Chapter 211

Balanced Design Analysis of Variance

Introduction

This procedure performs an analysis of variance on up to ten factors. The experimental design must be of the factorial type (no nested or repeated-measures factors) with no missing cells. If the data are balanced (equal-cell frequency), this procedure yields exact F-tests. If the data are not balanced, approximate F-tests are generated using the method of unweighted means (UWM).

The F-ratio is used to determine statistical significance. The tests are nondirectional in that the null hypothesis specifies that all means for a specified main effect or interaction are equal and the alternative hypothesis simply states that at least one is different.

Studies have shown that the properties of UWM F-tests are very good if the amount of unbalance in the cell frequencies is small. Despite that relative accuracy, you might well ask, "If the results are not always exact, why provide the method?" The answer is that the general linear models (GLM) solution (discussed in the General Linear Models chapter) sometimes requires more computer time and memory than is available. When there are several factors each with many levels, the GLM solution may not be obtainable. In these cases, UWM provides a very useful approximation. When the design is balanced, both procedures yield the same results, but the UWM method is much faster.

The procedure also calculates Friedman's two-way analysis of variance by ranks. This test is the nonparametric analog of the F-test in a randomized block design. (See Help File for details.)

Kinds of Research Questions

A large amount of research consists of studying the influence of a set of independent variables on a response (dependent) variable. Many experiments are designed to look at the influence of a single independent variable (factor) while holding other factors constant. These experiments are called single-factor experiments and are analyzed with the one-way analysis of variance (ANOVA). A second type of design considers the impact of one factor across several values of other factors. This experimental design is called the factorial design.

The factorial design is popular among researchers because it not only lets you study the individual effects of several factors in a single experiment, but it also lets you study their interaction. Interaction is present when the response variable fails to behave the same at values of one factor when a second factor is varied. Since factors seldom work independently, the study of their interaction becomes very important.

The Linear Model

We begin with an infinite population of individuals with many measurable characteristics. These individuals are separated into two or more treatment populations based on one or more of these characteristics. A random sample of the individuals in each population is drawn. A treatment is applied to each individual in the sample and an outcome is measured. The resulting data are analyzed using an analysis of variance table, which produces an F-test.

A mathematical model may be formulated that underlies each analysis of variance. This model expresses the response variable as the sum of parameters of the population. For example, a linear mathematical model for a two-factor experiment is

$$Y_{ijk} = m + a_i + b_j + (ab)_{ij} + e_{ijk}$$

where i = 1, 2, ..., I; j = 1, 2, ..., J; and k = 1, 2, ..., K. This model expresses the value of the response variable, Y, as the sum of five components:

m the mean.

a_i the contribution of the ith level of a factor A.

b_j the contribution of the jth level of a factor B.

$(ab)_{ij}$ the combined contribution of the ith level of a factor A and the jth level of a factor B.

e_{ijk} the contribution of the kth individual. This is often called the "error."

Note that this model is the sum of various constants. This type of model is called a linear model and becomes the mathematical basis for our discussion of the analysis of variance. Also note that this serves only as an example. Many linear models could be formulated for the two-factor experiment.

Assumptions

The following assumptions are made when using the F-test:

1. The response variable is continuous.

2. The e_{ijk} follow the normal probability distribution with mean equal to zero.

3. The variances of the e_{ijk} are equal for all values of i, j, and k.

4. The individuals are independent.

Limitations

There are few limitations when using these tests. Sample sizes may range from a few to several hundred. If your data are discrete with at least five unique values, you can assume that you have met the continuous variable assumption. Perhaps the greatest restriction is that your data comes from a random sample of the population. If you do not have a random sample, the F-test will not work.

The UWM procedure also requires that there are no missing cells. Because the concept of missing cells often gets confused with unbalanced data, we will give an example that discriminates between these two properties.

