis is similar to the analysis of single factor designed experiments mentioned in One Factor Designs. In single factor R-DOE analysis, the focus is on discovering whether change in the level of a factor affects reliability and how each of the factor levels are different from the other levels. The analysis models and calculations are similar to multi-factor R-DOE analysis.

Example

To illustrate single factor R-DOE analysis, consider the data in the table shown next, where 10 life data readings for a product are taken at each of the three levels of a certain factor, A.

Factor *A*					
Level 1		Level 2		Level 3	
F/S	Time	F/S	Time	F/S	Time
F	45	F	197	F	128
F	160	F	224	F	375
F	227	F	314	F	476
F	228	F	432	S	500
F	294	S	500	S	500
F	306	S	500	S	500
F	381	S	500	S	500
F	408	S	500	S	500
S	500	S	500	S	500
S	500	S	500	S	500

Factor A could be a stress that is thought to affect life or three different designs of the same product, or it could be the same product manufactured by three different machines or operators, etc. The goal of the experiment is to see if there is a change in life due to change in the levels of the factor. The design for this experiment is shown next.



The life of the product is assumed to follow the Weibull distribution. Therefore, the life characteristic to be used in the R-DOE analysis is the scale parameter, η. Since factor A has three levels, the model for the life characteristic, η, is:

$$\ln(\eta_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$$

where β_0 is the intercept, β_1 is the effect coefficient for the first level of the factor (β_1 is represented as "A[1]" in Weibull++ DOE folios) and β_2 is the effect coefficient for the second level of the factor (β_2 is represented as "A[2]" in Weibull++ DOE folios). Two indicator variables, x_1 and x_2, are the used to represent the three levels of factor A such that:

$$\begin{aligned} x_1 &= \quad 1,\ x_2 = 0 \text{ Level 1 Effect} \\ x_1 &= \quad 0,\ x_2 = 1 \text{ Level 2 Effect} \\ x_1 &= -1,\ x_2 = -1 \text{ Level 3 Effect} \end{aligned}$$

The following hypothesis test needs to be carried out in this example:

$$\begin{aligned} H_0 &: \quad \theta_i = 0 \\ H_1 &: \quad \theta_i \neq 0 \end{aligned}$$

where $\theta_i = [\beta_1, \beta_2]'$. The statistic for this test is:

$$LR = -2\ln\frac{L(\hat{\theta}_{(-i)})}{L(\hat{\theta})}$$

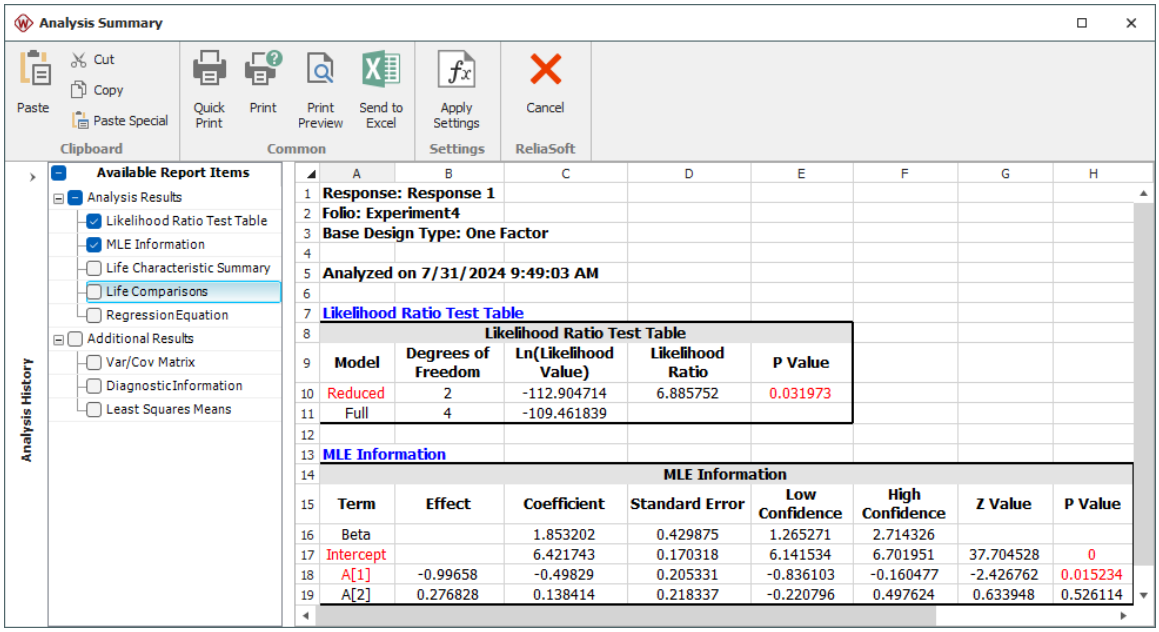
where $L(\hat{\theta})$ is the value of the likelihood function corresponding to the full model, and $L(\hat{\theta}_{(-i)})$ is the likelihood value for the reduced model. To calculate the statistic for this test, the MLE estimates of the parameters must be obtained.

MLE Estimates

Following the procedure used in the analysis of multi-factor R-DOE experiments, MLE estimates of the parameters are obtained by differentiating the log-likelihood function Λ:

$$\Lambda = \sum_{i\epsilon 1}^{F_e} \ln\left[\frac{Beta}{\eta_i}\left(\frac{t_i}{\eta_i}\right)^{Beta-1}\exp\left[-\left(\frac{t_i}{\eta_i}\right)^{Beta}\right]\right] + \sum_{i\epsilon 1}^{S_e}\left[-\left(\frac{t_i}{\eta_i}\right)^{Beta}\right] + \sum_{i\epsilon 1}^{FI}\ln\left[\exp\left[-\left(\frac{t_i^2}{\eta_i}\right)^{Beta}\right] - \exp\left[-\left(\frac{t_i^1}{\eta_i}\right)^{Beta}\right]\right]$$

Substituting η_i from the model for the life characteristic and setting the partial derivatives $\partial\Lambda/\partial\theta_i$ to zero, the parameter estimates are obtained as $\hat{Beta} = 1.8532$, $\hat{\beta}_0 = 6.4217$, $\hat{\beta}_1 = -0.4983$ and $\hat{\beta}_2 = 0.1384$. These parameters are shown in the MLE Information table in the analysis results, shown next.



Response: Response 1
Folio: Experiment4
Base Design Type: One Factor

Analyzed on 7/31/2024 9:49:03 AM

Likelihood Ratio Test Table

Model	Degrees of Freedom	Ln(Likelihood Value)	Likelihood Ratio	P Value
Reduced	2	-112.904714	6.885752	0.031973
Full	4	-109.461839		

MLE Information

Term	Effect	Coefficient	Standard Error	Low Confidence	High Confidence	Z Value	P Value
Beta		1.853202	0.429875	1.265271	2.714326		
Intercept		6.421743	0.170318	6.141534	6.701951	37.704528	0
A[1]	-0.99658	-0.49829	0.205331	-0.836103	-0.160477	-2.426762	0.015234
A[2]	0.276828	0.138414	0.218337	-0.220796	0.497624	0.633948	0.526114

Likelihood Ratio Test

Knowing the MLE estimates, the likelihood ratio test for the significance of factor A can be carried out. The likelihood value for the full model, $L(\hat{\theta})$, is the value of the likelihood function corresponding to the model $\ln(\eta_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$:

$$\begin{aligned} L(\hat{\theta}) = & \quad L(\hat{B}eta, \hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) \\ = & \quad 9.2E-50 \end{aligned}$$

The likelihood value for the reduced model, $L(\hat{\theta}_{(-i)})$, is the value of the likelihood function corresponding to the model $\ln(\eta_i) = \beta_0$:

$$\begin{aligned} L(\hat{\theta}_{(-i)}) = & \quad L(\hat{B}eta, \hat{\beta}_0) \\ = & \quad 2.9E-48 \end{aligned}$$

Then the likelihood ratio is:

$$\begin{aligned} LR = & \quad -2\ln\frac{L(\hat{\theta}_{(-i)})}{L(\hat{\theta})} \\ = & \quad 6.8858 \end{aligned}$$

If the null hypothesis, H_0, is true then the likelihood ratio will follow the chi-squared distribution. The number of degrees of freedom for this distribution is equal to the difference in the number of parameters between the full and the reduced model. In this case, this difference is 2. The p value corresponding to the likelihood ratio on the chi-squared distribution with two degrees of freedom is:

$$\begin{aligned} p\ value = & \quad 1 - P(\chi_2^2 < LR) \\ = & \quad 1 - 0.968 \\ = & \quad 0.032 \end{aligned}$$

Assuming that the desired significance is 0.1, since $p\ value < 0.1$, $H_0 : \theta_i = 0$ is rejected, it is concluded that, at a significance of 0.1, at least one of the parameters, β_1 or β_2, is non-zero. Therefore, factor A affects the life of the product. This result is shown in the Likelihood Ratio Test table in the analysis results.

Additional results for single factor R-DOE analysis obtained from the DOE folio include information on the life characteristic and comparison of life characteristics at different levels of the factor.

Life Characteristic Summary Results

Results in the Life Characteristic Summary table, include information about the life characteristic corresponding to each treatment level of the factor. If $\ln(\eta_i)$ is represented as $E(y_i)$, then the model for the life characteristic can be written as:

$$E(y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$$

The respective equations for all three treatment levels for a single replicate of the experiment can be expressed in matrix notation as:

$$E(y) = X\beta$$

where:

$$E(y) = \begin{bmatrix} E(y_1) \\ E(y_2) \\ E(y_3) \end{bmatrix} \quad X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

Knowing $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$, the predicted value of the life characteristic at any level can be obtained. For example, for the second level:

$$\begin{aligned} E(\hat{y}_2) &= \hat{\beta}_0 + \hat{\beta}_2 \\ or\ \ln(\hat{\eta}_2) &= \hat{\beta}_0 + \hat{\beta}_2 \\ &= 6.421743 + 0.138414 \\ &= 6.560157 \end{aligned}$$

Thus:

$$\begin{aligned} \hat{\eta}_2 &= \exp(6.560157) \\ &= 706.3828 \end{aligned}$$

The variance for the predicted values of life characteristic can be calculated using the following equation:

$$Var(y) = XVar(\hat{\beta})X'$$

where $Var(\hat{\beta})$ is the variance-covariance matrix for $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$. Substituting the required values:

$$Var(\hat{y}) = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} 0.0291 & -0.0174 & 0.0031 \\ -0.0174 & 0.0423 & -0.0154 \\ 0.0031 & -0.0154 & 0.0478 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{bmatrix}'$$

$$= \begin{bmatrix} 0.0364 & -0.0006 & -0.0009 \\ -0.0006 & 0.0829 & 0.0141 \\ -0.0009 & 0.0141 & 0.1167 \end{bmatrix}$$

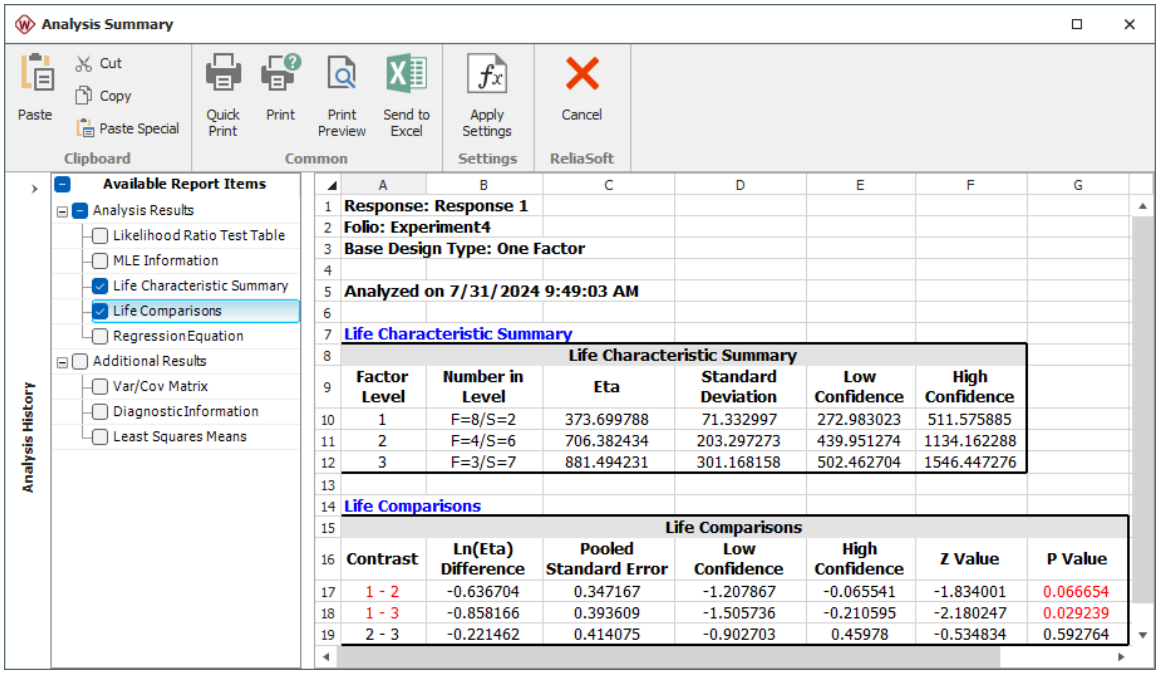
From the previous matrix, $Var(\hat{y}_2) = 0.0829$. Therefore, the 90% confidence interval ($\alpha = 0.1$) on $\hat{y}_2$ is:

$$\begin{aligned} CI\ on\ \hat{y}_2 &= E(\hat{y}_2) \pm z_{\alpha/2}\sqrt{Var(\hat{y}_2)} \\ &= E(\hat{y}_2) \pm z_{0.05}\sqrt{Var(\hat{y}_2)} \\ &= 6.560157 \pm 1.645\sqrt{0.0829} \\ &= 6.0867\ and\ 7.0336 \end{aligned}$$

Since $\hat{y}_2 = \ln(\hat{\eta}_2)$, the 90% confidence interval on $\hat{\eta}_2$ is:

$$\begin{aligned} CI\ on\ \hat{\eta}_2 &= \exp(6.0867)\ and\ \exp(7.0336) \\ &= 439.9\ and\ 1134.1 \end{aligned}$$

Results for other levels can be calculated in a similar manner and are shown next.



Response: Response 1
Folio: Experiment4
Base Design Type: One Factor

Analyzed on 7/31/2024 9:49:03 AM

Life Characteristic Summary

Life Characteristic Summary

Factor Level	Number in Level	Eta	Standard Deviation	Low Confidence	High Confidence
1	F=8/S=2	373.699788	71.332997	272.983023	511.575885
2	F=4/S=6	706.382434	203.297273	439.951274	1134.162288
3	F=3/S=7	881.494231	301.168158	502.462704	1546.447276

Life Comparisons

Life Comparisons

Contrast	Ln(Eta) Difference	Pooled Standard Error	Low Confidence	High Confidence	Z Value	P Value
1 - 2	-0.636704	0.347167	-1.207867	-0.065541	-1.834001	0.066654
1 - 3	-0.858166	0.393609	-1.505736	-0.210595	-2.180247	0.029239
2 - 3	-0.221462	0.414075	-0.902703	0.45978	-0.534834	0.592764

Life Comparisons Results

Results under Life Comparisons include information on how life is different at a level in comparison to any other level of the factor. For example, the difference between the predicted values of life at levels 1 and 2 is (in terms of the logarithmic transformation):

$$\begin{aligned} E(\hat{y}_1) - E(\hat{y}_2) &= 5.923453 - 6.560157 \\ &= -0.6367 \end{aligned}$$

The pooled standard error for this difference can be obtained as:

$$\begin{aligned} \textit{Pooled Std. Error} &= \sqrt{Var(\hat{y}_1 - \hat{y}_2)} \\ &= \sqrt{Var(\hat{y}_1) + Var(\hat{y}_2)} \\ &= \sqrt{0.0366 + 0.0831} \\ &= 0.3454 \end{aligned}$$

If the covariance between $\hat{y}_1$ and $\hat{y}_2$ is taken into account, then the pooled standard error is:

$$\begin{aligned} \textit{Pooled Std. Error} &= \sqrt{Var(\hat{y}_1 - \hat{y}_2)} \\ &= \sqrt{Var(\hat{y}_1) + Var(\hat{y}_2) - 2 \cdot Cov(\hat{y}_1, \hat{y}_2)} \\ &= \sqrt{0.0364 + 0.0829 - 2 \cdot (-0.0006)} \\ &= 0.3471 \end{aligned}$$

This is the value displayed by the Weibull++ DOE folio. Knowing the pooled standard error the confidence interval on the difference can be calculated. The 90% confidence interval on the difference in (logarithmic) life between levels 1 and 2 of factor A is:

$$\begin{aligned} &= \{E(\hat{y}_1) - E(\hat{y}_2)\} \pm z_{\alpha/2} \cdot \textit{Pooled Std. Error} \\ &= \{E(\hat{y}_1) - E(\hat{y}_2)\} \pm z_{0.05} \cdot \textit{Pooled Std. Error} \\ &= -0.6367 \pm 1.645 \cdot 0.3471 \\ &= -1.208 \textit{ and } -0.066 \end{aligned}$$

Since the confidence interval does not include zero it can be concluded that the two levels are significantly different at $\alpha = 0.1$. Another way to test for the significance of the difference in levels is to observe the p value. The z statistic corresponding to this difference is:

$$\begin{aligned} z_{(1-2)} &= \frac{E(\hat{y}_1) - E(\hat{y}_2)}{\textit{Pooled Std. Error}} \\ &= \frac{-0.6367}{0.3471} \\ &= -1.834 \end{aligned}$$

The p value corresponding to this statistic, based on the standard normal distribution, is:

$$\begin{aligned} p\ value = & \quad 2 \cdot (1 - P(Z < |-1.8335|) \\ = & \quad 2 \cdot (0.03336) \\ = & \quad 0.0667 \end{aligned}$$

Since $p\ value < \alpha$, it can be concluded that the levels are significantly different at $\alpha = 0.1$. The results for other levels can be calculated in a similar manner and are shown in the analysis results.

Measurement System Analysis

IN THIS CHAPTER

Introduction **397**

Gage Linearity and Bias Study **398**
Example of Linearity and Bias Study 398

Gage Repeatability and Reproducibility Study **405**
Gage R&R Study for Crossed Experiments 407
Example: Gage R&R Study for Crossed Experiment 409
Gage R&R Study for Nested Experiments 413
X-bar and R Charts in Gage R&R 418

Gage Agreement Study **426**
Accuracy Agreement Study 426
Precision Agreement Study 434

General Guidelines on Measurement System Analysis **437**

An important aspect of conducting design of experiments (DOE) is having a capable measurement system for collecting data. A measurement system is a collection of procedures, gages and operators that are used to obtain measurements. Measurement systems analysis (MSA) is used to evaluate the capacity of a measurement system from the following statistical properties: bias, linearity, stability, repeatability and reproducibility. Some of the applications of MSA are:

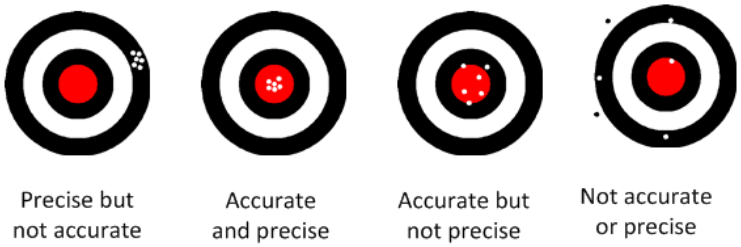
- Provide a criterion to accept new measuring equipment.
- Provide a comparison of one measuring device against another (gage agreement study).
- Provide a comparison for measuring equipment before and after repair.
- Evaluate the variance of components in a product/process.

Introduction

MSA studies the error within a measurement system. Measurement system error can be classified into three categories: accuracy, precision, and stability.

