

Chapter 210

One-Way Analysis of Variance

Introduction

This procedure performs an F-test from a one-way (single-factor) analysis of variance, Welch's test, the Kruskal-Wallis test, the van der Waerden Normal-Scores test, and the Terry-Hoeffding Normal-Scores test on data contained in either two or more variables or in one variable indexed by a second (grouping) variable. The one-way analysis of variance compares the means of two or more groups to determine if at least one group mean is different from the others. The F-ratio is used to determine statistical significance. The tests are non-directional in that the null hypothesis specifies that all means are equal and the alternative hypothesis simply states that at least one mean is different.

This procedure also checks the equal variance (homoscedasticity) assumption using the Levene test, Brown-Forsythe test, Conover test, and Bartlett test.

Kinds of Research Questions

One of the most common tasks in research is to compare two or more populations (groups). We might want to compare the income level of two regions, the nitrogen content of three lakes, or the effectiveness of four drugs. The first question that arises concerns which aspects (parameters) of the populations we should compare. We might consider comparing the means, medians, standard deviations, distributional shapes (histograms), or maximum values. We base the comparison parameter on our particular problem.

One of the simplest comparisons we can make is between the means of two or more populations. If we can show that the mean of one population is different from that of the other populations, we can conclude that the populations are different. Other aspects of the populations can (and should) also be considered, but the mean is usually the starting point.

If we are willing to make assumptions about other characteristics of the populations (such as that they are normally distributed and that their variances are equal), we can use the F-ratio to compare the means of random samples drawn from these populations. If these assumptions are violated, the nonparametric Kruskal-Wallis test may be used.

Assumptions

The statistical tests described in this chapter make certain assumptions. One reason for the popularity of the F-test is its robustness in the face of assumption violation. However, if an assumption is not even approximately met, the significance levels and the power of the F-test are invalidated. Unfortunately, in practice it often happens that more than one assumption is not met. Hence, steps should be taken to check the assumptions before important decisions are made. The reports include sections that investigate these assumptions.

One-Way Analysis of Variance Assumptions

The assumptions of the **one-way analysis of variance** are:

1. The data are continuous (not discrete).
2. The data follow the normal probability distribution. Each group is normally distributed about the group mean.
3. The variances of the populations are equal.
4. The groups are independent. There is no relationship among the individuals in one group as compared to another.
5. Each group is a simple random sample from its population. Each individual in the population has an equal probability of being selected in the sample.

Nonparametric Test Assumptions

The assumptions of the **Kruskal-Wallis, van der Waerden, and Terry-Hoeffding tests** are:

1. The variable of interest is continuous (not discrete). The measurement scale is at least ordinal.
2. The probability distributions of the populations are identical, except for location. Hence, we still require that the population variances are equal.
3. The groups are independent.
4. All groups are simple random samples from their respective populations. Each individual in the population has an equal probability of being selected in the sample.

Evaluate Assumptions by Studying the Residuals

The normality, equal variance, and independent assumptions can be evaluated visually by storing them back on the spreadsheet and then plotting them using other procedures such as Scatter Plots. The residuals and the means can be stored when the procedure is run.

Technical Details

This section documents many of the tests that are presented in this procedure. Observations are obtained from K independent groups. Within the i^{th} group, there are n_i observations of a single variable Y . We use Y_{ij} to represent the value of Y from the j^{th} individual of the i^{th} group.

F-Test

The F test assumes that the observations are normally distributed with a common variance, but possibly different means. The formula for the one-way analysis of variance (ANOVA) F-test is

$$F_{K-1, N-K} = \frac{MSB}{MSE}$$

where

$$MSB = \frac{1}{K-1} \sum_{i=1}^K n_i (\bar{Y}_i - \bar{Y})^2$$

$$MSE = \frac{1}{N-K} \sum_{i=1}^K \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$$

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^K \sum_{j=1}^{n_i} Y_{ij}$$

$$N = \sum_{i=1}^K n_i$$

$$\bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$$

The quantities MSB and MSE are often called the *mean square between* and *mean square error*, respectively. This F is compared to a tabulated F with $K-1$ numerator degrees of freedom and $N-K$ denominator degrees of freedom.

Welch's Test of Means Allowing for Unequal Group Variances

When it is not accurate to assume that the variances are equal, the alternative Welch test (1951) is often suggested. Some recommend that this procedure be used when the largest group variance is over four times as large as the smallest group variance. Other suggest that little power is lost if this test is always used, even when the variances are almost equal.