Let's assume that an experiment is designed to study the impact of education and region on income. Three regions are selected for this study. They are Boston, Chicago, and Denver. Two education levels are selected: high school and college. Hence, the experiment is a two-by-three factorial design with six treatment combinations (called "cells"). Suppose the researcher intends to sample ten individuals from each of the six treatment groups. If the experiment proceeds as planned, it will be balanced with no missing cells.

As long as there are ten individuals in each of the six cells, the design is said to be "balanced." Suppose that for one reason or another, two of the ten college people are lost from the Denver-college group. The design is now "unbalanced." Hence, an unbalanced design is one which has a differing number of individuals in each treatment group.

Suppose that instead of just two people, all ten individuals in the Denver-college group (cell) are lost from the study. Now the design has a missing cell." That is, one complete treatment combination is missing.

Again, the UWM procedure is exact for a balanced design, approximate for an unbalanced design with no missing cells, and impossible for a design with missing cells. Unfortunately, designs that are confounded, such as Latin squares and fractional factorials, have missing cells, so they cannot be analyzed with this procedure.

Multiple Comparison Procedures

The multiple comparison procedures are discussed in the One-Way Analysis of Variance chapter.

Friedman's Rank Test

Friedman's test is a nonparametric analysis that may be performed on data from a randomized block experiment. In this test, the original data are replaced by their ranks. It is used when the assumptions of normality and equal variance are suspect. In an experiment with b blocks and k treatments, the Friedman test statistic, Q, is calculated as follows:

$$Q = (k-1)\frac{12\sum_{j=1}^{k} R_j^2 - 3b^2 k(k+1)^2}{bk(k^2-1) - \sum(t^3 - t)}$$

The data within each of the b blocks are assigned ranks. The ranks are summed for each of the k groups. This rank sum is denoted as R_j. The factor involving t in the denominator is a correction for ties. The symbol t represents the number of times a single value is found within a block. When the *multiplicity* $\sum(t^3 - t)$ is included, the test is said to be corrected for ties. When this term is omitted, the test value is not corrected for ties.

This statistic is approximately distributed as a Chi-square with $k-1$ degrees of freedom.

The Q statistic is closely related to Kendall's coefficient of concordance, W, using the formula:

$$W = Q / b(k - 1)$$

In order to run this procedure, the first factor must be the blocking (random) factor and the second must be the treatment (fixed) factor.

Data Structure

The data must be entered in a format that puts the response in one variable and the values of each of the factors in other variables. An example of the data for a randomized-block design is shown next.

RNDBLOCK Dataset

Block	Treatment	Response
1	1	123
2	1	230
3	1	279
1	2	245
2	2	283
3	2	245
1	3	182
2	3	252
3	3	280
1	4	203
2	4	204
3	4	227

Comparisons

Comparisons are only generated for fixed factors. These options let you specify any planned comparisons that you want to run on this factor. A planned comparison is formulated in terms of the means as follows:

$$C_i = \sum_{j=1}^{J} w_{ij}\, m_j$$

In this equation, there are J levels in the factor, the means for each level of the factor are denoted m_j, and w_{ij} represents a set of J weight values for the ith comparison. The comparison value, C_i, is tested using a t-test. Note that if the w_{ij} sum to zero across j, the comparison is called a "contrast" of the means.

Comparisons may be specified by simply listing the weights. For example, suppose a factor has three levels (unique values). Further suppose that the first level represents a control group, the second a treatment at one dose, and the third a treatment at a higher dose. Three comparisons come to mind: compare each of the treatment groups of the control group and compare the two treatment groups to each other. These three comparisons would be

Control vs. Treatment 1	-1,1,0
Control vs. Treatment 2	-1,0,1
Treatment 1 vs. Treatment 2	0,-1,1

You might also be interested in comparing the control group with the average of both treatment groups. The weights for this comparison would be -2,1,1.