- **Accuracy** describes the difference between the measurement and the actual value of the part that is measured. It includes:
  - *Bias*: a measure of the difference between the true value and the observed value of a part. If the "true" value is unknown, it can be calculated by averaging several measurements with the most accurate measuring equipment available.
  - *Linearity*: a measure of how the size of the part affects the bias of a measurement system. It is the difference in the observed bias values through the expected range of measurement.
- **Precision** describes the variation you see when you measure the same part repeatedly with the same device. It includes the following two types of variation:
  - *Repeatability*: variation due to the measuring device. It is the variation observed when the same operator measures the same part repeatedly with the same device.
  - *Reproducibility*: variation due to the operators and the interaction between operator and part. It is the variation of the bias observed when different operators measure the same parts using the same device.
- **Stability**: a measure of how the accuracy and precision of the system perform over time.

The following picture illustrates accuracy and precision.



In this chapter, we will discuss how to conduct linearity and bias study and gage R&R (repeatability and reproducibility) analysis. The stability of a measurement system can be studied using statistical process control (SPC) charts.

Gage Linearity and Bias Study

Gage *linearity* tells you how accurate your measurements are across the expected range of the measurements. It answers the question, "Does my gage have the same accuracy for all sizes of objects being measured?"

Gage *bias* examines the difference between the observed average measurement and a reference value. It answers the question, "On average, how large is the difference between the values my gage yields and the reference values?"

Let's use an example to show what linearity is.

Example of Linearity and Bias Study

If a baby is 8.5 lbs and the reading of a scale is 8.9 lbs, then the bias is 0.4 lb. If an adult is 85 lbs and the reading from the same scale is 85.4 lbs, then the bias is still 0.4 lb. This scale does not seem to have a linearity issue. However, if the reading for the adult were 89 lbs, the bias would seem to increase as the weight increases. Thus, you might suspect that the scale has a linearity issue.

The following data set shows measurements from a gage linearity and bias study.

Part	Reference	Reading	Part	Reference	Reading
1	2	1.95	3	6	6.04
1	2	2.10	3	6	6.25
1	2	2.00	3	6	6.21
1	2	1.92	3	6	6.16
1	2	1.97	3	6	6.06
1	2	1.94	3	6	6.03
1	2	2.02	4	8	8.40
1	2	2.05	4	8	8.35
1	2	1.95	4	8	8.15
1	2	2.04	4	8	8.10
2	4	4.09	4	8	8.18

2	4	4.16	5	10	10.49
2	4	4.16	5	10	10.28
2	4	4.10	5	10	10.42
2	4	4.06	5	10	10.29
2	4	4.11	5	10	10.14
2	4	4.02	5	10	10.07

The first column is the part ID. The second column is the "true" value of each part, called *reference* or *master*. In a linearity study, the selected reference should cover the minimal and maximal value of the produced parts. The *Reading* column is the observed value from a measurement device. Each part was measured multiple times, and some parts have the same reference value.

The following linear regression equation is used for gage linearity and bias study:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

where:

- Y is the bias.
- X is the reference value.
- β_0 and β_1 are the coefficients.
- ε is error following a normal distribution
- $N(0, \sigma^2)$

First, we need to calculate the bias for each observation in the above table. Bias is the difference between Reading and Reference. The bias values are:

Part	Reference	Reading	Bias	Part	Reference	Reading	Bias
1	2	1.95	-0.05	3	6	6.04	0.04
1	2	2.1	0.1	3	6	6.25	0.25
1	2	2	0	3	6	6.21	0.21

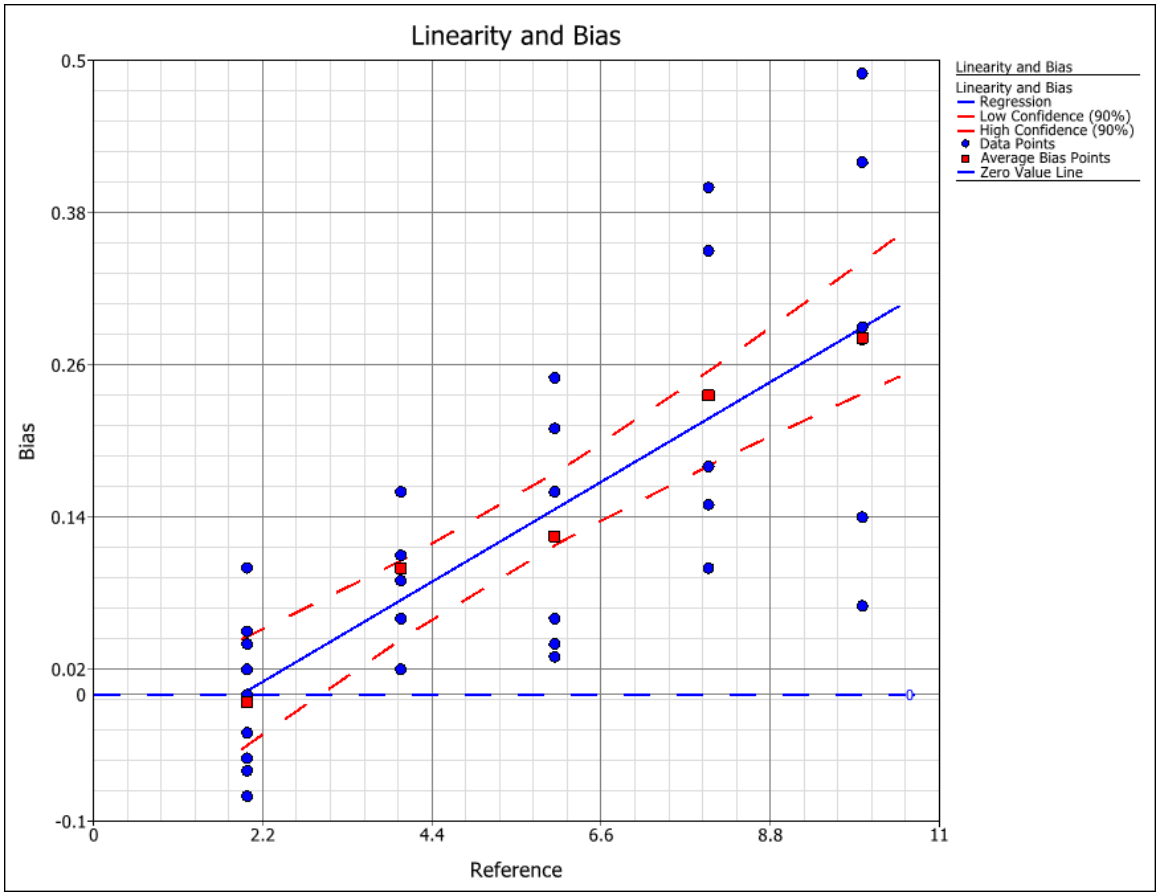
1	2	1.92	-0.08	3	6	6.16	0.16
1	2	1.97	-0.03	3	6	6.06	0.06
1	2	1.94	-0.06	3	6	6.03	0.03
1	2	2.02	0.02	4	8	8.4	0.4
1	2	2.05	0.05	4	8	8.35	0.35
1	2	1.95	-0.05	4	8	8.15	0.15
1	2	2.04	0.04	4	8	8.1	0.1
2	4	4.09	0.09	4	8	8.18	0.18
2	4	4.16	0.16	5	10	10.49	0.49
2	4	4.16	0.16	5	10	10.28	0.28
2	4	4.1	0.1	5	10	10.42	0.42
2	4	4.06	0.06	5	10	10.29	0.29
2	4	4.11	0.11	5	10	10.14	0.14
2	4	4.02	0.02	5	10	10.07	0.07

Results for Linearity Study

Using the Reference column as X and the Bias column as Y in the linear regression, we get the following results:

Source of Variation	Degrees of Freedom	Sum of Squares [Partial]	Mean Squares [Partial]	F Ratio	P Value
Reference	1	0.3748	0.3748	40.4619	3.83E-07
Residual	32	0.2964	0.0093		
Lack of Fit	3	0.01	0.0033	0.3388	0.7974
Pure Error	29	0.2864	0.0099		
Total	33	0.6712			

The calculated R-sq is 55.84% and R-sq(adj) is 54.46%. These values are not very high due to the large variation among the bias values. However, the p value of the lack of fit shows that the linear equation fits the data very well, and the following plot also shows there is a linear relation between reference and bias.



The estimated coefficients are:

Regression Information						
Term	**Coefficient**	**Standard Error**	**Low CI**	**High CI**	**T Value**	**P Value**
Intercept	-0.0685	0.0347	-0.1272	-0.0098	-1.9773	0.0567
Reference	0.0358	0.0056	0.0263	0.0454	6.361	3.83E-07

The linearity is defined by:

$$\text{linearity} = |\beta_1| \times \text{process variation}$$
$$= |\beta_1| \times 6 \times \text{process standard deviation}$$

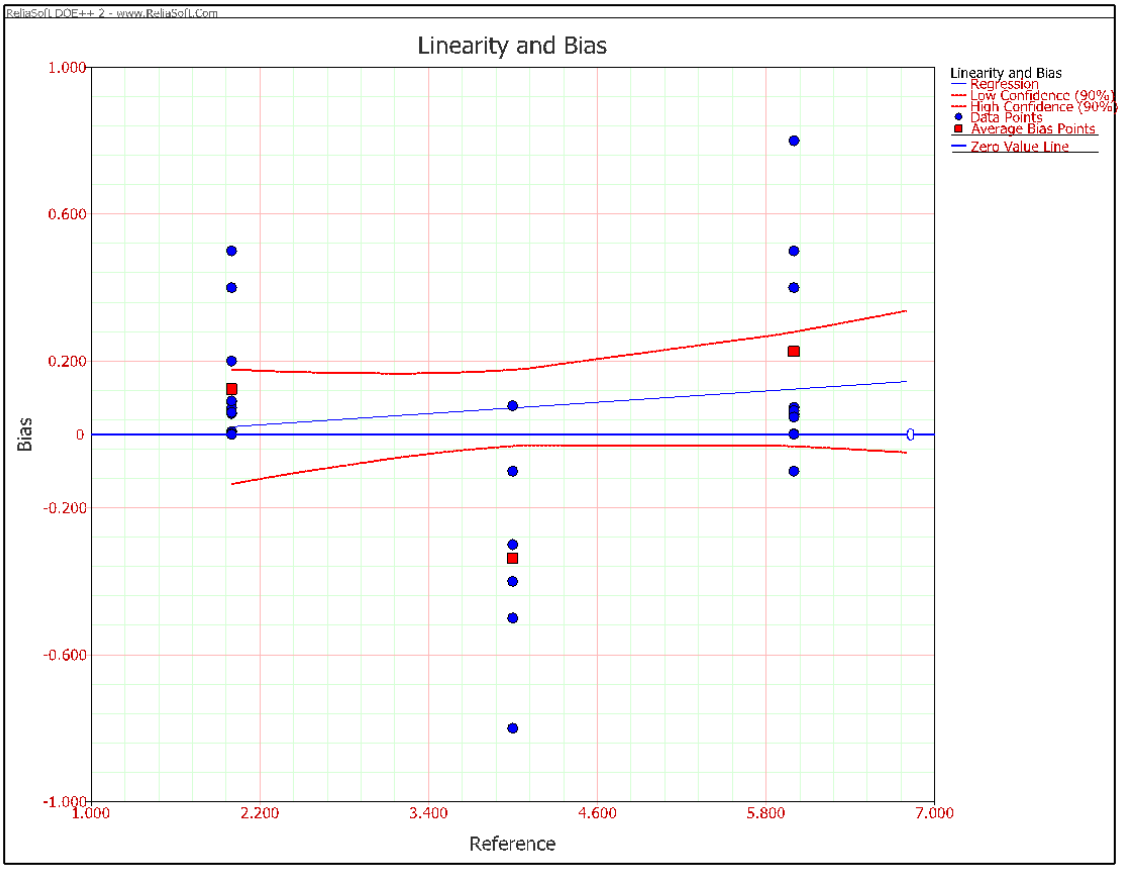
This means that when this gage is used for a process, the observed process variation will be $|\beta_1|$ times larger than the true process variation. This is because the observed value of a part is $|\beta_1|$ times larger/smaller than the true value plus a constant value of the intercept.

The percentage of linearity (% linearity) is defined by:

$$\%\text{linearity} = 100\left(\frac{\text{linearity}}{\text{process variation}}\right)\% = (100\beta_1)\%$$

% linearity shows the percentage of increase of the process variation due to the linearity of the gage. The smaller the linearity, the better the gage is.

If the linearity study shows no linear relation between reference and bias, you need to check the scatter plot of reference and bias to see if there is a non-linear relation. For example, the following plot shows a non-linear relationship between reference and bias.



Although the slope in the linear equation is almost 0 in the above plot, it does not mean the gage is accurate. The above figure shows an obvious V-shaped pattern between reference and

bias. This non-linear pattern requires further analysis to judge whether the gage's accuracy is acceptable.

Results for Bias Study

The bias study results are:

Reference	Bias	%Bias	Std of Mean	t	p
Average	0.1253	2.09%	0.017	7.3517	0.0000
2	-0.0060	0.10%	0.0183	0.3284	0.7501
4	0.1000	1.67%	0.0191	5.2223	0.0020
6	0.1250	2.08%	0.0385	3.2437	0.0229
8	0.2360	3.93%	0.0587	4.0203	0.0159
10	0.2817	4.70%	0.0652	4.3209	0.0076

- The Average row is the average of all the bias values while other rows are the reference values used in the study.
- The second column is the average bias for each reference value.
- The 3rd column is $|bias|/\text{process variation} \times 100\%$. Process variation is commonly defined as 6 times the process standard deviation. For this example, the process standard deviation is set to 1 and the process variation is 6.
- The 4th column is the standard deviation of the mean value of the bias for each reference value. If there are multiple parts having the same reference value, it is the pooled standard deviation of all the parts.

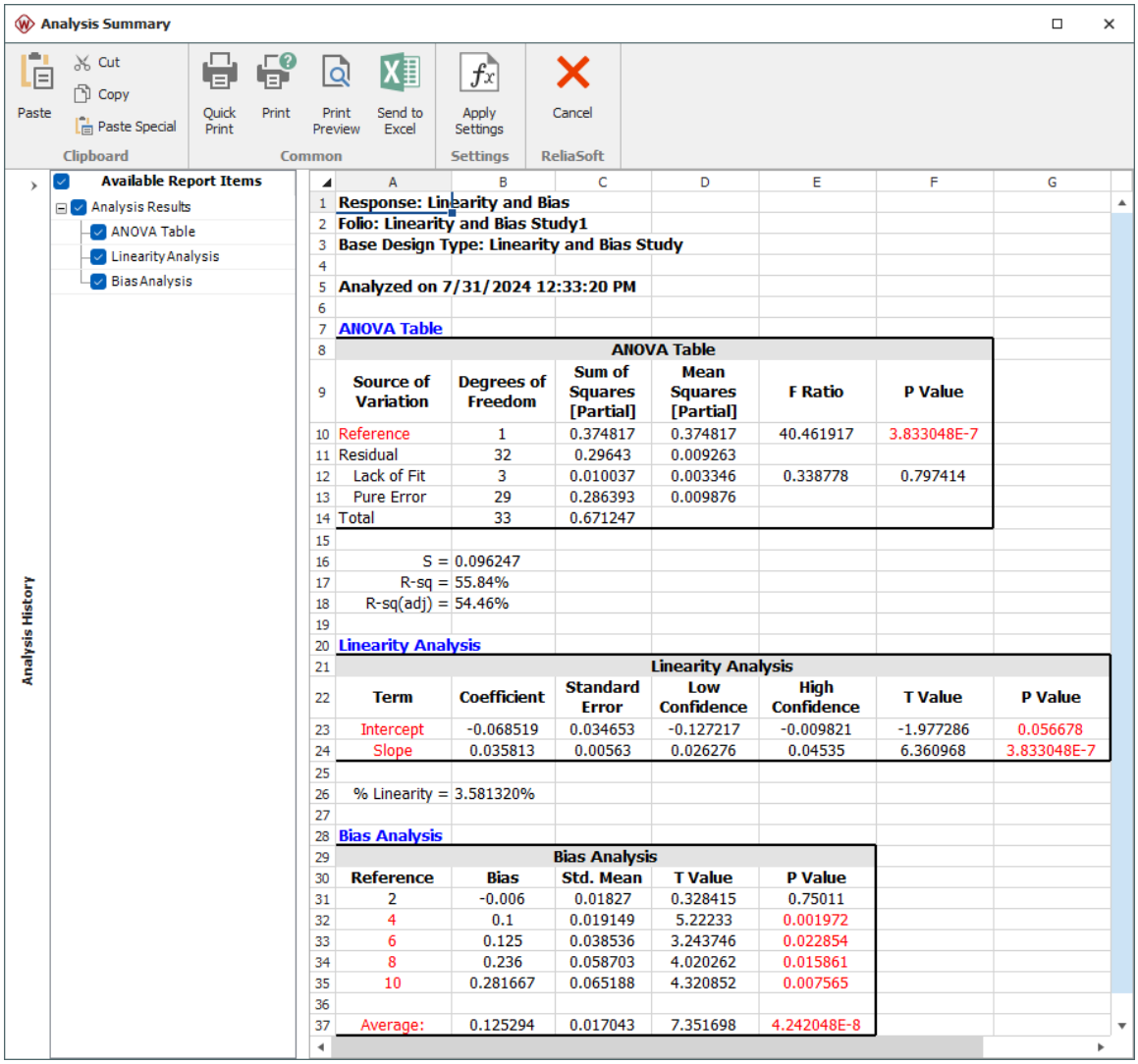
The T value is the ratio of the absolute value of the 2nd column and the 4th column. The p value is calculated from the T value and the corresponding degree of freedom for each reference value. If the p value is smaller than a given significance level, say 0.05, then the corresponding row has significant bias.

For this example, the p value column shows that bias appears for all the reference values except for the reference value of 2. The p value for Average row is very small, which means the average bias of all the readings is significant.

In some cases, such as the figure in the previous section, non-linearity occurs. Bias values are negative for some of the references and positive for others. Although each of the reference values can have significant bias, the average bias of all the references may not be significant.

When there are multiple parts for the same reference value, the standard deviation for that reference value is the pooled standard deviation of all the parts with the same reference value. The standard deviation for the average is calculated from the variance of all the parts.

There are no clear cut-off values for what percent of linearity and bias are acceptable. Users should make their decision based on their engineering feeling or experience. The results from the Weibull++ DOE folio is given in the following picture.