Continuing with the definitions in the last section, the formula for Welch's test is

$$W^* = \frac{\sum w_i (\bar{Y}_i - \hat{\mu})^2 / (K - 1)}{1 + [2(K - 2) / (K^2 - 1)] \sum h_i}$$

where

$$w_i = \frac{n_i}{s_i^2}$$

$$W = \sum w_i$$

$$\hat{\mu} = \sum w_i \bar{Y}_i / W$$

$$h_i = \frac{(1 - w_i / W)^2}{(n_i - 1)}$$

$$f = \frac{K^2 - 1}{3 \sum h_i}$$

$$s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$$

The value of W^* is compared to an F distribution with $K-1$ and f degrees of freedom.

Kruskal-Wallis Test

This discussion is based on the discussion in Gibbons and Chakraborti (2011). The Kruskal-Wallis test is a nonparametric substitute for the one-way ANOVA when the assumption of normality is not valid. Two key assumptions are that the group distributions are at least ordinal in nature and that they are identical, except for location. That is, this test assumes equal variances. This means that ties (repeated values) are not acceptable. When ties are present in your data, you should use the corrected version of this test.

The formula for the Kruskal-Wallis test is

$$H = \frac{12}{N(N + 1)} \sum_{i=1}^K \frac{R_i^2}{n_i} - 3(N + 1)$$

One-Way Analysis of Variance

The Kruskal-Wallis test corrected for ties is calculated by dividing H by a correction factor. The formula for the corrected version of H is

$$H_c = \frac{H}{1 - \frac{\sum t(t^2 - 1)}{N(N^2 - 1)}}$$

In both of the above formulas, R_i is the sum of the ranks of the i^{th} group, and t is the count of a particular tie. The summation is over all tied values.

The calculated value of H_c is compared to chi-squared distribution with $K-1$ degrees of freedom.

Normal Scores (Terry-Hoeffding and van der Waerden) Tests

This discussion is based on the discussion in Gibbons and Chakraborti (2011). These normal score tests are similar to the Kruskal-Wallis test, except that instead of being calculated on the ranks of the data, they are calculated on the normal scores. The **Terry-Hoeffding** (or Fisher-Yates-Terry-Hoeffding) test replaces the data values with the expected value of the normal order statistics. The **van der Waerden** test replaces the data values with the inverse normal scores (the $i/(N+1)^{\text{th}}$ quantile, where i indexes all data in all groups). Ties are resolved by substituting the average score of the tied values. These tests are equivalent in large samples.

The amazing feature of these tests is that, unlike the Kruskal-Wallis test, their asymptotic relative efficiency is one when the underlying data distribution is normal. That is, their power is as good as that of the F-test when the data are normal. When they data are not normal, their efficiency is often much better than the F-test.

The general formula for H , a normal scores test, is

$$H = \frac{1}{S^2} \sum_{i=1}^K n_i \bar{A}_i^2$$

where

$$S^2 = \frac{1}{N-1} \sum_{i=1}^K \sum_{j=1}^{n_i} A_{ij}^2$$

$$\bar{A}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} A_{ij}$$

The test assumes the H is distributed as a chi-squared random variable with $K-1$ degrees of freedom.

For the Terry-Hoeffding test, the value of A_{ij} is the expected value of the normal order statistic

$$A_{ij} = E(\xi_{ij})$$

where ξ_k is the k^{th} order statistic from a standard normal population.

One-Way Analysis of Variance

For the van der Waerden test, the value of A_{ij} is the inverse normal score

$$A_k = \Phi^{-1}\left(\frac{k}{N+1}\right)$$

where $\Phi^{-1}(x)$ is the inverse of the cumulative normal distribution.

Levene Test of Homogeneity (Equal Variance)

Levene (1960) presents a test of homogeneity (equal variance). The test does not assume that all populations are normally distributed and is recommended when the normality assumption is not viable.

Suppose K groups each have a normal distribution with possibly different means and standard deviations $\sigma_1, \sigma_2, \dots, \sigma_K$. Let n_1, n_2, \dots, n_K denote the number of subjects in each group, Y_{ij} denote response values, and N denote the total sample size of all groups. The test assumes that the data are obtained by taking a simple random sample from each of the K populations.