When a factor is quantitative, it might be of interest to divide the response pattern into linear, quadratic, cubic, and similar components. If the sample sizes are equal and the factor levels are equally spaced, these so-called components of trend may be studied by the use of simple contrasts. For example, suppose a quantitative factor has three levels: 5, 10, and 15. A contrast to test the linear and quadratic trend components would be

Linear trend	-1,0,1
Quadratic trend	1,-2,1

If the sample sizes for the groups are unequal (the design is unbalanced), adjustments must be made for the differing sample sizes.

NCSS will automatically generate some of the more common sets of contrasts or it will let you specify up to three custom contrasts yourself. The following common sets are designated available.

- **None**

    No comparisons are generated.

- **Standard Set**

    This option generates a standard set of contrasts in which the mean of the first level is compared to the average of the rest, the mean of the second group is compared to the average of those remaining, and so on.

    The following example displays the type of contrast generated by this option. Suppose there are four levels (groups) in the factor. The contrasts generated by this option are:

-3,1,1,1	Compare the first-level mean with the average of the rest.
0,-2,1,1	Compare the second-level mean with the average of the rest.
0,0,-1,1	Compare the third-level mean with the fourth-level mean.

- **Polynomial**

    This option generates a set of orthogonal contrasts that allow you to test various trend components from linear up to sixth order. These contrasts are appropriate even if the levels are unequally spaced or the group sample sizes are unequal. Of course, these contrasts are only appropriate for data that are at least ordinal. Usually, you would augment the analysis of this type of data with a multiple regression analysis.

    The following example displays the type of contrast generated by this option. Suppose there are four equally spaced levels in the factor and each group has two observations. The contrasts generated by this option are (scaled to whole numbers):

-3,-1,1,3	Linear component.
1,-1,-1,1	Quadratic component.
-1,3,-3,1	Cubic component.

- **Linear Trend**

    This option generates a set of orthogonal contrasts and retains only the linear component. This contrast is appropriate even if the levels are unequally spaced and the group sample sizes are unequal. See Orthogonal Polynomials above for more detail.

- **Linear-Quadratic Trend**

  This option generates the complete set of orthogonal polynomials, but only the results for the first two (the linear and quadratic) are reported.

- **Linear-Cubic Trend**

  This option generates the complete set of orthogonal polynomials, but only the results for the first three are reported.

- **Linear-Quartic Trend**

  This option generates the complete set of orthogonal polynomials, but only the results for the first four are reported.

- **Each with First**

  This option generates a set of nonorthogonal contrasts appropriate for comparing each of the remaining levels with the first level. The following example displays the type of contrast generated by this option. Suppose there are four levels (groups) in the factor. The contrasts generated by this option are:

-1,1,0,0	Compare the first- and second-level means.
-1,0,1,0	Compare the first- and third-level means.
-1,0,0,1	Compare the first- and fourth-level means.

- **Each with Last**

  This option generates a set of nonorthogonal contrasts appropriate for comparing each of the remaining levels with the last level. The following example displays the type of contrast generated by this option. Suppose there are four levels (groups) in the factor. The contrasts generated by this option are:

-1,0,0,1	Compare the first- and fourth-level means.
0,-1,0,1	Compare the second- and fourth-level means.
0,0,-1,1	Compare the third- and fourth-level means.

- **Custom**

  This option indicates that the contrasts listed in the corresponding three boxes of the Comparison panel should be used.

Example 1 – Analysis of a Randomized-Block Design

This section presents an example of how to run an analysis of the data contained in the Randomized Block dataset.

Setup

To run this example, complete the following steps:

1   Open the Randomized Block example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select Randomized Block and click OK.

2   Specify the Balanced Design Analysis of Variance procedure options

- Find and open the **Balanced Design Analysis of Variance** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Variables Tab

Response Variable(s) ....**Response**
Factor Variable 1 ....**Block**
  Type..**Random**
Factor Variable 2 ....**Treatment**
  Type..**Fixed**
  Comparisons...**Linear**

Reports Tab

Compute Power..**Checked**
Tukey-Kramer Test...**Checked**
Show Notes ....**Checked**

3   Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Expected Mean Squares

Expected Mean Squares
_____

Model Term	DF	Term Fixed?	Denominator Term	Expected Mean Square
A: Block	2	No	S	S+bsA
B: Treatment	3	Yes	AB	S+sAB+asB
AB	6	No	S	S+sAB
S	0	No		S

_____
Note: Expected Mean Squares are for the balanced cell-frequency case.

The Expected Mean Squares expressions are provided to show the appropriate error term for each factor. The correct error term for a factor is that term that is identical except for the factor being tested.