	A	B	C	D	E	F	G
1	**Response: Linearity and Bias**						
2	**Folio: Linearity and Bias Study1**						
3	**Base Design Type: Linearity and Bias Study**						
4							
5	**Analyzed on 7/31/2024 12:33:20 PM**						
6							
7	**ANOVA Table**						

ANOVA Table

Source of Variation	Degrees of Freedom	Sum of Squares [Partial]	Mean Squares [Partial]	F Ratio	P Value
Reference	1	0.374817	0.374817	40.461917	3.833048E-7
Residual	32	0.29643	0.009263		
Lack of Fit	3	0.010037	0.003346	0.338778	0.797414
Pure Error	29	0.286393	0.009876		
Total	33	0.671247			

S =	0.096247
R-sq =	55.84%
R-sq(adj) =	54.46%

Linearity Analysis

Term	Coefficient	Standard Error	Low Confidence	High Confidence	T Value	P Value
Intercept	-0.068519	0.034653	-0.127217	-0.009821	-1.977286	0.056678
Slope	0.035813	0.00563	0.026276	0.04535	6.360968	3.833048E-7

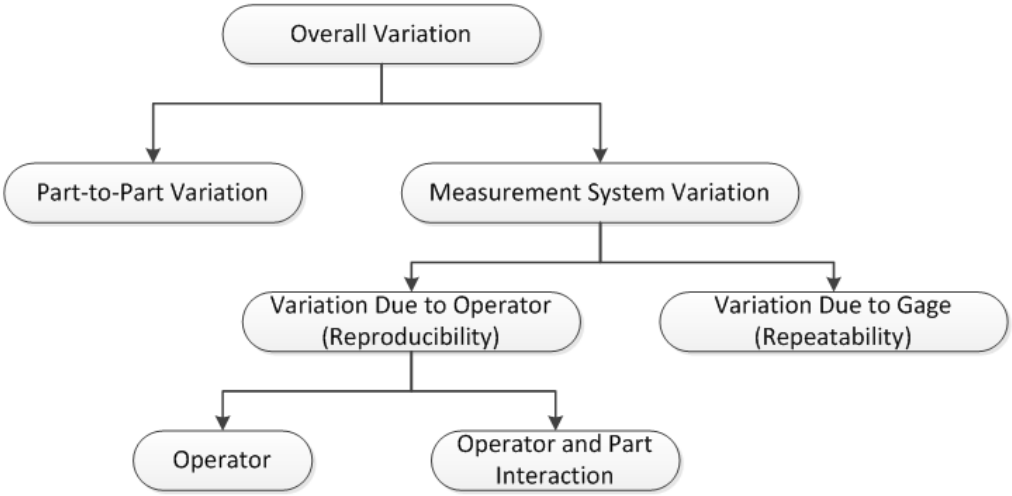
% Linearity =	3.581320%

Bias Analysis

Reference	Bias	Std. Mean	T Value	P Value
2	-0.006	0.01827	0.328415	0.75011
4	0.1	0.019149	5.22233	0.001972
6	0.125	0.038536	3.243746	0.022854
8	0.236	0.058703	4.020262	0.015861
10	0.281667	0.065188	4.320852	0.007565
Average:	0.125294	0.017043	7.351698	4.242048E-8

Gage Repeatability and Reproducibility Study

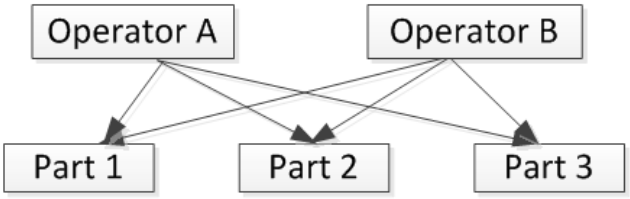
In the previous section, we discussed how to evaluate the accuracy of a measurement device by conducting a linearity and bias study. In this section, we will discuss how to evaluate the precision of a measurement device. Less variation means better precision. Gage repeatability and reproducibility (R&R) is a method for finding out the variations within a measurement system. Basically, there are 3 sources for variation: variation of the part, variation of the measurement device, and variation of operator. Variation caused by operator and interaction between operator and part is called reproducibility and variation caused by measurement device is called repeatability. The formal definitions of reproducibility and repeatability are given in the introduction of this chapter. In this section, we will briefly discuss how to calculate them. For more detail, please refer to Montgomery and Runger, 1993. The following picture shows the decomposition of variations for a product measured by a device.



Depending on how an experiment was conducted, there are two types of gage R&R study.

- When each part is measured multiple times by each operator, it is called a gage R&R crossed experiment.
- When each part is measured by only one operator, such as in destructive testing, this is called a gage R&R nested experiment.

The following picture represents a crossed experiment.



In the above picture, operator A and operator B measured the same three parts. In a nested experiment, each operator measures different parts, as illustrated below.

The X-bar and R chart methods and the ANOVA method have been used to provide an estimation of the variance for each variation source in a measurements system. The X-bar and R chart methods cannot calculate the variance of operator by part interaction. In Weibull++ DOE folios, we use the ANOVA method as discussed by Montgomery and Runger. The ANOVA method is the classical method for estimating variance components in designed experiments. It is more accurate than the X-bar and R chart methods.

In order to estimate variance, each part needs to be measured multiple times. For destructive testing, this is impossible. Therefore, some assumptions have to be made. Usually, for destructive testing, we need to assume that all the parts within the same batch are identical enough to claim that they are the same part. Nested design is the first option for destructive testing since each operator measures unique parts. If a part can be measured multiple times by different operators, then you would use crossed design.

From the above discussion, we know the total variability can be broken down into the following variance components:

$$\begin{aligned}\sigma^2_{total} &= \sigma^2_{part} + \sigma^2_{gage} \\ &= \sigma^2_{part} + \left(\sigma^2_{repeatability} + \sigma^2_{reproducibility}\right) \\ &= \sigma^2_{part} + \left[\sigma^2_{repeatability} + \left(\sigma^2_{operator} + \sigma^2_{operator\times part}\right)\right]\end{aligned}$$

In practice, $6\sigma_{gauge}$ is called gage variation. It is compared to the specification or tolerance of the product measured using this gage to get the so called precision-to-tolerance ratio (or P/T ratio), as given by:

$$\frac{P}{T} = \frac{6\sigma_{gage}}{USL - LSL}$$

where USL and LSL are the upper and lower specification limits of the product under study.

If the P/T ratio is 0.1 or less, this implies adequate gage capability. There are obvious dangers in relying too much on the P/T ratio. For example, the ratio may be made arbitrarily small by increasing the width of the specification tolerance [AIAG]. Therefore, other ratios are also often used. One is the gage to part variation ratio:

$$\frac{Gage}{Part} = \frac{\sigma_{gage}}{\sigma_{part}}$$

The other is the gage to total variation ratio:

$$\frac{Gage}{Total} = \frac{\sigma_{gage}}{\sigma_{total}}$$

The smaller the above two ratios, the higher the relative precision of the gage is. The calculations for obtaining the above variance components for nested design and for crossed design are different.

We should be aware that gage R&R study should be conducted only when gage linearity and bias are not found to be significant.

Gage R&R Study for Crossed Experiments

From a design of experiment point of view, the experiment for gage R&R study is a general level 2 factorial design. Denoting the measurement by operator i on part j at replication k as Y_{ijk} , we have the following ANOVA model:

$$Y_{ijk} = \mu + O_i + P_j + (OP)_{ij} + \varepsilon_{kij} \begin{cases} i = 1, 2, \ldots, o; \\ j = 1, 2, \ldots, p; \\ k = 1, 2, \ldots, n \end{cases}$$

where:

- O_i is the effect of the ith operator.
- P_j is the effect of the jth operator.
- $(OP)_{ij}$ represents the part and operator interaction.
- ε_{kij} is the random error that represents the repeatability.

Usually, all the effects in the above equation are assumed to be random effects that are normally distributed with mean of 0 and variance of σ_O^2, σ_P^2, σ_{OP}^2, and σ_e^2, respectively. When the operators in the study are the only operators who will work on the product, operator could be treated as fixed effect. However, as pointed out by Montgomery and Runger, it is usually desirable to regard the operators as representatives of a larger operator population, with the specific operators having been randomly selected for the gage R&R study. Therefore, the operator should always be treated as a random effect. The definitions of fixed and random effects are:

- **Fixed Effect**: An effect associated with a factor that has a limited number of levels or in which only a limited number of levels are of interest to the experimenter.
- **Random Effect**: An effect associated with a factor chosen at random from a population having a large or infinite number of possible values.

A model that has only fixed effect factors is called a fixed effect model; a model that has only random effect factors is called a random effect model; a model that has both random and fixed effect factors is called a mixed effect model.

For random and mixed effect models, variance components can be estimated using least squares estimation, maximum likelihood estimation (MLE), and restricted MLE (RMLE) methods. The general calculations for variance components and F test in the ANOVA table are beyond the discussion of this chapter. For detail, readers are referred to Searl 1971 and 1997. However, when the design is balanced, variance components can be estimated using the regular linear regression method discussed in the general level factorial design chapter. Weibull++ DOE folios use this method for balanced designs.

When a design is balanced, the expected mean squares for each effect in the above random effect model for gage R&R study using crossed design are:

Mean Squares	Degree of Freedom	Expected Mean Squares
MS_O	$o-1$	$\sigma^2 + np\sigma_O^2 + n\sigma_{OP}^2$
MS_P	$p-1$	$\sigma^2 + no\sigma_P^2 + n\sigma_{OP}^2$
MS_{OP}	$(o-1)(p-1)$	$\sigma_e^2 + n\sigma_{OP}^2$
MS_E	$(n-1)op$	σ_e^2

The mean squares in the first column can be estimated using the model given at the beginning of this section. Their calculations are the same regardless of whether the model is fixed, random, or mixed. The difference for fixed, random, and mixed models is the *expected mean*

squares. With the information in the above table, each variance component can be estimated by:

$$\hat{\sigma}_O^2 = (MS_O - MS_{OP}) / (pn);$$

$$\hat{\sigma}_P^2 = (MS_P - MS_{OP}) / (on);$$

$$\hat{\sigma}_{OP}^2 = (MS_{OP} - MS_e) / n$$

$$\hat{\sigma}_e^2 = MS_E$$

For the *F* test in the ANOVA table, the F ratio is calculated by:

$$\begin{cases} F_O = MS_O / MS_{OP} \text{ For operator} \\ F_P = MS_P / MS_{OP} \text{ For part} \\ F_{OP} = MS_{OP} / MS_E \text{ For operator by part interaction} \end{cases}$$

From the above *F* ratio, we can test whether the effect of operator, part, and their interaction are significant or not.

Example: Gage R&R Study for Crossed Experiment

A gage R&R study was conducted using a crossed experiment. The data set is given in the table below. The product tolerance is 2,000. We want to evaluate the precision of this gage using the P/T ratio, gage to part variation ratio and gage to total variation ratio.

Part	Operator	Response
1	A	405
1	A	232
1	A	476
1	B	389
1	B	234
1	B	456
1	C	684
1	C	674
1	C	634

2	A	409
2	A	609
2	A	444
2	B	506
2	B	567
2	B	435
2	C	895
2	C	779
2	C	645
3	A	369
3	A	332
3	A	399
3	B	426
3	B	471
3	B	433
3	C	523
3	C	550
3	C	520

First, using the regular linear regression method, the mean square for each term can be calculated and is given in the following table.

Source of Variation	Degrees of Freedom	Sum of Squares [Partial]	Mean Squares [Partial]	F Ratio	P Value
Part	2	105545.00	52772.00	5.0655	0.0801
Operator	2	332414.00	166207.00	15.9538	0.0124

Part * Operator	4	41672.00	10418.00	1.4924	0.2462
Residual	18	125655.00	6980.85		
Pure Error	18	125655.00	6980.85		
Total	26	605285.00			

All the effects are treated as random effects in the above table. The F ratios are calculated based on the equations given above. They are:

$$\begin{cases} F_O = MS_O/MS_{OP} = 166207/10418 = 15.9538 \\ F_P = MS_P/MS_{OP} = 52772/10418 = 5.0655 \\ F_{OP} = MS_{OP}/MS_E = 10418/6981 = 1.4924 \end{cases}$$

The p value column shows that the operator is the most significant effect since that has the smallest p value. This means that the variation among all the operators is relatively large.

Second, based on the equations for expected mean squares, we can calculate the variance components. They are given in the following table.

Source	Variance	% Contribution
Part	4706.00	15.61%
Reproducibility	18455.60	61.23%
Operator	17309.89	57.43%
Operator*Part	1145.72	3.80%
Repeatability	6980.85	23.16%
Total Gage R&R	25436.46	84.39%
Total Variation	30142.46	100.00%

The above table shows:

$$\hat{\sigma}_O^2 = (MS_O - MS_{OP}) / (pn) = \frac{166207 - 10418}{3 \times 3} = 17309.89$$

$$\hat{\sigma}_P^2 = (MS_P - MS_{OP}) / (on) = \frac{52772 - 10418}{3 \times 3} = 4706.00$$

$$\hat{\sigma}_{OP}^2 = (MS_{OP} - MS_e)/n = \frac{10418 - 6981}{3} = 1145.72$$

The repeatability is σ_e^2 for the random error. The reproducibility is the sum of σ_O^2 and σ_{OP}^2. The sum of repeatability and reproducibility is called the total gage R&R.

The last column in the above table shows the contribution of each variance component. For example, the contribution of the operator is 57.43%, which is calculated by:

$$\%\text{Contribution_Operator} = 17309.89/30142.46 = 57.43\%$$

The standard deviation for each effect is:

Source	Std (SD)
Part	68.600
Reproducibility	135.851
Operator	131.567
Operator*Part	33.848
Repeatability	83.551
Total Gage R&R	159.488
Total Variation	173.616

Since the product tolerance is 2,000, the P/T ratio is:

$$\frac{P}{T} = \frac{6\sigma_{gauge}}{USL - LSL} = \frac{6 \times 159.488}{2000} = 47.85\%$$

Since P/T ratio is much greater than 10%, this gage is not adequate for this product.

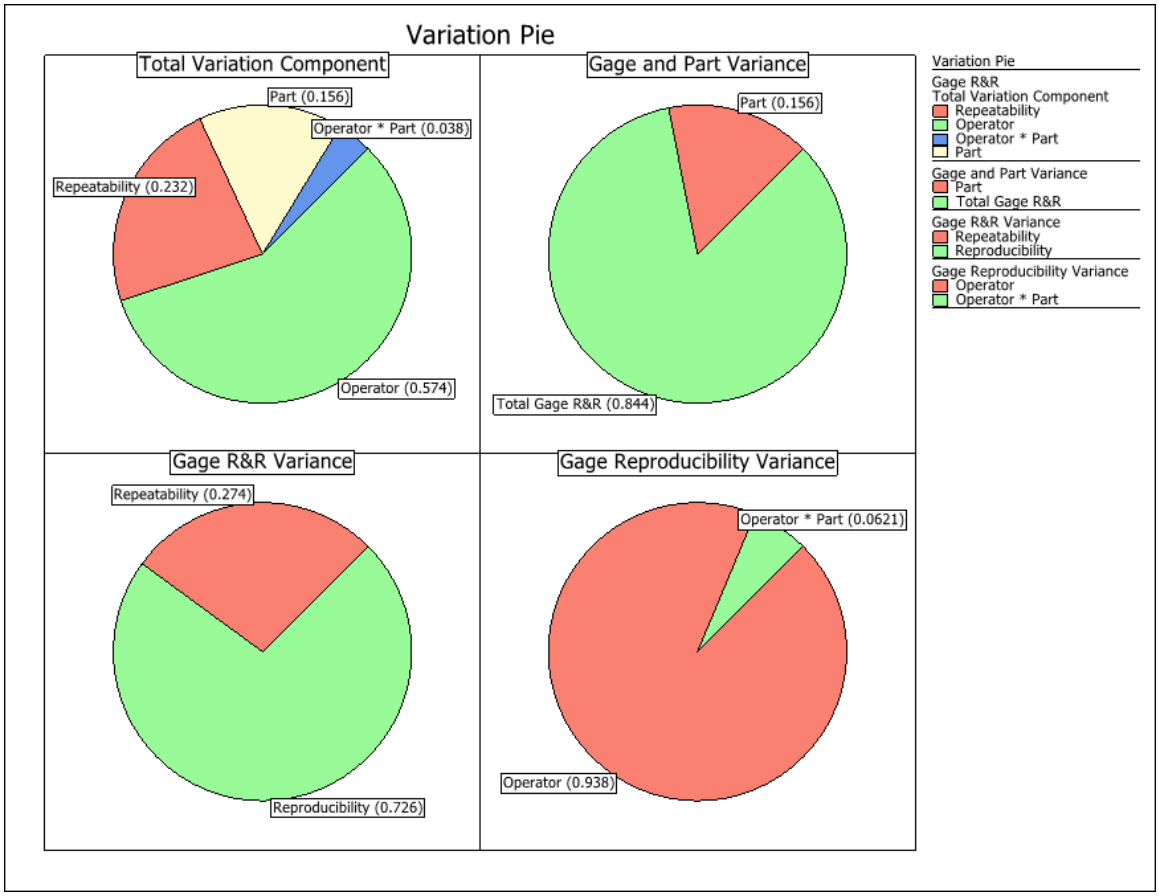
The gage to part variation ratio:

$$\frac{Gage}{Part} = \frac{\sigma_{gage}}{\sigma_{part}} = \frac{159.488}{68.600} = 232.49\%$$

The gage to total variation ratio:

$$\frac{Gage}{Total} = \frac{\sigma_{gage}}{\sigma_{total}} = \frac{159.488}{173.616} = 91.86\%$$

Clearly, all the ratios are too large. The operators should be trained and a new gage may need to be purchased. The pie chart plots for the contribution of each variance components are shown next.



In the above picture, the total variation component pie chart displays the ratio of each variance to the total variance. The gage and part variation chart displays the ratio of the gage variance to the total variance, and the ratio of the part to the total variance. The gage R&R variance is for the percentage of repeatability and reproducibility to the total gage variance. The gage reproducibility variance pie chart future decomposes reproducibility to operator variance, and operator and part interaction variance.

Gage R&R Study for Nested Experiments

When the experiment is nested, since the part is nested within each operator, we cannot assess the operator and part interaction. The regression model is:

$$Y_{ijk} = \mu + O_i + P_{j(i)} + \varepsilon_{kij} \begin{cases} i = 1, 2, \ldots, o; \\ j = 1, 2, \ldots, p; \\ k = 1, 2, \ldots, n \end{cases}$$

The estimated operator effect includes the operator effect and the operator and part interaction. For the general calculation on the above model, please refer to ["Applied Linear Statistical Models" by Kutner, Nachtsheim, Neter and Li]. When the nested experiment is balanced, its calculations for total sum of squares (SST), sum of squares of operator (SSO), and sum of square of error (SSE) are the same as those for the crossed design. The only difference is the sum of squares of part (SSP(O)). For nested designs, it is:

$$SS_{P(O)} = SS_P + SS_{OP}$$

SSP and SSOP are the sum of squares for part, and the sum of squares for part and operator interaction. They are calculated using a linear regression equation by including part and operator interaction in the model.

When the design is balanced, the expected mean squares for each effect in the above random effect model for gage R&R study nested design are:

Mean Squares	Degree of Freedom	Expected Mean Squares
MSO	$o-1$	$\sigma^2 + np\sigma_O^2 + n\sigma_{P(O)}^2$
MSP(O)	$o(p-1)$	$\sigma^2 + n\sigma_{P(O)}^2$
MSE	$(n-1)op$	σ_e^2

With the information in the above table, each variance component can be estimated by:

$$\hat{\sigma}_O^2 = \left(MS_O - MS_{P(O)}\right) / (pn);$$

$$\hat{\sigma}_{P(O)}^2 = \left(MS_{P(O)} - MS_E\right) / n;$$

$$\hat{\sigma}_e^2 = MS_E$$

For the F test in the ANOVA table, the F ratio is calculated by:

$$\begin{cases} F_O = MS_O / MS_{P(O)} \text{ For operator} \\ F_{P(O)} = MS_P / MS_E \text{ For part (nested in operator)} \end{cases}$$

Example: Gage R&R Study for Nested Experiment

For the example in the previous section, since it is a nested design, the part i measured by one operator is different from the part i measured by another operator. Therefore, when the design is nested, the design in fact should be:

Part	Operator	Response
1_1	A	405
1_1	A	232
1_1	A	476
2_1	B	389
2_1	B	234
2_1	B	456
3_1	C	684
3_1	C	674
3_1	C	634
1_2	A	409
1_2	A	609
1_2	A	444
2_2	B	506
2_2	B	567
2_2	B	435
3_2	C	895
3_2	C	779
3_2	C	645
1_3	A	369
1_3	A	332
1_3	A	399
2_3	B	426
2_3	B	471
2_3	B	433

3_3	C	523
3_3	C	550
3_3	C	520

We want to evaluate the precision of this gage using the P/T ratio, gage to part variation ratio, and gage to total variation ratio.

First, using the regular linear regression method for nested designs [Neter's book], all the mean squares for each term can be calculated. They are given in the following table.