The formula for the calculation of Levene's test is

$$W = \frac{(N - K) \sum_{i=1}^K n_i (Z_i - \bar{Z})^2}{(K - 1) \left\{ \sum_{i=1}^K \sum_{j=1}^{n_i} (Z_{ij} - \bar{Z}_i)^2 \right\}}$$

where

$$Z_{ij} = |Y_{ij} - \bar{Y}_i|$$

$$\bar{Z}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Z_{ij}$$

$$\bar{Z} = \frac{1}{N} \sum_{i=1}^K \sum_{j=1}^{n_i} Z_{ij}$$

$$\bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$$

If the assumptions are met, the distribution of this test statistic follows the F distribution with degrees of freedom $K - 1$ and $N - K$.

Brown-Forsythe Test of Homogeneity (Equal Variance)

Brown and Forsythe (1974) present a robust version of Levene's (1960) test of homogeneity (equal variance). The test does not assume that all populations are normally distributed and is recommended when the normality assumption is not viable.

Suppose K groups each have a normal distribution with possibly different means and standard deviations $\sigma_1, \sigma_2, \dots, \sigma_K$. Let n_1, n_2, \dots, n_K denote the number of subjects in each group, Y_{ij} denote response values, and N denote the total sample size of all groups. The test assumes that the data are obtained by taking a simple random sample from each of the K populations.

The formula for the calculation of Brown-Forsythe test is

$$W = \frac{(N - K) \sum_{i=1}^K n_i (Z_i - \bar{Z})^2}{(K - 1) \left\{ \sum_{i=1}^K \sum_{j=1}^{n_i} (Z_{ij} - \bar{Z}_i)^2 \right\}}$$

where

$$Z_{ij} = |Y_{ij} - \tilde{Y}_i|$$

$$\bar{Z}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Z_{ij}$$

$$\bar{Z} = \frac{1}{N} \sum_{i=1}^K \sum_{j=1}^{n_i} Z_{ij}$$

$$\tilde{Y}_i = \text{median}(Y_{i1}, Y_{i2}, Y_{i3}, \dots, Y_{in_i})$$

If the assumptions are met, the distribution of this test statistic follows approximately the F distribution with degrees of freedom $K - 1$ and $N - K$.

Conover Test of Homogeneity (Equal Variance)

Conover (1999) present a nonparametric test of homogeneity (equal variance) based on ranks. The test does not assume that all populations are normally distributed and is recommended when the normality assumption is not viable.

Suppose K groups each have a normal distribution with possibly different means and standard deviations $\sigma_1, \sigma_2, \dots, \sigma_K$. Let n_1, n_2, \dots, n_K denote the number of subjects in each group, Y_{ij} denote response values, and N denote the total sample size of all groups. The test assumes that the data are obtained by taking a simple random sample from each of the K populations.

The formula for the calculation of Conover test is

$$T = \frac{1}{D^2} \left[\sum_{i=1}^K \frac{S_i^2}{n_i} - N\bar{S}^2 \right]$$

One-Way Analysis of Variance

where

$$Z_{ij} = |Y_{ij} - \bar{Y}_i|$$

$$R_{ij} = \text{Rank}(Z_{ij})$$

$$S_i = \sum_{j=1}^{n_i} R_{ij}^2$$

$$\bar{S} = \frac{1}{N} \sum_{i=1}^K S_i$$

$$D^2 = \frac{1}{N-1} \left[\sum_{i=1}^K \sum_{j=1}^{n_i} R_{ij}^4 - N\bar{S}^2 \right]$$

If the assumptions are met, the distribution of this test statistic follows approximately the *chi-squared* distribution with degrees of freedom $K - 1$.

Bartlett Test of Homogeneity (Equal Variance)

Bartlett (1937) presents a test of homogeneity (equal variance). This test is discussed in several texts on statistical methods such as Winer (1991). The test assumes that all populations are normally distributed and is not recommended when the normality assumption is not viable.

Suppose K groups each have a normal distribution with possibly different means and standard deviations $\sigma_1, \sigma_2, \dots, \sigma_K$. Let n_1, n_2, \dots, n_K denote the number of subjects in each group and let N denote the total sample size of all groups. The test assumes that the data are obtained by taking a simple random sample from each of the K populations.

The formula for the calculation of Bartlett's test is

$$\chi_{K-1}^2 = \frac{(N-K)\log(S_p^2) - \sum_{i=1}^K (n_i-1)\log(S_i^2)}{1 + \frac{1}{3K-3} \left\{ \sum_{i=1}^K \left(\frac{1}{n_i-1} \right) - \frac{1}{N-K} \right\}}$$

where

$$S_p^2 = \frac{1}{N-K} \sum_{i=1}^K (n_i-1)S_i^2$$

$$S_i^2 = \frac{1}{n_i-1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

If the assumptions are met, the distribution of this test statistic follows the *chi-squared* distribution with degrees of freedom $K-1$.