Model Term

The source of variation or term in the model.

DF

The degrees of freedom. The number of observations used by this term.

Term Fixed?

Indicates whether the term is "fixed" or "random."

Denominator Term

Indicates the term used as the denominator in the F-ratio.

Expected Mean Square

This expression represents the expected value of the corresponding mean square if the design was completely balanced. "S" represents the expected value of the mean square error (experimental variance). The uppercase letters represent either the adjusted sum of squared treatment means if the factor is fixed, or the variance component if the factor is random. The lowercase letter represents the number of levels for that factor, and "*s*" represents the number of replications of the experimental layout.

These EMS expressions are provided to determine the appropriate error term for each factor. The correct error term for a factor is that term whose EMS is identical except for the factor being tested.

The appropriate error term for factor B is the AB interaction. The appropriate error term for AB is S (mean square error). Since there are zero degrees of freedom for S, the terms A and AB cannot be tested.

Analysis of Variance Table

Analysis of Variance Table

Model Term	DF	Sum of Squares	Mean Square	F-Ratio	P-Value	Power†
A: Block	2	10648.67	5324.333			
B: Treatment	3	4650.917	1550.306	1.0934	0.42136	0.17794
AB	6	8507.333	1417.889			
S	0	0				
Total (Adjusted)	11	23806.92				
Total	12					

† Power was calculated using the observed F-Ratio as the population effect size with a significance level of α = 0.05.

Model Term

The source of variation. The term in the model.

DF

The degrees of freedom. The number of observations used by the corresponding model term.

Sum of Squares

This is the sum of squares for this term. It is usually included in the ANOVA table for completeness, not for direct interpretation.

Mean Square

An estimate of the variation accounted for by this term. The sum of squares divided by the degrees of freedom.

F-Ratio

The ratio of the mean square for this term and the mean square of its corresponding error term. This is also called the F-test value.

P-Value

The significance level of the above F-ratio. The probability of an F-ratio larger than that obtained by this analysis. For example, to test at an alpha of 0.05, this probability would have to be less than 0.05 to make the F-ratio significant. Note that if the value is significant at the specified value of alpha, a star is placed to the right of the F-Ratio.

Power

Power is the probability of rejecting the hypothesis that the means are equal when they are in fact not equal. Power is one minus the probability of type II error (β). The power of the test depends on the sample size, the magnitudes of the variances, the alpha level, and the actual differences among the population means.

The power value calculated here assumes that the population standard deviation is equal to the observed standard deviation and that the differences among the population means are exactly equal to the difference among the sample means.

High power is desirable. High power means that there is a high probability of rejecting the null hypothesis when the null hypothesis is false. This is a critical measure of precision in hypothesis testing.

Generally, you would consider the power of the test when you accept the null hypothesis. The power will give you some idea of what actions you might take to make your results significant. If you accept the null hypothesis with high power, there is not much left to do. At least you know that the means are not different. However, if you accept the null hypothesis with low power, you can take one or more of the following actions:

1. Increase your alpha level. Perhaps you should be testing at alpha = 0.05 instead of alpha = 0.01. Increasing the alpha level will increase the power.

2. Increase your sample size, which will increase the power of your test if you have low power. If you have high power, an increase in sample size will have little effect.

3. Decrease the magnitude of the variance. Perhaps you can redesign your study so that measurements are more precise and extraneous sources of variation are removed.