Source of Variation	Degrees of Freedom	Sum of Squares [Partial]	Mean Squares [Partial]	F	P
Operator	2	332414.00	166207.00	6.77396	0.028917
Part(Operator)	6	147217.00	24536.17	3.514781	0.017648
Residual	18	125655.00	6980.85		
Pure Error	18	125655.00	6980.85		
Total	26	605285.00			

The F ratios are calculated based on the equations given above.

$$\begin{cases} F_O = MS_O/MS_{P(O)} = 166207/24536.17 = 6.774 \\ F_{P(O)} = MS_P/MS_E = 24536.17/6980.85 = 3.515 \end{cases}$$

The p value column shows that the operator and part (operator) both are significant at a significance level of 0.05.

Second, based on the equations for expected mean squares, we can calculate the variance components. They are given in the following table.

Source	Variance	% Contribution
Repeatability	6981	24.43%
Reproducibility	15741.2037	55.09%

Operator	15741.2037	55.09%
Part (Operator)	5851.7716	20.48%
Total Gage R&R	22722	79.52%
Total Variation	28574	100.00%

The standard deviation for each variation source is:

Source	Std (SD)
Repeatability	83.551
Reproducibility	125.464
Operator	125.464
Part (Operator)	76.497
Total Gage R&R	150.738
Total Variation	169.038

Since the product tolerance is 2,000, the P/T ratio is:

$$\frac{P}{T} = \frac{6\sigma_{gage}}{USL - LSL} = \frac{6 \times 150.738}{2000} = 45.22\%$$

Since the P/T ratio is much greater than 10%, this gage is not adequate for this product.

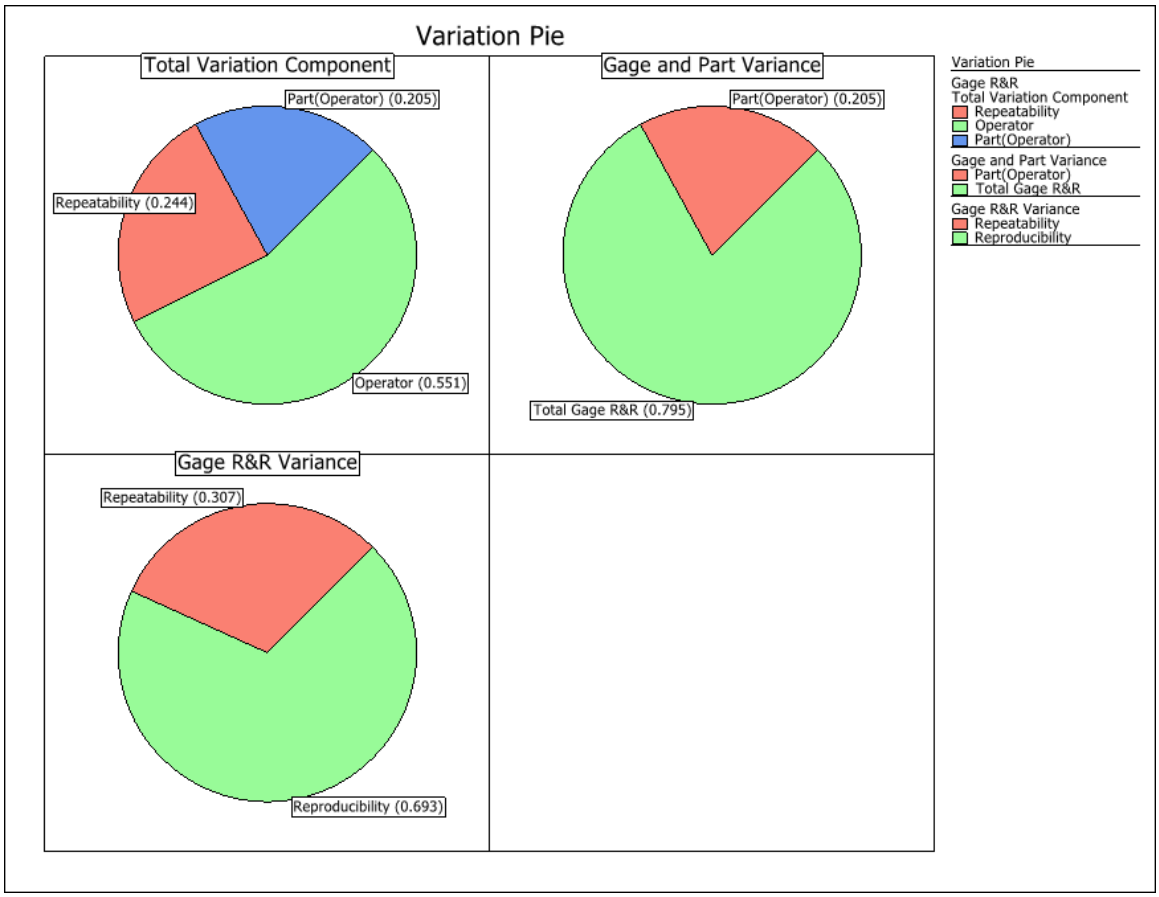
The gage to part variation ratio:

$$\frac{Gage}{Part} = \frac{\sigma_{gage}}{\sigma_{part(operator)}} = \frac{150.738}{76.497} = 197.05\%$$

The gage to total variation ratio:

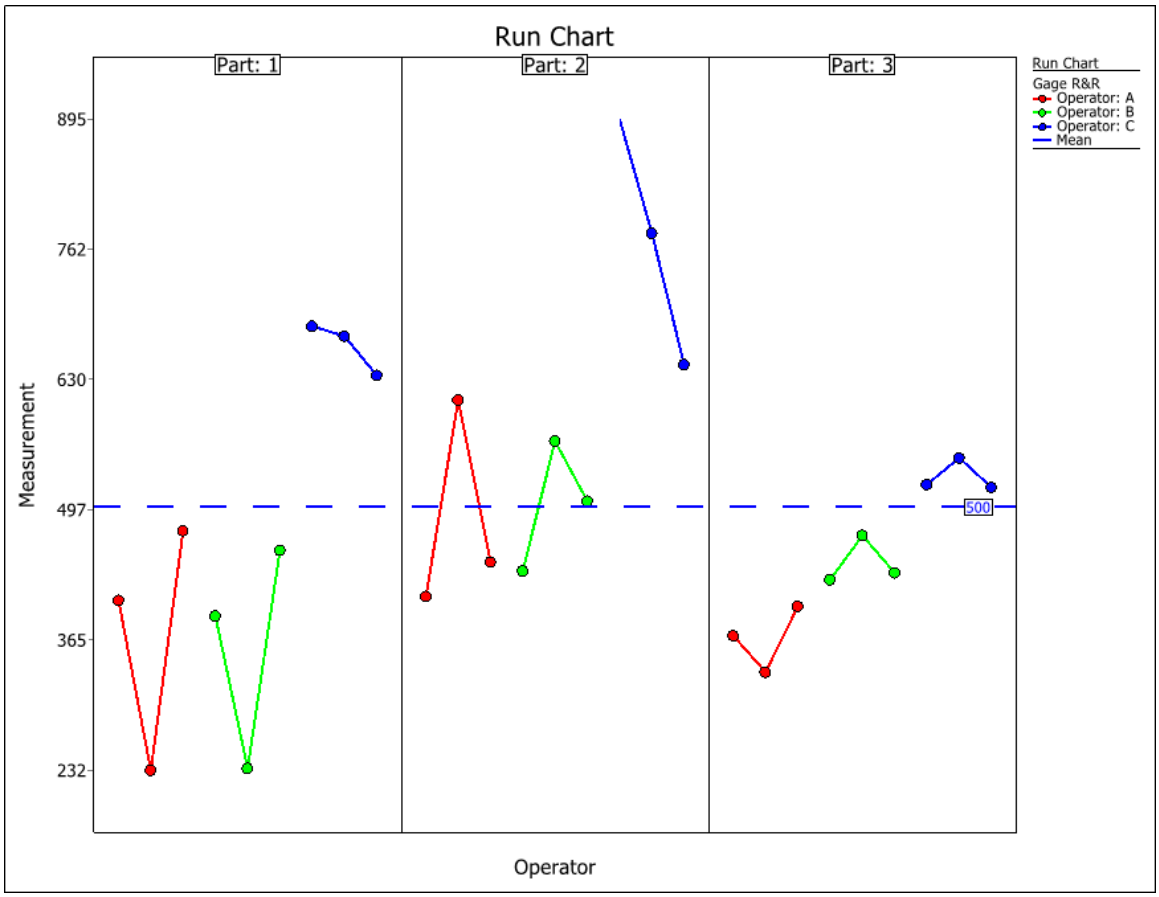
$$\frac{Gage}{Total} = \frac{\sigma_{gage}}{\sigma_{total}} = \frac{150.738}{169.038} = 89.17\%$$

The pie charts for all the variance components are shown next.

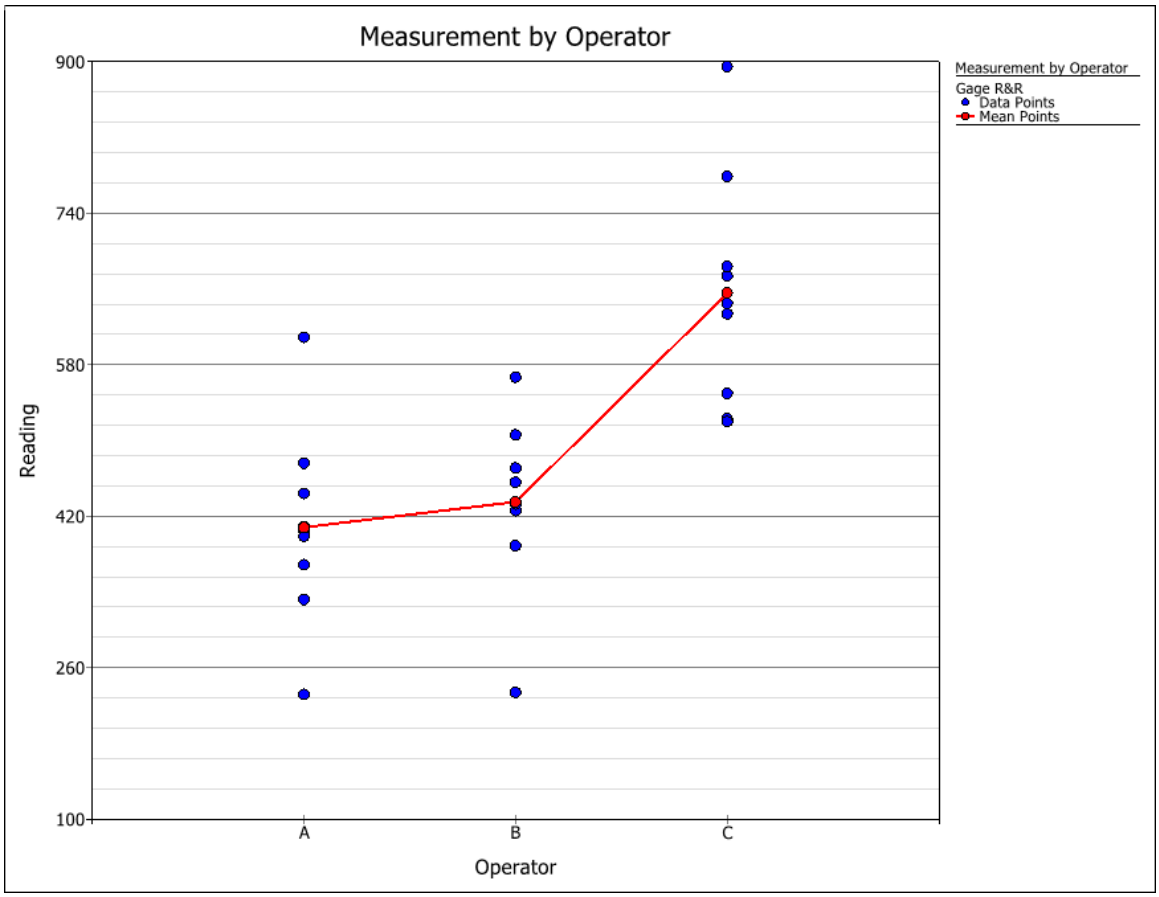


X-bar and R Charts in Gage R&R

X-bar and R charts are often used in gage R&R studies. Although Weibull++ DOE folios do not use them to estimate repeatability and reproducibility, they are included in the plot to visually display the data. Along with X-bar and R charts, other plots are also used in DOE folios. For example, the following is a run chart for the example of gage R&R study using crossed design.

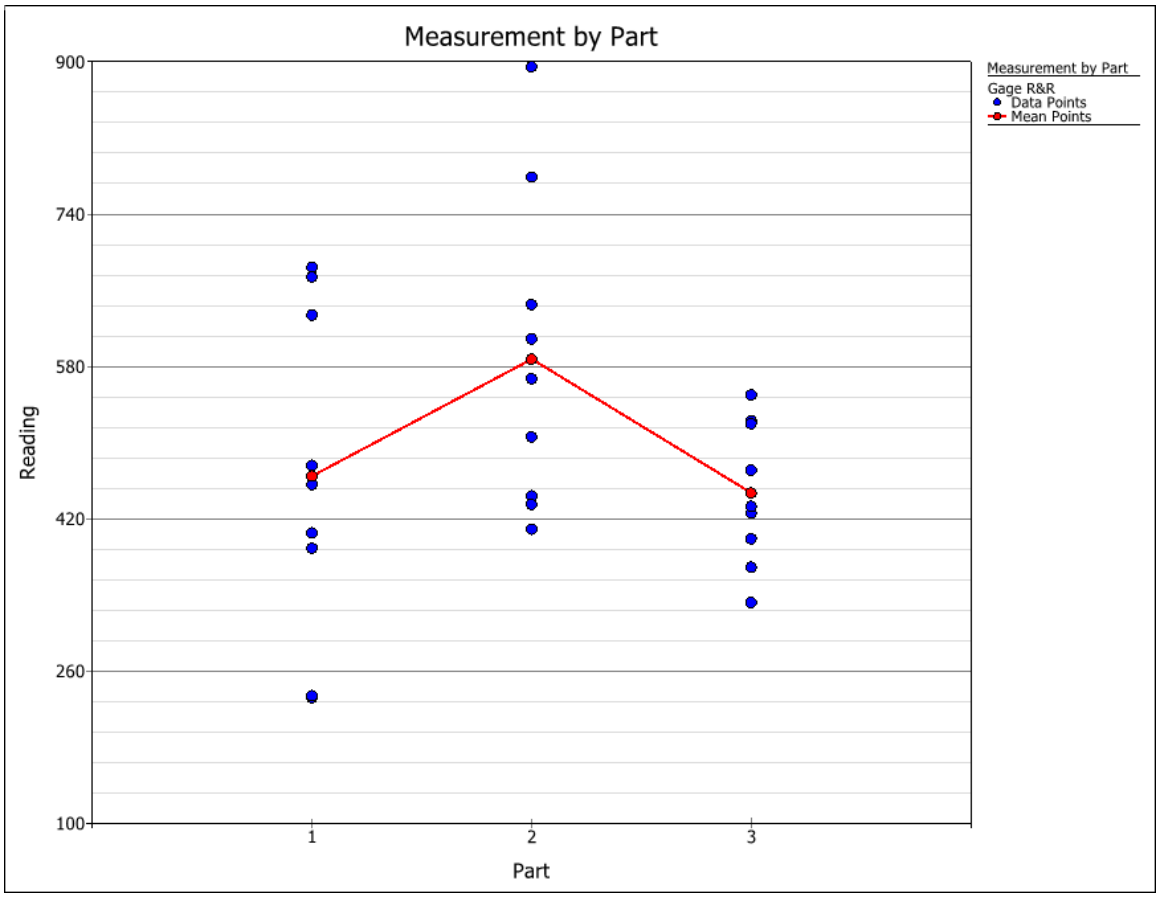


Each column in the above figure is the 9 measurements of a part by all the operators. In the above plot, we see that all readings by operator C (the blue points) are above the mean line. This indicates that operator C's readings are different from the calculated mean. Part 3 (the last column in the plot) has the least variation among these 3 parts. These two conclusions also can be inferred from the following two plots.

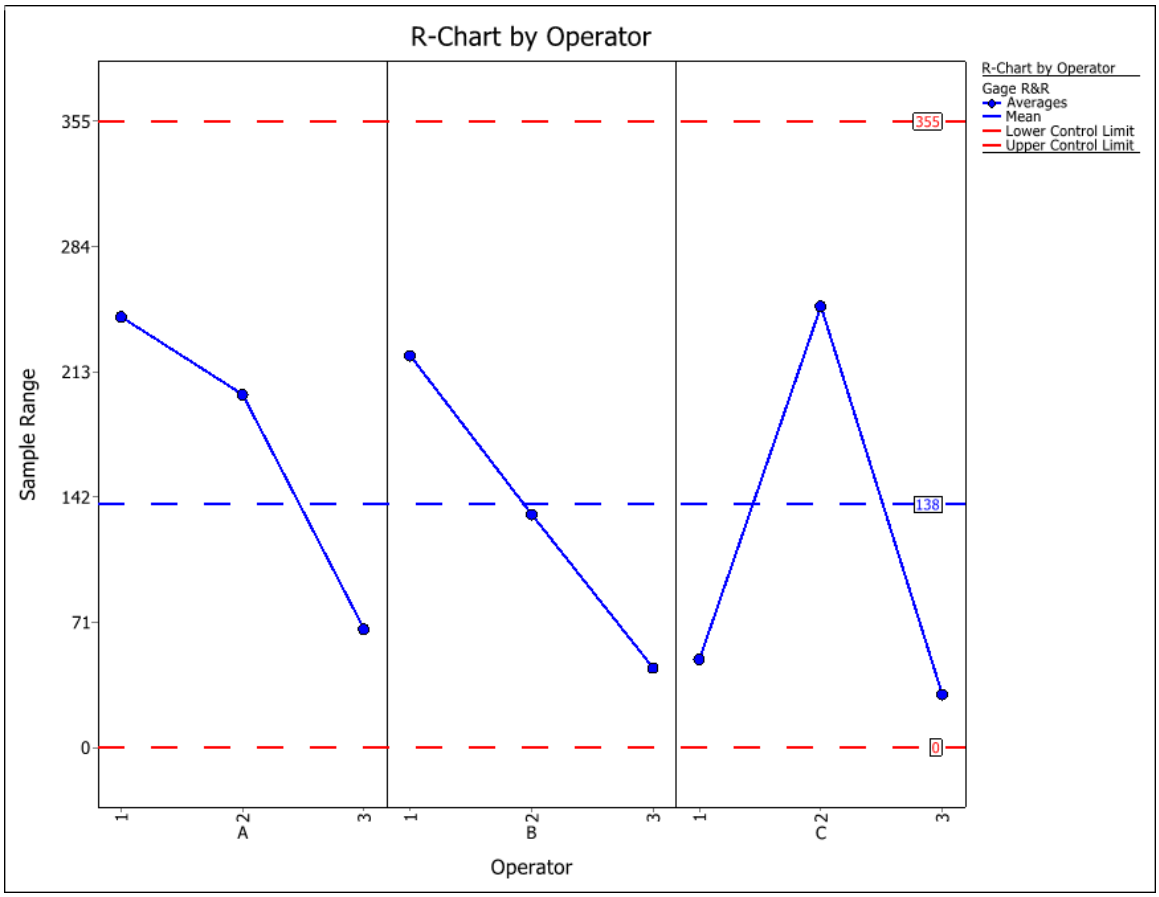


The above plot shows that operator C's readings are much higher than the other two operators.

The above plot shows part 3 has less variation compared to parts 1 and 2.



Now let's talk about X-bar and R charts. The X-bar chart is used to see how the mean reading changes among the parts; the R chart is used to check the repeatability. When the number of readings of each part by the same operator is greater than 10, an s chart is used to replace the R chart. The R chart is accurate only when the sample size is small (<10). For this example, the sample size is 3, so the R chart is used, as shown next.



In the above plot, the x-axis is operator and the y-axis is the range for each part measured by each operator.

The step-by-step calculation for the R chart (n $\leq$10) is given below.