Normality Tests

Shapiro-Wilk Normality Test

This test for normality has been found to be the most powerful test in most situations. It is the ratio of two estimates of the variance of a normal distribution based on a random sample of n observations. The numerator is proportional to the square of the best linear estimator of the standard deviation. The denominator is the sum of squares of the observations about the sample mean. The test statistic W may be written as the square of the Pearson correlation coefficient between the ordered observations and a set of weights which are used to calculate the numerator. Since these weights are asymptotically proportional to the corresponding expected normal order statistics, W is roughly a measure of the straightness of the normal quantile-quantile plot. Hence, the closer W is to one, the more normal the sample is.

The probability values for W are valid for sample sizes greater than 3. The test was developed by Shapiro and Wilk (1965) for sample sizes up to 20. **NCSS** uses the approximations suggested by Royston (1992) and Royston (1995) which allow unlimited sample sizes. Note that Royston only checked the results for sample sizes up to 5000 but indicated that he saw no reason larger sample sizes should not work. W may not be as powerful as other tests when ties occur in your data. The test is not calculated when a frequency variable is specified.

Anderson-Darling Normality Test

This test, developed by Anderson and Darling (1954), is a popular among those tests that are based on EDF statistics. In some situations, it has been found to be as powerful as the Shapiro-Wilk test.

Skewness Normality Test

This skewness test was reported by D'Agostino (1990). Skewness implies a lack of symmetry. One characteristic of the normal distribution is that it has no skewness. Hence, one type of non-normality is skewness.

The formula for the sample skewness is

$$\sqrt{b_1} = \frac{m_3}{m_2^{3/2}}$$

where the sample moments are defined by

$$m^k = \frac{1}{N} \sum_{i=1}^K \sum_{j=1}^{n_i} (d_{ij} - \bar{d})^k$$

$$d_{ij} = Y_{ij} - \bar{Y}_i$$

$$\bar{d} = \frac{1}{N} \sum_{i=1}^K \sum_{j=1}^{n_i} d_{ij}$$

One-Way Analysis of Variance

The distribution of the Z_5 is well-approximated by the normal distribution in sample sizes of at least 8

$$Z_5 = D \ln(V/A + \{(V/A)^2 + 1\}^{1/2})$$

where

$$A = \sqrt{\frac{2}{W^2 - 1}}$$

$$D = \sqrt{\frac{1}{\ln(W)}}$$

$$W^2 = -1 + \sqrt{2(B - 1)}$$

$$B = \frac{3(N^2 + 27N - 70)(N + 1)(N + 3)}{(N - 2)(N + 5)(N + 7)(N + 9)}$$

$$V = \sqrt{b_1} \sqrt{\frac{(N + 1)(N + 3)}{6(N - 2)}}$$

A two-sided test of the significance of Z_5 is based on the standard normal distribution.

Kurtosis Normality Test

This kurtosis test was reported by D'Agostino (1990). Kurtosis measures the heaviness of the tails of the distribution. For the normal distribution, the kurtosis is three.

Using the same sample moment definitions as those of the skewness, the formula for the sample kurtosis is

$$b_2 = \frac{m_4}{m_2^2}$$

A two-sided test of whether the kurtosis is three based on the approximate normality of Z_k can be computed as follows

$$Z_k = \frac{\left(1 - \frac{2}{9A}\right) - \left(\frac{1 - 2/A}{1 + x\sqrt{2/(A-4)}}\right)^{1/3}}{\sqrt{\frac{2}{9A}}}$$

One-Way Analysis of Variance

where

$$A = 6 + \frac{8}{\beta} \left(\frac{2}{\beta} + \sqrt{1 + \frac{4}{\beta}} \right)$$

$$\beta = \frac{6(N^2 - 5N + 2)}{(N + 7)(N + 9)} \sqrt{\frac{6(N + 3)(N + 5)}{N(N - 2)(N - 3)}}$$

$$x = \frac{b_2 - M}{\sqrt{V}}$$

$$M = \frac{3(N - 1)}{N + 1}$$

$$V = \frac{24N(N - 2)(N - 3)}{(N^2 + 1)(N + 3)(N + 5)}$$

Omnibus Normality Test

A third normality test, also developed by D'Agostino (1990), combines the skewness and kurtosis tests into a single measure. The test statistic C is based on the previous two tests. Its formula is given by

$$C = Z_s^2 + Z_k^2$$

This C is distributed as a chi-squared random variable with two degrees of freedom when the population is normally distributed.