Friedman's Rank Test Section

Friedman's Rank Test

Test Results

Ties	Friedman's Statistic (Q)	DF	P-Value	Concordance (W)
Ignored	3.4	3	0.33397	0.38
Correction	3.4	3	0.33397	0.38

Correction Factor for Ties: Multiplicity = 0

Group Details

Treatment	Number of Blocks	Median	Ranks Mean	Ranks Sum
1	3	230	2.00	6
2	3	245	3.33	10
3	3	252	3.00	9
4	3	204	1.67	5

Ties

- **Ignored**

  Statistics on this row are not adjusted for ties.

- **Correction**

  Statistics on this row are adjusted for ties.

Friedman's Statistic (Q)

The value of Friedman's Q statistic. This statistic is approximately distributed as a Chi-square random variable with degrees of freedom equal to *k-1*, where *k* is the number of treatments. The Chi-square approximation grows closer as the number of blocks is increased.

DF

The degrees of freedom. The degrees of freedom is equal to *k-1*, where *k* is the number of treatments.

P-Value

The significance level of the *Q* statistic. If this value is less than a specified alpha level (often 0.05), the null hypothesis of equal medians is rejected.

Concordance (W)

The value of Kendall's coefficient of concordance, which measures the agreement between observers of samples. This value ranges between zero and one. A value of one indicates perfect concordance. A value of zero indicates no agreement or independent samples.

Multiplicity

The value of the correction factor for ties: $\sum(t^3 - t)$.

Treatment

The level of the treatment (fixed factor) whose statistics are reported on this line.

Number Blocks

The number of levels (categories) of the block variable (random factor).

Median

The median value of responses at this treatment level.

Mean of Ranks

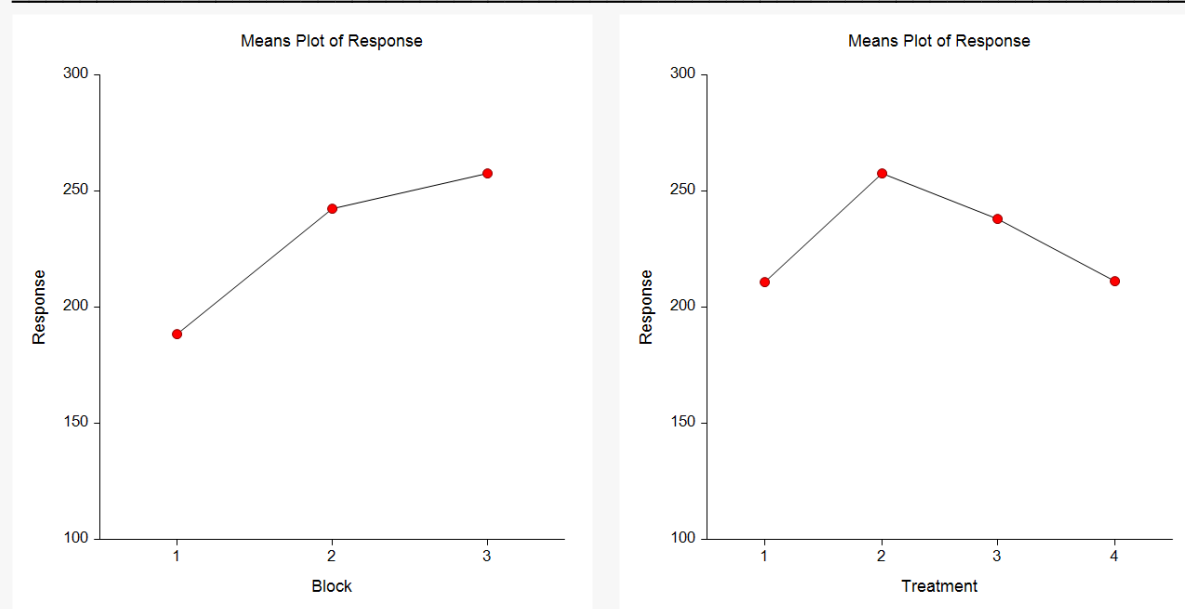The average of the ranks at this treatment level.