Step 1: calculate the range of each part for each operator.

$$R_{i,j} = \max\left(X_{ijk}\right) - \min\left(X_{ijk}\right)$$

$R_{i,j}$ is the range of the reading for the ith part and the jth operator. k is the trial number.

Step 2: calculate the average range for each operator.

$$\bar{R}_j = average(R_{i,j})$$

Step 3: calculate the overall average range for all the operators.

$$\bar{\bar{R}} = average(\bar{R}_j)$$

This is the central line in the R chart.

Step 4: calculate the upper control limit (UCL) and the lower control limit (LCL) for the R chart.

$$UCL = D4 * \bar{\bar{R}};$$
$$LCL = D3 * \bar{\bar{R}}$$

D3 and D4 are from the following table:

n	A2	D3	D4	d2
2	1.88	0	3.267	1.128
3	1.023	0	2.575	1.693
4	0.729	0	2.282	2.059
5	0.577	0	2.115	2.326
6	0.483	0	2.004	2.534
7	0.419	0.076	1.924	2.704
8	0.373	0.136	1.864	2.847
9	0.337	0.184	1.816	2.97
10	0.308	0.223	1.777	3.078

The calculation results for this example are:

	Operator A					
Part Number	**T1**	**T2**	**T3**	$\bar{R}_{ij}$	$\bar{\bar{X}}_{ij}$	$\bar{\bar{X}}_j$
1	405	232	476	244	371	
2	409	609	444	200	487.3333	408.3333
3	369	332	399	67	366.6667	
	Operator B					
1	389	234	456	222	359.6667	
2	506	567	435	132	502.6667	435.2222
3	426	471	433	45	443.3333	

	Operator C					
1	684	674	634	50	664	
2	895	779	645	250	773	656
3	523	550	520	30	531	
			$\bar{\bar{R}}$	137.7778	$\bar{\bar{X}}$	499.8519

From the above table, we know that the three values for the R chart are:

$$UCL = D4 * \bar{\bar{R}} = 2.575 \times 137.778 = 354.7784$$
$$\bar{\bar{R}} = 137.7778$$
$$LCL = D3 * \bar{\bar{R}} = 0$$

The step by step calculation for the X-bar chart for sample size *n*, where n is less than or equal to 10, is given below.

Step 1: Calculate the average of the reading for part i, by operator j.

$$\bar{X}_{ij} = average\,(X_{ijk})$$

Step 2: Calculate the overall mean of operator j.

$$\bar{\bar{X}}_j = average\,(\bar{X}_{ij})$$

Step 3: Calculate the overall mean of all the observations:

$$\bar{\bar{X}} = average(\bar{\bar{X}}_j)$$

$\bar{\bar{X}}$ is the central line of the X-bar chart.

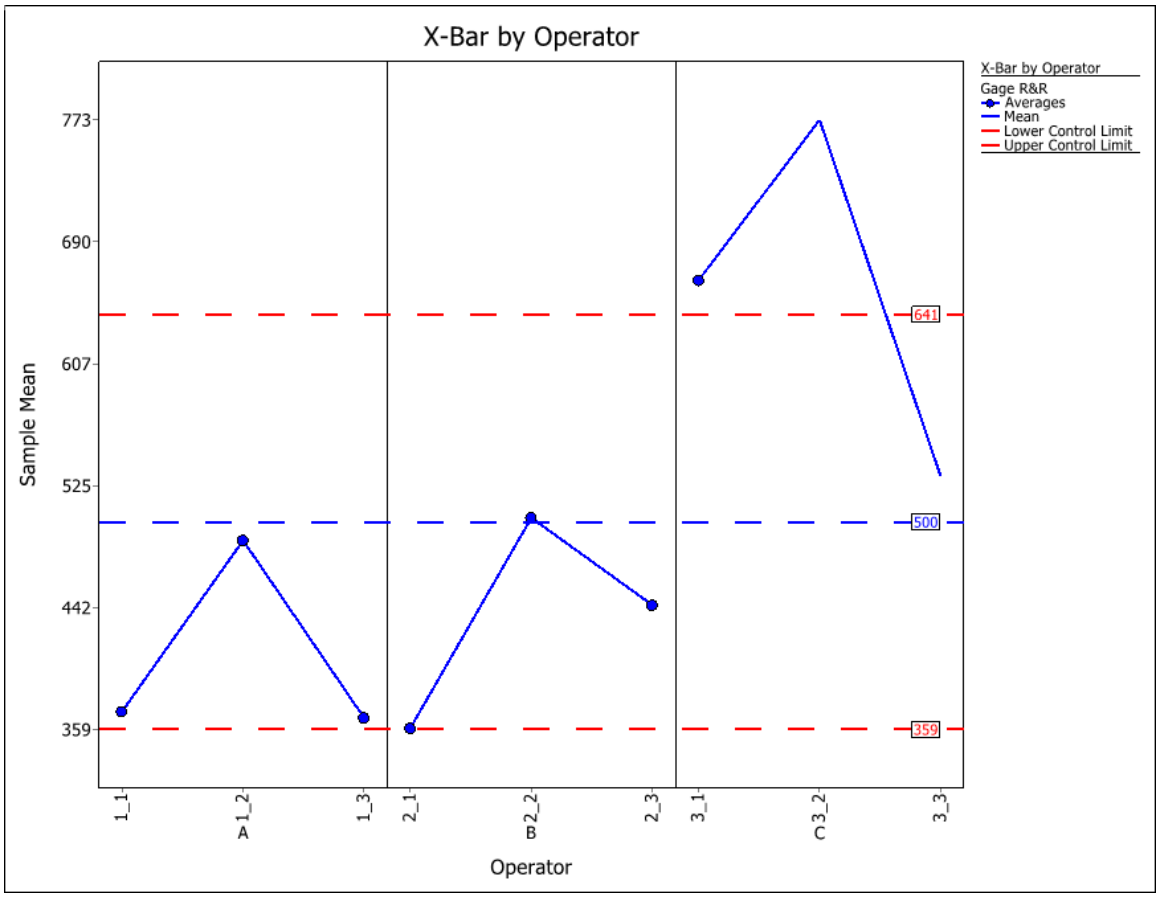
The above table gives the values of $\bar{X}_{ij}$, $\bar{\bar{X}}_j$, and $\bar{\bar{X}}$.

Step 4: Calculate the UCL and LCL.

$$UCL = \bar{\bar{X}} + A2 * \bar{\bar{R}}$$

$$LCL = \bar{\bar{X}} - A2 * \bar{\bar{R}}$$

A2 is from the above constant value table. The X-bar chart for this example is:



When the sample size (the reading of the same part by the same operator) is greater than 10, the more accurate s chart is used to replace the R chart. The calculation for the UCL and LCL in the X-bar chart is also updated using the sample standard deviation s.

The step by step calculations for the s chart are given below.

Step 1: Calculate the standard deviation for each part of each operator.

$$S_{ij} = \frac{1}{n-1}\sum_{k}(x_{ijk} - \bar{x}_{ij})$$

Step 2: Calculate the average of these standard deviations.

$$\bar{S} = average(S_{ij})$$

The above equation is only valid for balanced designs.

$\bar{S}$ is the central line for the s chart.

Step 3: Calculate the UCL and LCL.

$$UCL = \bar{S} + 3\frac{\bar{S}}{c_4} \times c_5 ; LCL = \bar{S} - 3\frac{\bar{S}}{c_4} \times c_5$$

where:

$$c_4 = \sqrt{\frac{2}{n-1}} \frac{(\frac{n}{2}-1)!}{(\frac{n-1}{2}-1)!} = \left(\frac{2}{n-1}\right)^{1/2} \frac{\Gamma(n/2)}{\Gamma(\frac{n-1}{2})} ;$$

$$c_5 = \sqrt{1-c_4^2}$$

For the X-bar chart, the central line is the same as before. Only the UCL and LCL need to use the following equations when n>10.

$$UCL = \bar{\bar{x}} + 3\frac{\bar{S}}{c_4\sqrt{n}};$$
$$UCL = \bar{\bar{x}} - 3\frac{\bar{S}}{c_4\sqrt{n}}$$

From the above calculation, it can be seen the calculation for the s chart is much more complicated than the calculation for the R chart. This is why the R chart was often used in the past, before computers were in common use.

Gage Agreement Study

In the above sections, we discussed how to evaluate a gage's accuracy and precision. Accuracy is assessed using a linearity and bias study, while precision is evaluated using a gage R&R study. Often times, we need to compare two measurement devices. For instance, can an old device be replaced by a new one, or can an expensive one be replaced by a cheap one, without loss of the accuracy and precision of the measurements? The study used for comparing the accuracy and precision of two gages is called a gage agreement study.

Accuracy Agreement Study

One way to compare the accuracy of two gages is to conduct a linearity and bias study for each gage by the same operator, and then compare the percentages of the linearity and bias. This provides a rough idea of how close the accuracies of the two gages are. However, it is difficult to quantify how close they should be in order to claim there is no significant difference between them. Therefore, a formal statistical method is needed. Let's use the following example to explain how to compare the accuracy of two devices.

Example: Compare the Accuracy of Two Gages Using a Paired t-Test

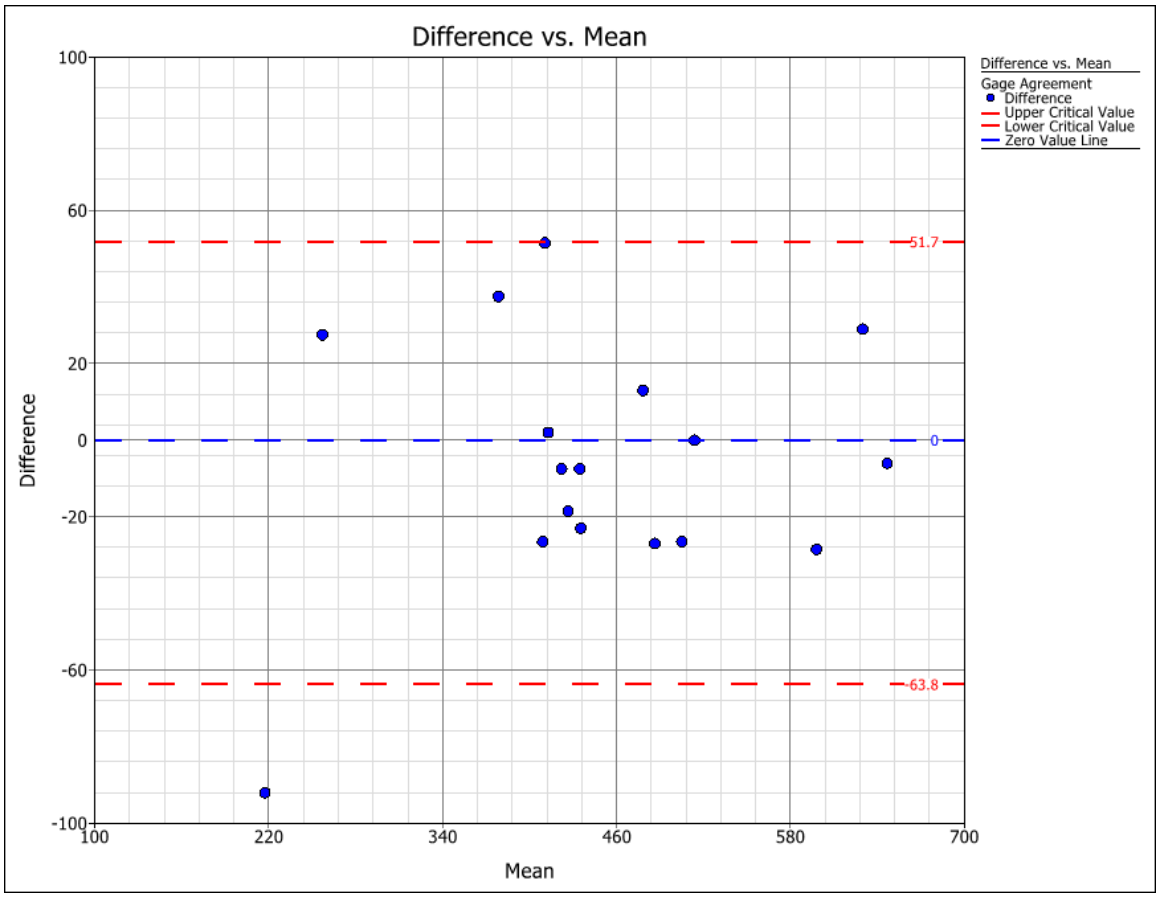
There are two gages: Gage 1 and Gage 2. There are 17 subjects/parts. For each subject, there are two readings from each gage.

	Gage 1		Gage 2	
Subject	**1st Reading**	**2nd Reading**	**1st Reading**	**2nd Reading**
1	494	490	512	525
2	395	397	430	415
3	516	512	520	508
4	434	401	428	444
5	476	470	500	500
6	557	611	600	625
7	413	415	364	460
8	442	431	380	390
9	650	638	658	642
10	433	429	445	432
11	417	420	432	420
12	656	633	626	605
13	267	275	260	227
14	478	492	477	467
15	178	165	259	268
16	423	372	350	370
17	427	421	451	443

If their bias and linearity are the same, then the difference between the average readings for the same subject by the two devices should be almost the same. In other words, the differences should be around 0, with a constant standard deviation. We can test if this hypothesis is true or not. The differences of the readings are given in the table below.

Subject	Gage 1		Gage 2		Difference	Grand Average
	Number of Read-ing	Average Reading	Number of Read-ing	Average Reading		
1	2	492	2	518.5	-26.5	505.25
2	2	396	2	422.5	-26.5	409.25
3	2	514	2	514	0	514
4	2	417.5	2	436	-18.5	426.75
5	2	473	2	500	-27	486.5
6	2	584	2	612.5	-28.5	598.25
7	2	414	2	412	2	413
8	2	436.5	2	385	51.5	410.75
9	2	644	2	650	-6	647
10	2	431	2	438.5	-7.5	434.75
11	2	418.5	2	426	-7.5	422.25
12	2	644.5	2	615.5	29	630
13	2	271	2	243.5	27.5	257.25
14	2	485	2	472	13	478.5
15	2	171.5	2	263.5	-92	217.5
16	2	397.5	2	360	37.5	378.75
17	2	424	2	447	-23	435.5

The difference vs. mean plot is shown next.



The above plot shows that all the values are within the control limits (significant level = 0.05) except for one point, and are evenly distributed around the central 0 line.

The paired t-test is used to test if the two gages have the same bias (i.e., if the "difference" has a mean value of 0). The paired t-test is conducted using the Difference column. The calculation is given below.

Step 1: Calculate the mean value of this column.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

For this example, n is 17.

Step 2: Calculate the standard deviation of this column.

$$Var(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$Std(x) = \sqrt{Var(x)}$$

Step 3: Conduct the t-test.

$$t_0 = \frac{\bar{x}}{std(x)/\sqrt{n}}$$

Step 4: Calculate the p value.

$$p = InvT(df = n - 1, t_0)$$

The calculation will be summarized in the following table.

Mean (Gage 1-Gage 2)	Std. Mean	Lower Bound	Upper Bound	T Value	P Value
6.02941	8.053186092	-23.101404	11.04257999	0.748698924	0.464904

Since the p value is 0.464904, which is greater than the significant level of 0.05, the two gages have the same bias.

The paired t-test is valid only when there is no trend or pattern in the difference vs. mean plot. If the points show a pattern such as a linear pattern, the conclusion from the paired t-test may not be valid.

Example: Compare the Accuracy of Two Gages Using Linear Regression

The data set for a gage agreement study is given in the table below.

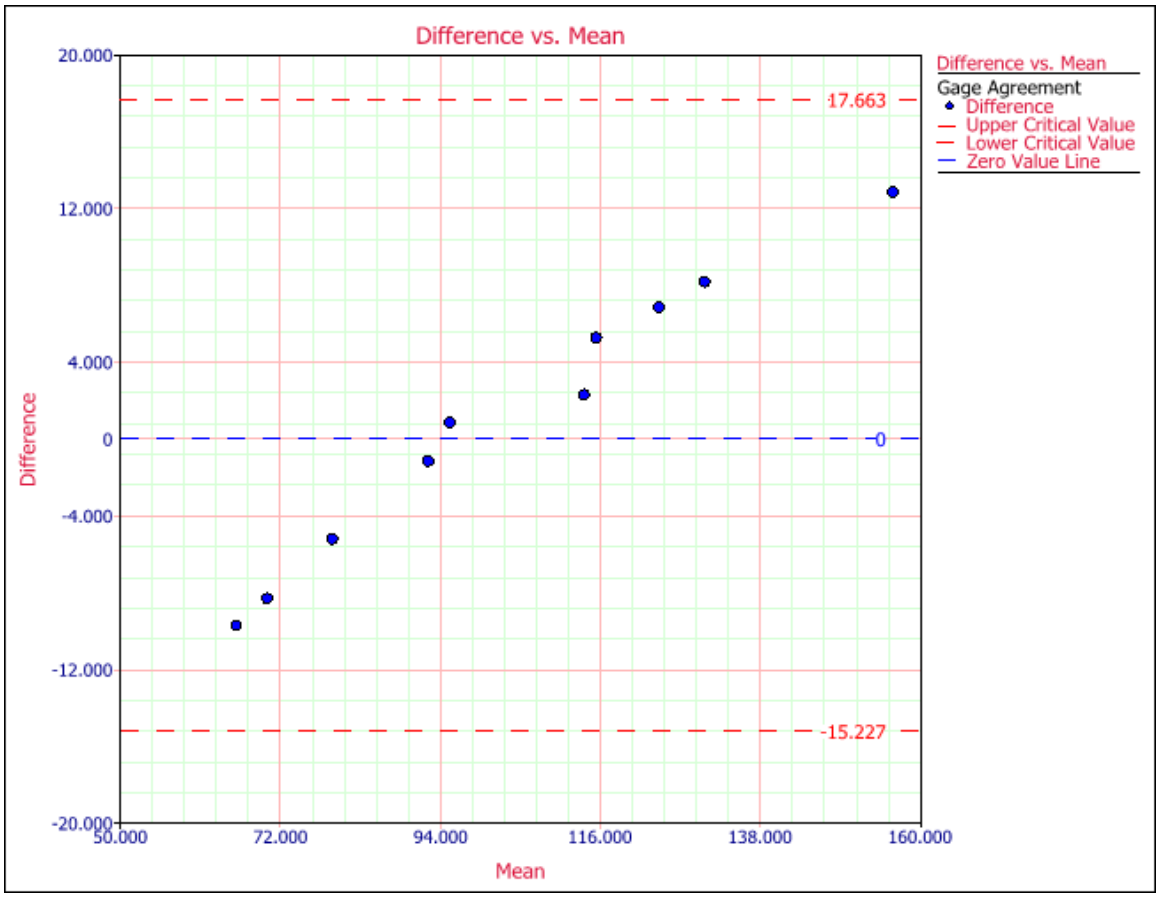
Subject	Gage 1		Gage 2	
	1st Reading	2nd Reading	1st Reading	2nd Reading
1	66.32	65.80	74.30	74.39
2	95.51	95.94	94.74	94.93
3	61.93	60.27	70.81	70.75
4	163.08	162.33	149.91	149.75
5	76.60	76.56	82.00	81.53

6	127.35	127.68	120.58	120.70
7	93.07	90.51	92.96	92.88
8	134.39	134.49	126.24	126.23
9	115.54	114.33	112.27	112.96
10	117.92	118.26	112.41	113.18

The differences of the readings are given in the table below.

Subject	Gage 1		Gage 2		Difference	Grand Average
	Number of Read-ing	Average Reading	Number of Read-ing	Average Reading		
1	2	66.06	2	74.35	-8.29	70.20
2	2	95.72	2	94.83	0.89	95.28
3	2	61.10	2	70.78	-9.67	65.94
4	2	162.70	2	149.83	12.87	156.26
5	2	76.58	2	81.77	-5.19	79.17
6	2	127.52	2	120.64	6.88	124.08
7	2	91.79	2	92.92	-1.13	92.35
8	2	134.44	2	126.24	8.20	130.34
9	2	114.94	2	112.61	2.32	113.77
10	2	118.09	2	112.79	5.30	115.44

The difference vs. mean plot shows a clear linear pattern, although all the points are within the control limits.