Sum of Ranks

The sum of the ranks at this treatment level.
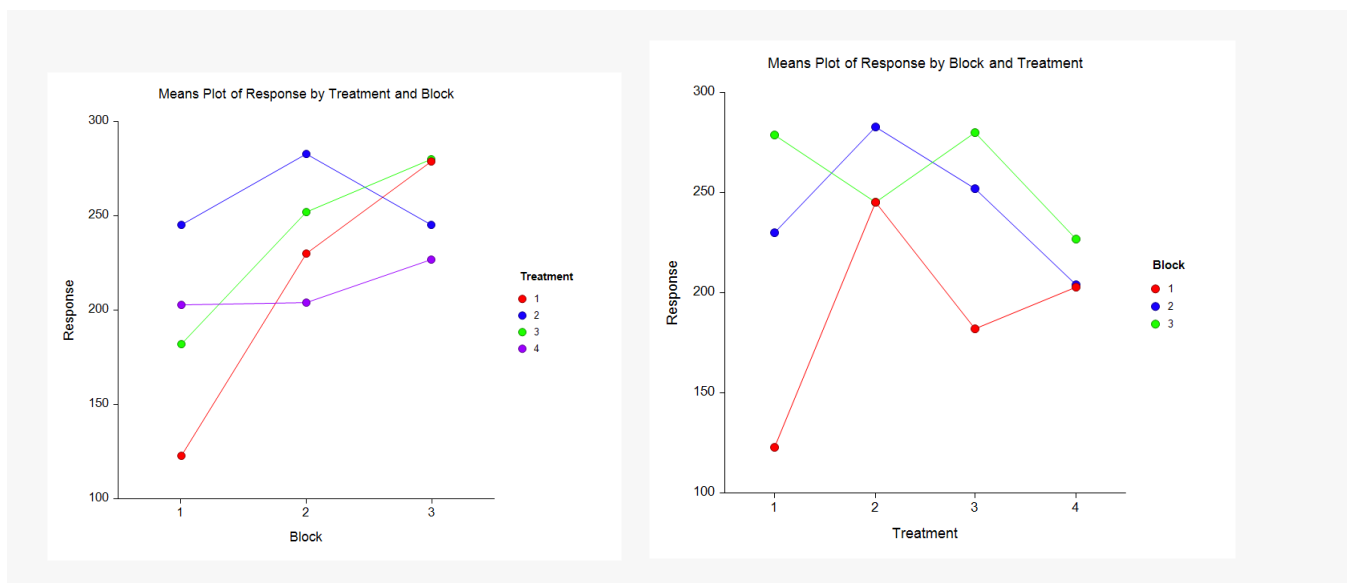
Means and Effects and Means Plots

Means and Effects

Term	Count	Mean	Standard Error	Effect
All	12	229.4167		229.4167
A: Block				
1	4	188.25	0	-41.16667
2	4	242.25	0	12.83333
3	4	257.75	0	28.33333
B: Treatment				
1	3	210.6667	21.74005	-18.75
2	3	257.6667	21.74005	28.25
3	3	238	21.74005	8.583333
4	3	211.3333	21.74005	-18.08333
AB: Block, Treatment				
1, 1	1	123	0	-46.5
1, 2	1	245	0	28.5
1, 3	1	182	0	-14.83333
1, 4	1	203	0	32.83333
2, 1	1	230	0	6.5
2, 2	1	283	0	12.5
2, 3	1	252	0	1.166667
2, 4	1	204	0	-20.16667
3, 1	1	279	0	40
3, 2	1	245	0	-41
3, 3	1	280	0	13.66667
3, 4	1	227	0	-12.66667

Means Plot(s)

Means Plot of Response

Means Plot of Response

Means Plot of Response by Treatment and Block

Means Plot of Response by Block and Treatment

Term

The label for this line of the report.

Count

The number of observations in the mean.