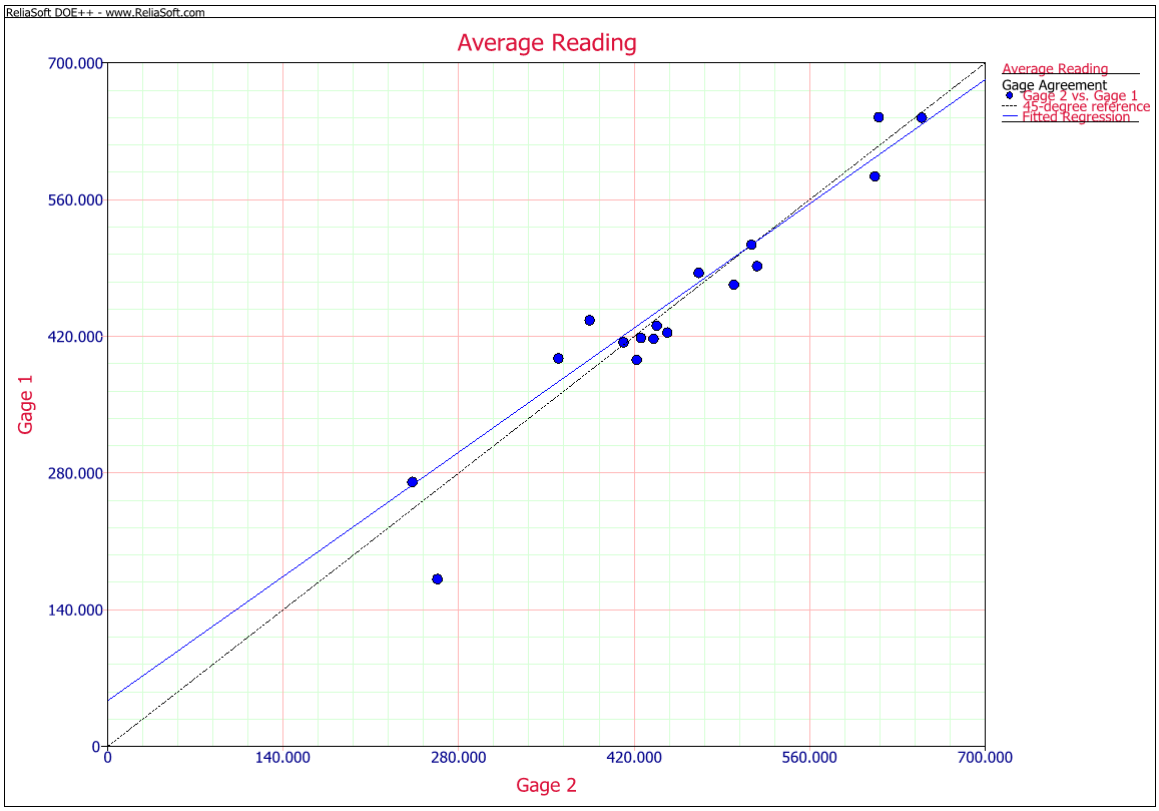


The paired t-test results are:

Mean (Gage 1-Gage 2)	Std. Mean	Lower Bound	Upper Bound	T Value	P Value
1.218	7.3804	17.9137	-15.4777	0.5219	0.6144

Since the p value is large, we cannot reject the null hypothesis. The conclusion is that the bias is the same for these two gages. However, the linear pattern in the above plot makes us suspect that this conclusion may not be accurate. We need to compare both the bias and the linearity. The F-test used in linear regression can do the work.

If the two gages have the same accuracy (linearity and bias), then the average readings from Gage 1 and the average readings from Gage 2 should be on a 45 degree line that passes the origin in the average reading plot. However, the following plot shows the points are not very close to the 45 degree line.



We can fit a linear regression equation:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

where Y is the average reading for each part from Gage 1, and X is the average reading for each part from Gage 2. If the two gages agree with each other, then β_0 should be 0 and β_1 should be one. Using the data in this example, the calculated regression coefficients are:

Term	Coefficient	Standard Error	Low Confidence	High Confidence	T Value	P Value
β_0	22.3873	1.2471	19.5114	25.2633	17.9508	9.51E-08
β_1	0.775	0.0114	0.7487	0.8013	19.7265	4.54E-08

The p values in the above results show that β_0 is not 0 and β_1 is not 1. These tests are for each individual coefficient. For β_0, the t value is:

$$t_0 = \frac{|\beta_0|}{1.2471} = 17.9508$$

For β_1, the t value is

$$t_1 = \frac{|\beta_1 - 1|}{0.0114} = 19.7265$$

The p value is calculated using the above t values and the degree of freedom of error of 8.

Since we want to test these two coefficients simultaneously, using an F-test is more appropriate. The null hypothesis for the F-test is:

$$H0: (\beta_0, \beta_1) = (0, 1)$$
$$H1: (\beta_0, \beta_1) \neq (0, 1)$$

Under the null hypothesis, the statistic is:

$$f = \frac{\left[(\hat{\beta}_0, \hat{\beta}_1) - (0,1)\right] \Sigma^{-1} \left[(\hat{\beta}_0, \hat{\beta}_1) - (0,1)\right]'}{2} \sim F(2, dfErr)$$

For this example:

$$f = 200.5754$$

The result for the F-test is given below.

Simultaneous Coefficient Test		
Test	**F Value**	**P Value**
β_0= 0 and β_1= 1	200.5754	3.43E-08

Since the p value is almost 0 in the above table, we have enough evidence to reject the null hypothesis. Therefore, these two gages have different accuracy.

This example shows that the paired t-test and the regression coefficient test give different conclusion. This is because the t-test cannot catch the difference between the linearity of these two gages, while the simultaneous regression coefficient test can.

Precision Agreement Study

A gage agreement experiment should be conducted by the same operator, so the gage reproducibility caused by operator is removed. Only repeatability caused by gages is calculated and compared. Therefore precision agreement study is comparing the repeatability of each gage.

Let's use the first example in the above accuracy agreement study for a precision agreement study.

First, we need to calculate the repeatability of each gage. Repeatability is also the pure error which is the variation of the multiple readings for the same part by the same operator. The result of Gage 2 is given in the following table.

Subject	1st Reading	2nd Reading	Sum of Square (SS)
1	512	525	84.5
2	430	415	112.5
3	520	508	72
4	428	444	128
5	500	500	0
6	600	625	312.5
7	364	460	4608
8	380	390	50
9	658	642	128
10	445	432	84.5
11	432	420	72
12	626	605	220.5
13	260	227	544.5
14	477	467	50
15	259	268	40.5
16	350	370	200
17	451	443	32
		Total SS	6739.5
		Repeatability	396.4412

The repeatability is calculated by the following steps.

Step 1: For each subject, calculate the sum of square (SS) of the repeated readings for the same gage. For example, for subject 1, the SS under this gage is:

$$SS_i = \sum_{i=1}^{n_i} (x_i - \bar{x})^2 = \left[(512 - 518.5)^2 + (525 - 518.5)^2\right] = 84.5$$

Step 2: Add the SS of all the subjects together.

$$SS = \sum_{i=1} SS_i$$

Step 3: Find the degree of freedom.

$$df = \sum_{i=1}^{n} (n_i - 1)$$

n_i is the number of repeated reading for subject i. n is the total number of subjects.

Step 4: Calculate the variance (repeatability). For Gage 2, it is:

$$\hat{s}_2^2 = \frac{SS}{df} = 396.4412$$

Repeating the above procedure, we can get the repeatability for gage 1. It is 234.2941. We can then compare the repeatability of these two gages. If these two variances are the same, then the ratio of them follows an F distribution with degree of freedom of df_1 and df_2. df_1 is the degree of freedom for Gage 1 (the numerator in the F ratio) and Gage 2 (the denominator in the F ratio).

The results are:

Gage	Repeatability Variance	Degrees of Freedom	F Ratio	Lower Bound	Upper Bound	P Value
Gage 1	234.2941	17	0.591	0.22101	1.5799	0.1440
Gage 2	396.4412	17				

The p value in the range of (risk level)/2 = 0.025 and 1-(risk level)/2 = 0.975. Therefore, we cannot reject the null hypothesis that these two gages have the same precision.

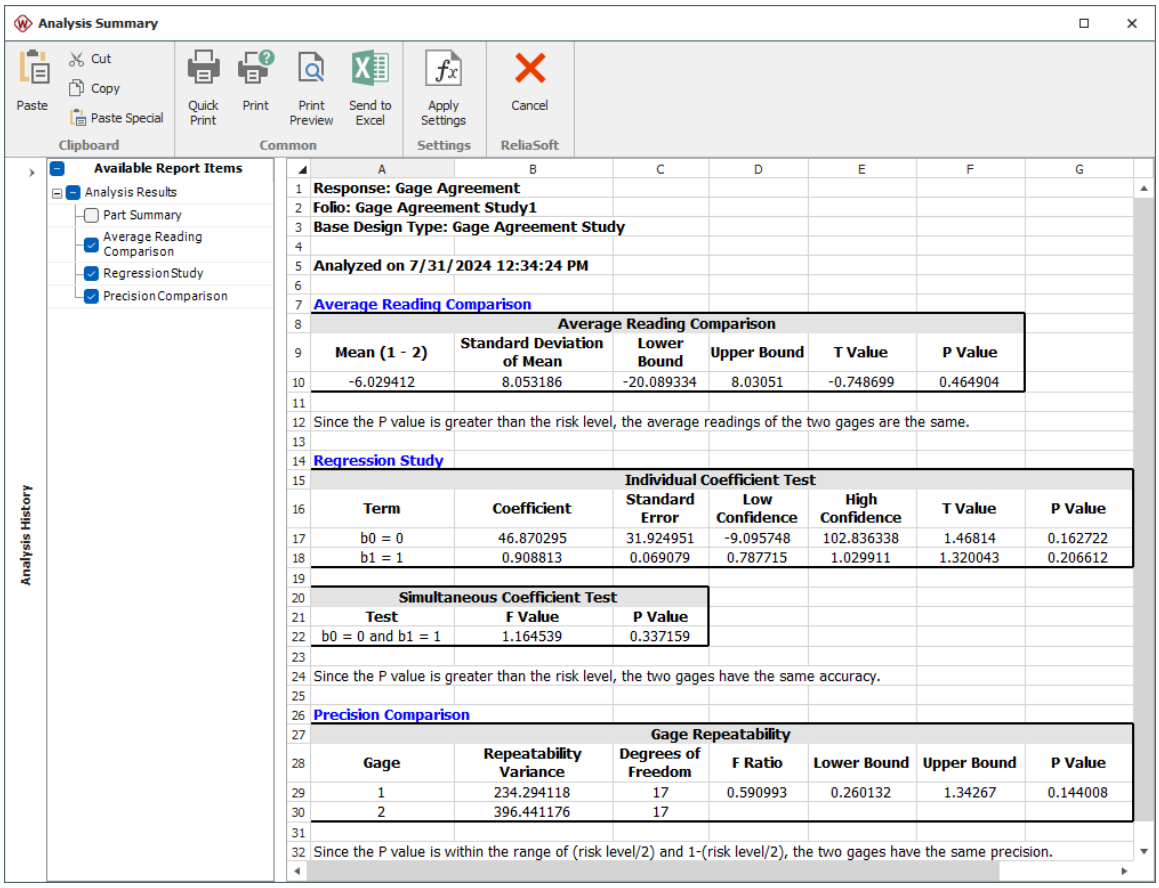
The bounds in the above table are calculated by:

$$f_U = invF(df_1, df_2, \alpha/2)$$

$$f_L = invF(df_1, df_2, 1 - \alpha/2)$$

$$\text{Upper bound} = F_{ratio}/f_L$$
$$\text{Lower bound} = F_{ratio}/f_U$$

For this example $f_U = 2.6733$ and $f_L = 0.374069$. Therefore, the upper bound is 1.5799 and the lower bound is 0.22101. Since the bounds include 1, it means the two gages have the same repeatability. The results from the DOE folio are given below.



Response: Gage Agreement
Folio: Gage Agreement Study1
Base Design Type: Gage Agreement Study

Analyzed on 7/31/2024 12:34:24 PM

Average Reading Comparison

Average Reading Comparison					
Mean (1 - 2)	Standard Deviation of Mean	Lower Bound	Upper Bound	T Value	P Value
-6.029412	8.053186	-20.089334	8.03051	-0.748699	0.464904

Since the P value is greater than the risk level, the average readings of the two gages are the same.

Regression Study

Individual Coefficient Test						
Term	Coefficient	Standard Error	Low Confidence	High Confidence	T Value	P Value
b0 = 0	46.870295	31.924951	-9.095748	102.836338	1.46814	0.162722
b1 = 1	0.908813	0.069079	0.787715	1.029911	1.320043	0.206612

Simultaneous Coefficient Test		
Test	F Value	P Value
b0 = 0 and b1 = 1	1.164539	0.337159

Since the P value is greater than the risk level, the two gages have the same accuracy.

Precision Comparison

Gage Repeatability						
Gage	Repeatability Variance	Degrees of Freedom	F Ratio	Lower Bound	Upper Bound	P Value
1	234.294118	17	0.590993	0.260132	1.34267	0.144008
2	396.441176	17				

Since the P value is within the range of (risk level/2) and 1-(risk level/2), the two gages have the same precision.

General Guidelines on Measurement System Analysis

The experiments for MSA should be designed experiments. The experiment should be designed and conducted based on DOE principals. Here are some of the guidelines for preparation prior to conducting MSA [AIAG].

1. Whenever possible, the operators chosen should be selected from those who normally operate the gage. If these operators are not available, then personnel should be properly trained in the correct usage of the gage.

2. The sample parts must be selected from the process which represents its entire operating range. This is sometimes done by taking one sample per day for several days. The collected samples will be treated as if they represent the full range of product variation. Each part must be numbered for identification.

3. The gage must have a graduation that allows at least one-tenth of the expected process variation of the characteristic to be read directly. Process variation is usually defined as 6 times the process standard deviation. For example, if the process variation is 0.1, the equipment should read directly to an increment no larger than 0.01.

The manner in which a study is conducted is very important if reliable results are to be obtained. To minimize the possibility of getting inaccurate results, the following steps are suggested:

1. The measurements should be made in a random order. The operators should be unaware of which numbered part is being checked in order to avoid any possible bias. However, the person conducting the study should know which numbered part is being checked and record the data accordingly, such as Operator A, Part 1, first trial.

2. In reading the gage, the readings should be estimated to the nearest number that can be obtained. At a minimum, readings should be made to one-half of the smallest graduation. For example, if the smallest graduation is 0.01, then the estimate for each reading should be rounded to the nearest 0.005.

Appendices

IN THIS CHAPTER

ANOVA Calculations in Multiple Linear Regression **439**
Total Sum of Squares 440
Model Sum of Squares 441
Error Sum of Squares 441
Mean Squares 441
Calculation of the Statistic F_0 442

Use of Regression to Calculate Sum of Squares **442**

Plackett-Burman Designs **445**
12-Run Design 445
20-Run Design 445
24-Run Design 446

Taguchi Orthogonal Arrays **446**
Two Level Designs 446
Three Level Designs 448
Mixed Level Designs 449

Alias Relations for Taguchi Orthogonal Arrays **452**

Box-Behnken Designs **455**

DOE Glossary **455**

DOE References **461**

ANOVA Calculations in Multiple Linear Regression

The sum of squares for the analysis of variance in multiple linear regression is obtained using the same relations as those in simple linear regression, except that the matrix notation is

preferred in the case of multiple linear regression. In the case of both the simple and multiple linear regression models, once the observed and fitted values are available, the sum of squares are calculated in an identical manner. The difference between the two models lies in the way the fitted values are obtained. In a simple linear regression model, the fitted values are obtained from a model having only one predictor variable. In multiple linear regression analysis, the model used to obtained the fitted values contains more than one predictor variable.

Total Sum of Squares

Recall from Simple Linear Regression Analysis that the total sum of squares, SS_r , is obtained using the following equation:

$$\begin{aligned} SS_r &= \sum_{i=1}^{n}(y_i - \bar{y})^2 \\ &= \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} \end{aligned}$$

The first term, $\sum_{i=1}^{n} y_i^2$, can be expressed in matrix notation using the vector of observed values, **y**, as:

$$\sum_{i=1}^{n} y_i^2 = y'y$$

If **J** represents an *n x n* square matrix of ones, then the second term, $\left(\sum_{i=1}^{n} y_i\right)^2 / n$, can be expressed in matrix notation as:

$$\frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} = \left(\frac{1}{n}\right) y'Jy$$

Therefore, the total sum of squares in matrix notation is:

$$\begin{aligned} SS_r &= y'y - \left(\frac{1}{n}\right) y'Jy \\ &= y'\left[I - \left(\frac{1}{n}\right) J\right] y \end{aligned}$$

where **I** is the identity matrix of order n .

Model Sum of Squares

Similarly, the model sum of squares or the regression sum of squares, SS_R, can be obtained in matrix notation as:

$$\begin{aligned} SS_R &= \sum_{i=1}^{n} \hat{y}_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n} \\ &= \hat{y}'\hat{y} - (\frac{1}{n}) y' J y \\ & y'[H - (\frac{1}{n})J] y \end{aligned}$$

where H is the hat matrix and is calculated using $H = X(X'X)^{-1}X'$.

Error Sum of Squares

The error sum of squares or the residual sum of squares, SS_E, is obtained in the matrix notation from the vector of residuals, e, as:

$$\begin{aligned} SS_E &= e'e \\ &= (y - \hat{y})'(y - \hat{y}) \\ &= y'(I - H)y \end{aligned}$$

Mean Squares

Mean squares are obtained by dividing the sum of squares with their associated degrees of freedom. The number of degrees of freedom associated with the total sum of squares, SS_r, is ($n-1$) since there are *n* observations in all, but one degree of freedom is lost in the calculation of the sample mean, $\bar{y}$. The total mean square is:

$$MS_r = \frac{SS_r}{n-1}$$

The number of degrees of freedom associated with the regression sum of squares, SS_R, is k. There are (k+1) degrees of freedom associated with a regression model with (k+1) coefficients, β_0, β_1, β_2 .... β_k. However, one degree of freedom is lost because the deviations, ($\hat{y}_i - \bar{y}$), are subjected to the constraints that they must sum to zero ($\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = 0$). The regression mean square is:

$$MS_R = \frac{SS_R}{k}$$

The number of degrees of freedom associated with the error sum of squares is $n-(k+1)$, as there are n observations in all, but $k+1$ degrees of freedom are lost in obtaining the estimates of β_0, β_1, β_2 ...β_k to calculate the predicted values, $\hat{y}_i$. The error mean square is:

$$MS_E = \frac{SS_E}{n-(k+1)}$$

The error mean square, MS_E, is an estimate of the variance, σ^2, of the random error terms, ϵ_i.

$$\hat{\sigma}^2 = MS_E$$

Calculation of the Statistic F_0

Once the mean squares MS_R and MS_E are known, the statistic to test the significance of regression can be calculated as follows:

$$F_0 = \frac{MS_R}{MS_E}$$

Use of Regression to Calculate Sum of Squares

This appendix explains the reason behind the use of regression in Weibull++ DOE folios in all calculations related to the sum of squares. A number of textbooks present the method of direct summation to calculate the sum of squares. But this method is only applicable for balanced designs and may give incorrect results for unbalanced designs. For example, the sum of squares for factor A in a balanced factorial experiment with two factors, A and B, is given as follows:

$$\begin{aligned} SS_A &= \sum_{i=1}^{n_a} n_b n(\bar{y}_{i..} - \bar{y}_{...})^2 \\ &= \sum_{i=1}^{n_a} \frac{y_{i..}^2}{n_b n} - \frac{y_{...}^2}{n_a n_b n} \end{aligned}$$

where n_a represents the levels of factor A, n_b represents the levels of factor B, and n represents the number of samples for each combination of A and B. The term $\bar{y}_{i..}$ is the mean value for the ith level of factor A, $y_{i..}$ is the sum of all observations at the ith level of factor A and $y_{...}$ is the sum of all observations.

The analogous term to calculate SS_A in the case of an unbalanced design is given as:

$$SS_A = \sum_{i=1}^{n_a} \frac{y_{i..}^2}{n_{i.}} - \frac{y_{...}^2}{n_{..}}$$

where $n_{i.}$ is the number of observations at the ith level of factor A and $n_{..}$ is the total number of observations. Similarly, to calculate the sum of squares for factor B and interaction AB, the formulas are given as:

$$SS_B = \sum_{j=1}^{n_b} \frac{y_{.j.}^2}{n_{.j}} - \frac{y_{...}^2}{n_{..}}$$

$$SS_{AB} = \sum_{i=1}^{n_a}\sum_{j=1}^{n_b} \frac{y_{ij.}^2}{n_{ij}} - \sum_{i=1}^{n_a} \frac{y_{i..}^2}{n_{i.}} - \sum_{j=1}^{n_b} \frac{y_{.j.}^2}{n_{.j}} + \frac{y_{...}^2}{n_{..}}$$

Applying these relations to the unbalanced data of the last table, the sum of squares for the interaction AB is:

$$\begin{aligned} SS_{AB} &= \sum_{i=1}^{n_a}\sum_{j=1}^{n_b} \frac{y_{ij.}^2}{n_{ij}} - \sum_{i=1}^{n_a} \frac{y_{i..}^2}{n_{i.}} - \sum_{j=1}^{n_b} \frac{y_{.j.}^2}{n_{.j}} + \frac{y_{...}^2}{n_{..}} \\ &= \left(6^2 + 4^2 + \frac{(42+6)^2}{2} + 12^2\right) - \left(\frac{10^2}{2} + \frac{60^2}{3}\right) \\ &\quad - \left(\frac{54^2}{3} + \frac{16^2}{2}\right) + \frac{70^2}{5} \\ &= -22 \end{aligned}$$

which is obviously incorrect since the sum of squares cannot be negative. For a detailed discussion on this refer to Searle (1997, 1971).

Factor A	Factor B		Row
	Level 1	Level 2	Total
Level 1	6	4	10
Level 2	6, 42	12	60
Column Total	54	16	70

The correct sum of squares can be calculated as shown next. The y and X matrices for the design of the last table can be written as:

$$y = \begin{bmatrix} 6 \\ 4 \\ 6 \\ 12 \\ 42 \end{bmatrix} \text{ and } X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix}$$

Then the sum of squares for the interaction AB can be calculated as:

$$SS_{AB} = y'[H - (1/5)J]y - y'[H_{-AB} - (1/5)J]y$$

where H is the hat matrix and J is the matrix of ones. The matrix $H_{\sim AB}$ can be calculated using $H_{\sim AB} = X_{\sim AB}(X'_{\sim AB}X_{\sim AB})^{-1}X'_{\sim AB}$ where $X_{\sim AB}$ is the design matrix, X, excluding the last column that represents the interaction effect AB. Thus, the sum of squares for the interaction AB is:

$$\begin{aligned} SS_{AB} = & \quad y'[H-(1/5)J]y - y'[H_{\sim AB}-(1/5)J]y \\ = & \quad 368 - 339.4286 \\ = & \quad 28.5714 \end{aligned}$$

This is the value that is calculated by the DOE folio (see the first figure below, for the experiment design and the second figure below for the analysis).

B1 | 1

	Standard Order	A:Factor 1	B:Factor 2	Response 1
1	1	1	1	6
2	2	1	2	4
3	3	2	1	6
4	4	2	2	12
5	5	2	1	42

	A	B	C	D	E	F
7	ANOVA Table					
8	ANOVA Table					
9	**Source of Variation**	**Degrees of Freedom**	**Sum of Squares [Partial]**	**Mean Squares [Partial]**	**F Ratio**	**P Value**
10	Model	3	368	122.666667	0.1893	0.894857
11	A:Factor 1	1	193.142857	193.142857	0.29806	0.681864
12	B:Factor 2	1	56	56	0.08642	0.817979
13	A • B	1	28.571429	28.571429	0.044092	0.868237
14	Residual	1	648	648		
15	Pure Error	1	648	648		
16	Total	4	1016			

Plackett-Burman Designs

12-Run Design

Run	Factors										
	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*J*	*K*	*L*
1	1	1	-1	1	1	1	-1	-1	-1	1	-1
2	-1	1	1	-1	1	1	1	-1	-1	-1	1
3	1	-1	1	1	-1	1	1	1	-1	-1	-1
4	-1	1	-1	1	1	-1	1	1	1	-1	-1
5	-1	-1	1	-1	1	1	-1	1	1	1	-1
6	-1	-1	-1	1	-1	1	1	-1	1	1	1
7	1	-1	-1	-1	1	-1	1	1	-1	1	1
8	1	1	-1	-1	-1	1	-1	1	1	-1	1
9	1	1	1	-1	-1	-1	1	-1	1	1	-1
10	-1	1	1	1	-1	-1	-1	1	-1	1	1
11	1	-1	1	1	1	-1	-1	-1	1	-1	1
12	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

20-Run Design

Run	Factors																		
	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*J*	*K*	*L*	*M*	*N*	*O*	*P*	*Q*	*R*	*S*	*T*
1	1	1	-1	-1	1	1	1	1	-1	1	-1	1	-1	-1	-1	-1	1	1	-1
2	-1	1	1	-1	-1	1	1	1	1	-1	1	-1	1	-1	-1	-1	-1	1	1
3	1	-1	1	1	-1	-1	1	1	1	1	-1	1	-1	1	-1	-1	-1	-1	1
4	1	1	-1	1	1	-1	-1	1	1	1	1	-1	1	-1	1	-1	-1	-1	-1
5	-1	1	1	-1	1	1	-1	-1	1	1	1	1	-1	1	-1	1	-1	-1	-1
6	-1	-1	1	1	-1	1	1	-1	-1	1	1	1	1	-1	1	-1	1	-1	-1
7	-1	-1	-1	1	1	-1	1	1	-1	-1	1	1	1	1	-1	1	-1	1	-1
8	-1	-1	-1	-1	1	1	-1	1	1	-1	-1	1	1	1	1	-1	1	-1	1
9	1	-1	-1	-1	-1	1	1	-1	1	1	-1	-1	1	1	1	1	-1	1	-1
10	-1	1	-1	-1	-1	-1	1	1	-1	1	1	-1	-1	1	1	1	1	-1	1
11	1	-1	1	-1	-1	-1	-1	1	1	-1	1	1	-1	-1	1	1	1	1	-1
12	-1	1	-1	1	-1	-1	-1	-1	1	1	-1	1	1	-1	-1	1	1	1	1
13	1	-1	1	-1	1	-1	-1	-1	-1	1	1	-1	1	1	-1	-1	1	1	1
14	1	1	-1	1	-1	1	-1	-1	-1	-1	1	1	-1	1	1	-1	-1	1	1
15	1	1	1	-1	1	-1	1	-1	-1	-1	-1	1	1	-1	1	1	-1	-1	1
16	1	1	1	1	-1	1	-1	1	-1	-1	-1	-1	1	1	-1	1	1	-1	-1
17	-1	1	1	1	1	-1	1	-1	1	-1	-1	-1	-1	1	1	-1	1	1	-1
18	-1	-1	1	1	1	1	-1	1	-1	1	-1	-1	-1	-1	1	1	-1	1	1
19	1	-1	-1	1	1	1	1	-1	1	-1	1	-1	-1	-1	-1	1	1	-1	1
20	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

24-Run Design

Run	Factors																					
	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*J*	*K*	*L*	*M*	*N*	*O*	*P*	*Q*	*R*	*S*	*T*	*U*	*V*	*W*	*X*
1	1	1	1	1	1	-1	1	-1	1	1	-1	-1	1	1	-1	-1	1	-1	1	-1	-1	-1	-1
2	-1	1	1	1	1	1	-1	1	-1	1	1	-1	-1	1	1	-1	-1	1	-1	1	-1	-1	-1
3	-1	-1	1	1	1	1	1	-1	1	-1	1	1	-1	-1	1	1	-1	-1	1	-1	1	-1	-1
4	-1	-1	-1	1	1	1	1	1	-1	1	-1	1	1	-1	-1	1	1	-1	-1	1	-1	1	-1
5	-1	-1	-1	-1	1	1	1	1	1	-1	1	-1	1	1	-1	-1	1	1	-1	-1	1	-1	1
6	1	-1	-1	-1	-1	1	1	1	1	1	-1	1	-1	1	1	-1	-1	1	1	-1	-1	1	-1
7	-1	1	-1	-1	-1	-1	1	1	1	1	1	-1	1	-1	1	1	-1	-1	1	1	-1	-1	1
8	1	-1	1	-1	-1	-1	-1	1	1	1	1	1	-1	1	-1	1	1	-1	-1	1	1	-1	-1
9	-1	1	-1	1	-1	-1	-1	-1	1	1	1	1	1	-1	1	-1	1	1	-1	-1	1	1	-1
10	-1	-1	1	-1	1	-1	-1	-1	-1	1	1	1	1	1	-1	1	-1	1	1	-1	-1	1	1
11	1	-1	-1	1	-1	1	-1	-1	-1	-1	1	1	1	1	1	-1	1	-1	1	1	-1	-1	1
12	1	1	-1	-1	1	-1	1	-1	-1	-1	-1	1	1	1	1	1	-1	1	-1	1	1	-1	-1
13	-1	1	1	-1	-1	1	-1	1	-1	-1	-1	-1	1	1	1	1	1	-1	1	-1	1	1	-1
14	-1	-1	1	1	-1	-1	1	-1	1	-1	-1	-1	-1	1	1	1	1	1	-1	1	-1	1	1
15	1	-1	-1	1	1	-1	-1	1	-1	1	-1	-1	-1	-1	1	1	1	1	1	-1	1	-1	1
16	1	1	-1	-1	1	1	-1	-1	1	-1	1	-1	-1	-1	-1	1	1	1	1	1	-1	1	-1
17	-1	1	1	-1	-1	1	1	-1	-1	1	-1	1	-1	-1	-1	-1	1	1	1	1	1	-1	1
18	1	-1	1	1	-1	-1	1	1	-1	-1	1	-1	1	-1	-1	-1	-1	1	1	1	1	1	-1
19	-1	1	-1	1	1	-1	-1	1	1	-1	-1	1	-1	1	-1	-1	-1	-1	1	1	1	1	1
20	1	-1	1	-1	1	1	-1	-1	1	1	-1	-1	1	-1	1	-1	-1	-1	-1	1	1	1	1
21	1	1	-1	1	-1	1	1	-1	-1	1	1	-1	-1	1	-1	1	-1	-1	-1	-1	1	1	1
22	1	1	1	-1	1	-1	1	1	-1	-1	1	1	-1	-1	1	-1	1	-1	-1	-1	-1	1	1
23	1	1	1	1	-1	1	-1	1	1	-1	-1	1	1	-1	-1	1	-1	1	-1	-1	-1	-1	1
24	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

Taguchi Orthogonal Arrays

Two Level Designs

L4 (2^3)

Run	Columns		
	1	2	3
1	1	1	1
2	1	2	2
3	2	1	2
4	2	2	1

L8 (2^7)

Run	Columns						
	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2

L12 (2^11)

Run	Columns										
	1	2	3	4	5	6	7	8	9	10	11
1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	2	2	2	2	2	2
3	1	1	2	2	2	1	1	1	2	2	2
4	1	2	1	2	2	1	2	2	1	1	2
5	1	2	2	1	2	2	1	2	1	2	1
6	1	2	2	2	1	2	2	1	2	1	1
7	2	1	2	2	1	1	2	2	1	2	1
8	2	1	2	1	2	2	2	1	1	1	2
9	2	1	1	2	2	2	1	2	2	1	1
10	2	2	2	1	1	1	1	2	2	1	2
11	2	2	1	2	1	2	1	1	1	2	2
12	2	2	1	1	2	1	2	1	2	2	1

L16 (2^15)

Run	Columns														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2
3	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2
4	1	1	1	2	2	2	2	2	2	2	2	1	1	1	1
5	1	2	2	1	1	2	2	1	1	2	2	1	1	2	2
6	1	2	2	1	1	2	2	2	2	1	1	2	2	1	1
7	1	2	2	2	2	1	1	1	1	2	2	2	2	1	1
8	1	2	2	2	2	1	1	2	2	1	1	1	1	2	2
9	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
10	2	1	2	1	2	1	2	2	1	2	1	2	1	2	1
11	2	1	2	2	1	2	1	1	2	1	2	2	1	2	1
12	2	1	2	2	1	2	1	2	1	2	1	1	2	1	2
13	2	2	1	1	2	2	1	1	2	2	1	1	2	2	1
14	2	2	1	1	2	2	1	2	1	1	2	2	1	1	2
15	2	2	1	2	1	1	2	1	2	2	1	2	1	1	2
16	2	2	1	2	1	1	2	2	1	1	2	1	2	2	1

Three Level Designs

L9 (3^4)

Run	Columns			
	1	2	3	4
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	2	1	2	3
5	2	2	3	1
6	2	3	1	2
7	3	1	3	2
8	3	2	1	3
9	3	3	2	1

L27 (3^13)

Run	Columns												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	2	2	2	2	2	2	2	2	2
3	1	1	1	1	3	3	3	3	3	3	3	3	3
4	1	2	2	2	1	1	1	2	2	2	3	3	3
5	1	2	2	2	2	2	2	3	3	3	1	1	1
6	1	2	2	2	3	3	3	1	1	1	2	2	2
7	1	3	3	3	1	1	1	3	3	3	2	2	2
8	1	3	3	3	2	2	2	1	1	1	3	3	3
9	1	3	3	3	3	3	3	2	2	2	1	1	1
10	2	1	2	3	1	2	3	1	2	3	1	2	3
11	2	1	2	3	2	3	1	2	3	1	2	3	1
12	2	1	2	3	3	1	2	3	1	2	3	1	2
13	2	2	3	1	1	2	3	2	3	1	3	1	2
14	2	2	3	1	2	3	1	3	1	2	1	2	3
15	2	2	3	1	3	1	2	1	2	3	2	3	1
16	2	3	1	2	1	2	3	3	1	2	2	3	1
17	2	3	1	2	2	3	1	1	2	3	3	1	2
18	2	3	1	2	3	1	2	2	3	1	1	2	3
19	3	1	3	2	1	3	2	1	3	2	1	3	2
20	3	1	3	2	2	1	3	2	1	3	2	1	3
21	3	1	3	2	3	2	1	3	2	1	3	2	1
22	3	2	1	3	1	3	2	2	1	3	3	2	1
23	3	2	1	3	2	1	3	3	2	1	1	3	2
24	3	2	1	3	3	2	1	1	3	2	2	1	3
25	3	3	2	1	1	3	2	3	2	1	2	1	3
26	3	3	2	1	2	1	3	1	3	2	3	2	1
27	3	3	2	1	3	2	1	2	1	3	1	3	2

Mixed Level Designs

L8 (2^4 4^1)

Run	Columns				
	1	2	3	4	5
1	1	1	1	1	1
2	2	2	2	2	1
3	1	1	2	2	2
4	2	2	1	1	2
5	1	2	1	2	3
6	2	1	2	1	3
7	1	2	2	1	4
8	2	1	1	2	4

L16 (2^12 4^1)

Run	Columns												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	2	2	2	2	2	2	2	2	1
3	2	2	2	2	1	1	1	1	2	2	2	2	1
4	2	2	2	2	2	2	2	2	1	1	1	1	1
5	1	1	2	2	1	1	2	2	1	1	2	2	2
6	1	1	2	2	2	2	1	1	2	2	1	1	2
7	2	2	1	1	1	1	2	2	2	2	1	1	2
8	2	2	1	1	2	2	1	1	1	1	2	2	2
9	1	2	1	2	1	2	1	2	1	2	1	2	3
10	1	2	1	2	2	1	2	1	2	1	2	1	3
11	2	1	2	1	1	2	1	2	2	1	2	1	3
12	2	1	2	1	2	1	2	1	1	2	1	2	3
13	1	2	2	1	1	2	2	1	1	2	2	1	4
14	1	2	2	1	2	1	1	2	2	1	1	2	4
15	2	1	1	2	1	2	2	1	2	1	1	2	4
16	2	1	1	2	2	1	1	2	1	2	2	1	4

L16 (2^9 4^2)

Run	Columns										
	1	2	3	4	5	6	7	8	9	10	11
1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2	2	2	1	2
3	2	2	2	1	1	1	2	2	2	1	3
4	2	2	2	2	2	2	1	1	1	1	4
5	1	2	2	1	2	2	1	2	2	2	1
6	1	2	2	2	1	1	2	1	1	2	2
7	2	1	1	1	2	2	2	1	1	2	3
8	2	1	1	2	1	1	1	2	2	2	4
9	2	1	2	2	1	2	2	1	2	3	1
10	2	1	2	1	2	1	1	2	1	3	2
11	1	2	1	2	1	2	1	2	1	3	3
12	1	2	1	1	2	1	2	1	2	3	4
13	2	2	1	2	2	1	2	2	1	4	1
14	2	2	1	1	1	2	1	1	2	4	2
15	1	1	2	2	2	1	1	1	2	4	3
16	1	1	2	1	1	2	2	2	1	4	4

L16 (2^6 4^3)

Run	Columns								
	1	2	3	4	5	6	7	8	9
1	1	1	1	1	1	1	1	1	1
2	1	1	2	2	2	2	1	2	2
3	2	2	1	1	2	2	1	3	3
4	2	2	2	2	1	1	1	4	4
5	2	2	1	2	1	2	2	1	2
6	2	2	2	1	2	1	2	2	1
7	1	1	1	2	2	1	2	3	4
8	1	1	2	1	1	2	2	4	3
9	1	2	2	2	2	1	3	1	3
10	1	2	1	1	1	2	3	2	4
11	2	1	2	2	1	2	3	3	1
12	2	1	1	1	2	1	3	4	2
13	2	1	2	1	2	2	4	1	4
14	2	1	1	2	1	1	4	2	3
15	1	2	2	1	1	1	4	3	2
16	1	2	1	2	2	2	4	4	1

L16 (2^3 4^4)

Run	Columns						
	1	2	3	4	5	6	7
1	1	1	1	1	1	1	1
2	1	2	2	1	2	2	2
3	2	1	2	1	3	3	3
4	2	2	1	1	4	4	4
5	2	2	1	2	1	2	1
6	2	1	2	2	2	1	2
7	1	2	2	2	3	4	4
8	1	1	1	2	4	3	3
9	1	2	2	3	1	3	4
10	1	1	1	3	2	4	3
11	2	2	1	3	3	1	2
12	2	1	2	3	4	2	1
13	2	1	2	4	1	4	2
14	2	2	1	4	2	3	1
15	1	1	1	4	3	2	4
16	1	2	2	4	4	1	3

L18 (2^1 3^7)

Run	Columns							
	1	2	3	4	5	6	7	8
1	1	1	1	1	1	1	1	1
2	1	1	2	2	2	2	2	2
3	1	1	3	3	3	3	3	3
4	1	2	1	1	2	2	3	3
5	1	2	2	2	3	3	1	1
6	1	2	3	3	1	1	2	2
7	1	3	1	2	1	3	2	3
8	1	3	2	3	2	1	3	1
9	1	3	3	1	3	2	1	2
10	2	1	1	3	3	2	2	1
11	2	1	2	1	1	3	3	2
12	2	1	3	2	2	1	1	3
13	2	2	1	2	3	1	3	2
14	2	2	2	3	1	2	1	3
15	2	2	3	1	2	3	2	1
16	2	3	1	3	2	3	1	2
17	2	3	2	1	3	1	2	3
18	2	3	3	2	1	2	3	1