Mean

The value of the sample mean.

Standard Error

The standard error of the mean. Note that these standard errors are the square root of the mean square of the error term for this term divided by the count. *These standard errors are not the same as the simple standard errors calculated separately for each group.* The standard errors reported here are those appropriate for testing in multiple comparisons.

Effect

The additive component that this term contributes to the mean.

Means Plot(s)

These plots display the means for each factor and two-way interaction. Note how easily you can see patterns in the plots.

Multiple Comparison Reports

Tukey-Kramer Multiple Comparison Test
───

Comparison Term: B: Treatment
───

Group	Count	Mean	Different From Groups
1	3	210.6667	
2	3	257.6667	
3	3	238	
4	3	211.3333	
───

α = 0.05, Error Term = AB, DF = 6, MSE = 1417.889, Critical Value = 4.8952

Notes:
This report provides multiple comparison tests for all pairwise differences between the means.

These sections present the results of the multiple-comparison procedures selected. These reports all use a uniform format that will be described by considering Tukey-Kramer Multiple-Comparison Test. The reports for the other procedures are similar. For more information on the interpretation of various multiple-comparison procedures, turn to the section by that name in the One-way Analysis of Variance chapter.

Group

The label for this group.

Count

The number of observations in the mean.

Mean

The value of the sample mean.

Different From Groups

A list of those groups that are significantly different from this group according to this multiple-comparison procedure. All groups not listed are not significantly different from this group.

Alpha

The level of significance that you selected.

Error Term

The term in the ANOVA model that is used as the error term.

DF

The degrees of freedom of the error term.

MSE

The value of the mean square error.

Critical Value

The value of the test statistic that is "just significant" at the given value of alpha. This value depends on which multiple-comparison procedure you are using. It is based on the t-distribution or the studentized range distribution. It is the value of t, F, or q in the corresponding formulas.

Planned Comparison Reports

This section presents the results of any planned comparisons that were selected.

Planned Comparison: Linear Trend

Comparison Term: B: Treatment

Test Results

Comparison Value	Standard Error	T-Statistic	Two-Sided P-Value	Reject H0 at α = 0.05?	95% Confidence Interval Limits for the Comparison Value	
					Lower	Upper
-3.950387	21.74005	0.1817	0.86179	No	-57.14637	49.24559

α = 0.05, Error Term = AB, DF = 6, MSE = 1417.889

Group Details

Group	Comparison Coefficient	Count	Mean
1	-0.6708204	3	210.6667
2	-0.2236068	3	257.6667
3	0.2236068	3	238
4	0.6708204	3	211.3333

Notes:
This section presents the results of the planned comparison testing whether a first-order polynomial (a linear trend) is significant.

Comparison Value

The value of the comparison. This is formed by multiplying the comparison coefficient by the mean for each group and summing.

Standard Error

This is the standard error of the estimated comparison value. It is the denominator of the T-Value (above).

T-Statistic

The t-test used to test whether the above Comparison Value is significantly different from zero.

$$t_f = \frac{\sum_{i=1}^{k} c_i M_i}{\sqrt{MSE \sum_{i=1}^{k} \frac{c_i^2}{n_i}}}$$

where **MSE** is the mean square error, **f** is the degrees of freedom associated with **MSE**, **k** is the number of groups, **c_i** is the comparison coefficient for the ith group, **M_i** is the mean of the ith group, and **n_i** is the sample size of the ith group.

Two-Sided P-Value

The significance level of the above T-Value. The Comparison is statistically significant if this value is less than the specified alpha.

Reject H0 at α = 0.05?

The decision based on the specified value of the multiple-comparison alpha.

Group

The label for this group.

Comparison Coefficient

The coefficient (weight) used for this group.

Count

The number of observations in the mean.

Mean

The value of the sample mean.

Alpha

The level of significance that you selected.

Error Term

The term in the ANOVA model that is used as the error term.

DF

The degrees of freedom of the error term.

MSE

The value of the mean square error.