Alias Relations for Taguchi Orthogonal Arrays

For L8 (2^7):

	1	2	3	4	5	6	7
1		3	2	5	4	7	6
2			1	6	7	4	5
3				7	6	5	4
4					1	2	3
5						3	2
6							1

For L16 (2^15):

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1		3	2	5	4	7	6	9	8	11	10	13	12	15	14
2			1	6	7	4	5	10	11	8	9	14	15	12	13
3				7	6	5	4	11	10	9	8	15	14	13	12
4					1	2	3	12	13	14	15	8	9	10	11
5						3	2	13	12	15	14	9	8	11	10
6							1	14	15	12	13	10	11	8	9
7								15	14	13	12	11	10	9	8
8									1	2	3	4	5	6	7
9										3	2	5	4	7	6
10											1	6	7	4	5
11												7	6	5	4
12													1	2	3
13														3	2
14															1

For L32 (2^31) - first 16 columns:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1		3	2	5	4	7	6	9	8	11	10	13	12	15	14	17
2			1	6	7	4	5	10	11	8	9	14	15	12	13	18
3				7	6	5	4	11	10	9	8	15	14	13	12	19
4					1	2	3	12	13	14	15	8	9	10	11	20
5						3	2	13	12	15	14	9	8	11	10	21
6							1	14	15	12	13	10	11	8	9	22
7								15	14	13	12	11	10	9	8	23
8									1	2	3	4	5	6	7	24
9										3	2	5	4	7	6	25
10											1	6	7	4	5	26
11												7	6	5	4	27
12													1	2	3	28
13														3	2	29
14															1	30
15																31

For L32 (2^31) - remaining columns:

	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	16	19	18	21	20	23	22	25	24	27	26	29	28	31	30
2	19	16	17	22	23	20	21	26	27	24	25	30	31	28	29
3	18	17	16	23	22	21	20	27	26	25	24	31	30	29	28
4	21	22	23	16	17	18	19	28	29	30	31	24	25	26	27
5	20	23	22	17	16	19	18	29	28	31	30	25	24	27	26
6	23	20	21	18	19	16	17	30	31	28	29	26	27	24	25
7	22	21	20	19	18	17	16	31	30	29	28	27	26	25	24
8	25	26	27	28	29	30	31	16	17	18	19	20	21	22	23
9	24	27	26	29	28	31	30	17	16	19	18	21	20	23	22
10	27	24	25	30	31	28	29	18	19	16	17	22	23	20	21
11	26	25	24	31	30	29	28	19	18	17	16	23	22	21	20
12	29	30	31	24	25	26	27	20	21	22	23	16	17	18	19
13	28	31	30	25	24	27	26	21	20	23	22	17	16	19	18
14	31	28	29	26	27	24	25	22	23	20	21	18	19	16	17
15	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16
16	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
17		3	2	5	4	7	6	9	8	11	10	13	12	15	14
18			1	6	7	4	5	10	11	8	9	14	15	12	13
19				7	6	5	4	11	10	9	8	15	14	13	12
20					1	2	3	12	13	14	15	8	9	10	11
21						3	2	13	12	15	14	9	8	11	10
22							1	14	15	12	13	10	11	8	9
23								15	14	13	12	11	10	9	8
24									1	2	3	4	5	6	7
25										3	2	5	4	7	6
26											1	6	7	4	5
27												7	6	5	4
28													1	2	3
29														3	2
30															1

Box-Behnken Designs

Factors	Runs (excluding center points)	Design Matrix
3	12	$\begin{bmatrix} \pm1 & \pm1 & 0 \\ \pm1 & 0 & \pm1 \\ 0 & \pm1 & \pm1 \\ 0 & 0 & 0 \end{bmatrix}$
4	24	$\left[\begin{array}{cccc} \pm1 & \pm1 & 0 & 0 \\ 0 & 0 & \pm1 & \pm1 \\ 0 & 0 & 0 & 0 \\ \hdashline \pm1 & 0 & 0 & \pm1 \\ 0 & \pm1 & \pm1 & 0 \\ 0 & 0 & 0 & 0 \\ \hdashline \pm1 & 0 & \pm1 & 0 \\ 0 & \pm1 & 0 & \pm1 \\ 0 & 0 & 0 & 0 \end{array}\right]$
5	40	$\left[\begin{array}{ccccc} \pm1 & \pm1 & 0 & 0 & 0 \\ 0 & 0 & \pm1 & \pm1 & 0 \\ 0 & \pm1 & 0 & 0 & \pm1 \\ \pm1 & 0 & \pm1 & 0 & 0 \\ 0 & 0 & 0 & \pm1 & \pm1 \\ 0 & 0 & 0 & 0 & 0 \\ \hdashline 0 & \pm1 & \pm1 & 0 & 0 \\ \pm1 & 0 & 0 & \pm1 & 0 \\ 0 & 0 & \pm1 & 0 & \pm1 \\ \pm1 & 0 & 0 & 0 & \pm1 \\ 0 & \pm1 & 0 & \pm1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$

This table indicates that all combinations of plus and minus levels are to be run. Dashed lines indicate how the design can be separated into blocks.

DOE Glossary

- **Alias**: Two or more effects are said to be aliased in an experiment if these effects cannot be distinguished from each other. This happens when the columns of the design matrix corresponding to these effects are identical. As a result, the aliased effects are estimated by the same linear combination of observations instead of each effect being estimated by a unique

combination.

- **ANOVA**: ANOVA is the acronym for Analysis of Variance. It refers to the procedure of splitting the variability of a data set to conduct various significance tests.
- **ANOVA Model**: The regression model where all factors are treated as qualitative factors. ANOVA models are used in the analysis of experiments to identify significant factors by investigating each level of the factors individually.
- **Balanced Design**: An experiment in which equal number of observations are taken for each treatment.
- **Blocking**: Separation of experiment runs based on the levels of a nuisance factor. Blocking is used to deal with known nuisance factors. You should block what you can and randomize what you cannot. *See also* Nuisance Factors, Randomization.
- **Center Point**: The experiment run that corresponds to the mid-level of all the factor ranges.
- **Coded Values**: The factor values that are such that the upper limit of the investigated range of the factor becomes +1 and the lower limit becomes -1. Using coded values makes the experiments with all factors at two levels orthogonal.
- **Confidence Interval**: A closed interval where a certain percentage of the population is likely to lie. For example, a 90% confidence interval with a lower limit of A and an upper limit of B implies that 90% of the population lies between the values of A and B.
- **Confounding**: Confounding occurs in a design when certain effects cannot be distinguished from the block effect. This happens when full factorial designs are run using incomplete blocks. In such designs the same linear combination of observations estimates the block effect and the confounded effects. *See also* Incomplete Blocks.
- **Contrast**: Any linear combination of two or more factor level means such that the coefficients in the combination add up to zero. The difference between the means at any two levels of a factor is an example of a contrast.
- **Control Factors**: The factors affecting the response that are easily manipulated and set by the operator. *See also* Noise Factors.
- **Cross Array Design**: The experiment design in which every treatment of the inner array is replicated for each run of the outer array. *See also* Inner Array, Outer Array.
- **Curvature Test**: The test that investigates if the relation between the response and the factors is linear by using center points. *See also* Center Point.

- **Defining Relation**: For two level fractional factorial experiments, the equation that is used to obtain the fraction from the full factorial experiment. The equation shows which of the columns of the design matrix in the fraction are identical to the first column. For example, the defining relation $I=ABC$ can be used to obtain a half-fraction of the two level full factorial experiment with three factors A, B and C. The effects used in the equation are called the *generators* or *words*.
- **Degrees of Freedom**: The number of independent observations made in excess of the unknowns.
- **Design Matrix**: The matrix whose columns correspond to the levels of the variables (and their interactions) at which observations are recorded.
- **Design Resolution**: The number of factors in the smallest word in a defining relation. Design resolution indicates the degree of aliasing in a fractional factorial design. *See also* Defining Relation, Word.
- **Error**: The natural variations that occur in a process, even when all the factors are maintained at the same level. *See also* Residual.
- **Error Sum of Squares**: The variation in the data not captured by the model. The error sum of squares is also called the residual sum of squares. *See also* Model Sum of Squares, Total Sum of Squares.
- **Extra Sum of Squares**: The increase in the model sum of squares when a term is added to the model.
- **Factorial Experiment**: The experiment in which all combinations of the factor levels are run.
- **Fractional Factorial Experiment**: The experiment where only a fraction of the combinations of the factor levels are run.
- **Factor**: The entity whose effect on the response is investigated in the experiment.
- **Fitted Value**: The estimate of an observation obtained using the model that has been fit to all the observations.
- **Fixed Effects Model**: The ANOVA model used in the experiments where only a limited number of the factor levels are of interest to the experimenter. *See also* Random Effects Model.

- **Full Model**: The model that includes all the main effects and their interactions. In Weibull++ DOE folios, a full model is the model that contains all the effects that are specified by the user. *See also* Reduced Model.
- **Generator**: See Word.
- **Hierarchical Model**: In Weibull++ DOE folios, a model is said to be hierarchical if, corresponding to every interaction, the main effects of the related factors are included in the model.
- **Incomplete Blocks**: Blocks that do not contain all the treatments of a factorial experiment.
- **Inner Array**: The experiment design used to investigate the control factors under Taguchi's philosophy to design a robust system. *See also* Robust System, Outer Array, Cross Array.
- **Interactions**: Interaction between factors means that the effect produced by a change in a factor on the response depends on the level of the other factor(s).
- **Lack-of-Fit Sum of Squares**: The portion of the error sum of squares that represents variation in the data not captured because of using a reduced model. *See also* Reduced Model, Pure Error Sum of Squares.
- **Least Squares Means**: The predicted mean response value for a given factor level while the remaining factors in the model are set to the coded value of zero.
- **Level**: The setting of a factor used in the experiment.
- **Main Effect**: The change in the response due to a change in the level of a factor.
- **Mean Square**: The sum of squares divided by the respective degrees of freedom.
- **Model Sum of Squares**: The portion of the total variability in the data that is explained by the model. *See also* Error Sum of Squares, Total Sum of Squares.
- **Multicollinearity**: A model with strong dependencies between the independent variables is said to have multicollinearity.
- **New Observations**: Observations that are not part of the data set used to fit the model.
- **Noise Factors**: Those nuisance factors that vary uncontrollably or naturally and can only be controlled for experimental purposes. For example, ambient temperature, atmospheric pressure and humidity are examples of noise factors.
- **Nuisance Factors**: Factors that have an effect on the response but are not of primary interest to the investigator.

- **Orthogonal Array**: An array in which all the columns are orthogonal to each other. Two columns are said to be orthogonal if the sum of the terms resulting from the product of the columns is zero.
- **Orthogonal Design**: An experiment design is orthogonal if the corresponding design matrix is such that the sum of the terms resulting from the product of any two columns is zero. In orthogonal designs the analysis of an effect does not depend on what other effects are included in the model.
- **Outer Array**: The experiment design used to investigate noise factors under Taguchi's philosophy to design a robust system. *See also* Robust System, Inner Array, Cross Array.
- **Partial Sum of Squares**: The type of extra sum of squares that is calculated assuming that all terms other than the given term are included in the model. The partial sum of squares is also referred to as the adjusted sum of squares. *See also* Extra Sum of Squares, Sequential Sum of Squares.
- **Prediction Interval**: The confidence interval on new observations.
- **Pure Error Sum of Squares**: The portion of the error sum of squares that represents variation due to replicates. *See also* Lack-of-Fit Sum of Squares.
- **Qualitative Factor**: The factor where the levels represent different categories and no numerical ordering is implied. These factors are also called categorical factors.
- **Random Effects Model**: The ANOVA model used in the experiments where the factor levels to be investigated are randomly selected from a large or infinite population. *See also* Fixed Effects Model.
- **Randomization**: Conducting experiment runs in a random order to cancel out the effect of unknown nuisance factors. *See also* Blocking.
- **Randomized Complete Block Design**: An experiment design where each block contains one replicate of the experiment and runs within the block are subjected to randomization.
- **Reduced Model**: A model that does not contain all the main effects and interactions. In Weibull++ DOE folios, a reduced model is the model that does not contain all the effects specified by the user. *See also* Full Model.
- **Regression Model**: A model that attempts to explain the relationship between two or more variables.
- **Repeated Runs**: Experiment runs corresponding to the same treatment that are conducted at the same time.

- **Replicated Runs**: Experiment runs corresponding to the same treatment that are conducted in a random order.
- **Residual**: An estimate of error which is obtained by calculating the difference between an observation and the corresponding fitted value. *See also* Error, Fitted Value.
- **Residual Sum of Squares**: See Error Sum of Squares.
- **Response**: The quantity that is investigated in an experiment to see which of the factors affect it.
- **Robust System**: A system that is insensitive to the effects of noise factors.
- **Rotatable Design**: A design is rotatable if the variance of the predicted response at any point depends only on the distance of the point from the design center point.
- **Screening Designs**: Experiments that use only a few runs to filter out important main effects and lower order interactions by assuming that higher order interactions are unimportant.
- **Sequential Sum of Squares**: The type of extra sum of squares that is calculated assuming that all terms preceding the given term are included in the model. *See also* Extra Sum of Squares, Partial Sum of Squares.
- **Signal to Noise Ratio**: The ratios defined by Taguchi to measure variation in the response caused by the noise factors.
- **Standard Order**: The order of the treatments such that factors are introduced one by one with each new factor being combined with the preceding terms.
- **Sum of Squares**: The quantity that is used to measure either a part or all of the variation in a data set.
- **Total Sum of Squares**: The sum of squares that represent all of the variation in a data set.
- **Transformation**: The mathematical function that makes the data follow a given characteristic. In the analysis of experiments transformation is used on the response data to make it follow the normal distribution.
- **Treatment**:The levels of a factor in a single factor experiment are also referred to as treatments. In experiments with many factors a combination of the levels of the factors is referred to as a treatment.
- **Word**: The effect used in the defining relation. For example, for the defining relation $I=ABC$, the word is ABC.

DOE References

1. AIAG (2010), *Measurement Systems Analysis (MSA)*, Automotive Industry Action Group, 4th edition.

2. Box, G. E. P., and Behnken, D. W. (1960), "Some New Three Level Designs for the Study of Quantitative Variables," *Technometrics*, Vol. 2, No. 4, pp. 455-475.

3. Box, G. E. P., and Draper, N. R. (1987), *Empirical Model Building and Response Surfaces*, John Wiley & Sons, Inc., New York.

4. Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978), *Statistics for Experimenters*, John Wiley & Sons, Inc., New York.

5. Cook, R. D. and Nachtsheim, C. J. (1980), "A Comparison of Algorithms for Constructing Exact D-Optimal Designs," *Technometrics*, vol. 22, no. 3, 315-324.

6. Derringer, G., and Suich, R. (1980), "Simultaneous Optimization of Several Response Variables," *Journal of Quality Technology*, Vol. 12, pp. 214-219.

7. Draper, N., and Smith H. (1998), *Applied Regression Analysis*, John Wiley & Sons, Inc., New York.

8. Dykstra, O. (1971), "The augmentation of experimental data to maximize |X'X|," *Technometrics*, vol. 13, no. 3, 682-688.

9. Fisher, R. A. (1966), *The Design of Experiments*, Hafner Publishing Company, New York.

10. Fedorov, V. V. (1972), "Theory of Optimal Experiments (Review)", *Biometrika*, cvol. 59, no. 3, 697-698. Translated and edited by W. J. Studden and E. M. Klimko.

11. Fries, A., and Hunter, W. G. (1980), "Minimum Aberration 2k-p Designs," *Technometrics*, Vol. 22, pp. 601-608.

12. Galil, Z. and Kiefer, J. (1980), "Time and Space Saving Computer Methods, Related to Mitchell's DETMAX, for Finding D-Optimal Designs", *Technometrics*, vol. 22, no. 3, 301-313.

13. Guo, H., Niu, P., and Szidarovszky, F. (2012), "A Simple Method for Power Calculation in Experiments for Treatment Comparison," *The IEEE International Conference on Industrial Engineering and Engineering Management*, Dec, 2012.

14. Hamada, M., and Balakrishnan N. (1998), "Analyzing Unreplicated Factorial Experiments: A Review with Some New Proposals," *Statistica Sinica*, Vol. 8, pp. 1-41.

15. Johnson, M. E. and Nachtsheim, C. J. (1983), “Some Guidelines for Constructing Exact D-Optimal Designs on Convex Design Spaces,” *Technometrics* , vol. 25, no. 3, 271-277.

16. Khuri, A. I., and Cornell, J. A. (1996), *Response Surfaces: Designs and Analyses*, Dekker, New York.

17. Kutner, M. H., Nachtsheim, C.J., Neter, J., and Li, W. (2005), *Applied Linear Statistical Models*, McGraw-Hill/Irwin, New York.

18. Lenth, R. V. (1989), "Quick and Easy Analysis of Unreplicated Factorials," *Technometrics*, Vol. 31, pp. 469-473.

19. Meeker, William Q., and Escobar, Luis A. (1998), *Statistical Methods for Reliability Data*, John Wiley & Sons, Inc., New York.

20. Montgomery, Douglas C. (2001), *Design and Analysis of Experiments*, John Wiley & Sons, Inc., New York.

21. Montgomery, Douglas C., and Peck, E. A. (1992), *Introduction to Linear Regression Analysis*, John Wiley & Sons, Inc., New York.

22. Montgomery, Douglas C., and Runger, George C. (1991), *Applied Statistics and Probability for Engineers*, John Wiley & Sons, Inc., New York.

23. Montgomery, Douglas C., and Runger, George C. (1993a), "Gauge capability analysis and designed experiments. Part I: Basic methods," *Quality Engineering*, 6, 1, 115-135.

24. Montgomery, Douglas C., and Runger, George C.(1993b), "Gauge capability analysis and designed experiments. Part II: Experimental design models and variance component estimation," *Quality Engineering*, 6, 2, 289-305.

25. Myers, R. H., and Montgomery, D. C. (1995), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, John Wiley & Sons, Inc., New York.

26. Plackett, R. L., and Burman, J. P. (1946), "The Design of Optimum Multifactorial Experiments," *Biometrika*, Vol. 33, No. 4, pp. 305-325.

27. ReliaSoft Corporation (2007a), *Accelerated Life Testing Reference*, ReliaSoft Publishing, Tucson, AZ.

28. ReliaSoft Corporation (2007b), *Life Data Analysis Reference*, ReliaSoft Publishing, Tucson, AZ.

29. Ross, S. (1987), *Introduction to Probability and Statistics for Engineers and Scientists*, John Wiley & Sons, Inc., New York.

30. Sahai, Hardeo, and Ageel, Mohammed I. (2000), *The Analysis of Variance*, Birkhauser, Boston.

31. Searle, S. R. (1997), *Linear Models*, John Wiley & Sons, Inc., New York.

32. Searle, S. R. (1971), "Topics in Variance Component Estimation," *Biometrics*, Vol. 27, No. 1, pp. 1-76.

33. Taguchi, G. (1991), *Introduction to Quality Engineering, Asian Productivity Organization*, UNIPUB, White Plains, New York.

34. Taguchi, G. (1987), *System of Experimental Design*, UNIPUB/Kraus International, White Plains, New York.

35. Taguchi, Genichi, Chowdhury, Subir, and Wu, Yuin. (2005), *Taguchi's Quality Engineering Handbook*, John Wiley & Sons, Inc., Hoboken, New Jersey.

36. Tukey, J. W. (1951), "Quick and Dirty Methods in Statistics, Part II, Simple Analysis for Standard Designs," *Proceedings of the Fifth Annual Convention*, American Society for Quality Control, pp. 189-197.

37. Wu, C. F. Jeff, and Hamada, Michael (2000), *Experiments: Planning, Analysis and Parameter Design Optimization*, John Wiley & Sons, Inc., New York.