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 come from a random sample of the population. If you do not have a random sample, your significance levels will be incorrect.

Multiple Comparison Procedures

Given that the analysis of variance (ANOVA) test finds a significant difference among treatment means, the next task is to determine which treatments are different. Multiple comparison procedures (MCPs) are methods that pinpoint which treatments are different.

The discussion to follow considers the following experiment. Suppose an experiment studies how two gasoline additives influence the miles per gallon obtained. Three types of gasoline were studied. The first sample received additive W, the second received additive V, and the third did not receive an additive (the control group).

If the F-test from an ANOVA for this experiment is significant, we do not know which of the three possible pairs of groups are different. MCPs can help solve this dilemma.

Multiple Comparison Considerations

Whenever MCPs are to be used, the researcher needs to contemplate the following issues.

Exploration Versus Decision-Making

When conducting exploration (or data snooping), you make several comparisons to discover the underlying factors that influence the response. In this case, you do not have a set of planned comparisons to make. In contrast, in a decision-making mode, you would try to determine which treatment is preferred. In the above example, because you do not know which factors influence gasoline additive performance, you should use the exploration mode to identify those. A decision-making emphasis would choose the gasoline that provides the highest miles per gallon.

Choosing a Comparison Procedure

You should consider two items here. First, will you know before or after experimentation which comparisons are of interest? Second, are you interested in some or all possible comparisons? Your choice of an MCP will depend on how you answer these two questions.

Error Rates

You will need to consider two types of error rates: comparisonwise and experimentwise.

1. Comparisonwise error rate. In this case, you consider each comparison of the means as if it were the only test you were conducting. This is commonly denoted as α . The conceptual unit is the comparison. Other tests that might be conducted are ignored during the calculation of the error rate. If we perform several tests, the probability of a type I error on each test is α .
2. Experimentwise, or familywise, error rate. In this situation, the error rate relates to a group of independent tests. This is the probability of making one or more type I errors in a group of independent comparisons. We will denote this error rate as α_f .

One-Way Analysis of Variance

The relationship between these two error rates is:

$$\alpha_f = 1 - (1 - \alpha)^c$$

where c is the total number of comparisons in the family. The following table shows these error rates for a few values of c and α . The body of the table consists of the calculated values of α_f .

Calculated Experimentwise Error Rates

α	k				
	2	3	5	10	20
0.20	0.360	0.488	0.672	0.893	0.988
0.10	0.190	0.271	0.410	0.651	0.878
0.05	0.098	0.143	0.226	0.401	0.642
0.02	0.040	0.059	0.096	0.183	0.332
0.01	0.020	0.030	0.049	0.096	0.182

As you can see, the possibility of at least one erroneous result goes up markedly as the number of tests increases. For example, to obtain an α_f of 0.05 with a c of 5, you would need to set α to 0.01.

Multiple Comparison Procedures

All of the multiple comparison procedures (MCPs) considered here assume that there is independence between treatments or samples, equal variance for each treatment, and normality (except the Kruskal-Wallis Z , which does not need normality). In addition, unless stated otherwise, the significance tests are assumed to be two-tailed.

Let \bar{y}_i and n_i represent the mean and sample size of the i^{th} treatment group. Let s^2 represent the mean square error for these means based on ν degrees of freedom. Let k be the number of treatments being compared for a factor or interaction.

Alpha

This is the α_f , or α , specified for the multiple comparison test. It may be comparisonwise or experimentwise, depending on the test. This alpha can range from 0.01 to 0.10.

One-Way Analysis of Variance

Bonferroni (All Pairs)

The Bonferroni MCP chooses the comparisonwise error rate in such a way as to control the desired experimentwise α_f . With k means and with an interest in all-possible pairs, the comparisonwise error rate is defined as $\alpha = \alpha_f / (k(k - 1))$. The significance test for any pair would be as follows, where $t_{\alpha, v}$ is a Student's t with v degrees of freedom

$$\frac{|\bar{y}_i - \bar{y}_j|}{\sqrt{s^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}} \geq t_{\alpha, v}$$

Generally, this MCP is run after the fact to find out which pairs are different.

Bonferroni (Versus Control)

If one of the treatments is a control group and you want to compare all of the other means to the mean of this control group, there are $k - 1$ comparisons. Again, you should choose the comparisonwise error rate in such a way as to achieve the overall or experimentwise α . The comparisonwise error rate is $\alpha = \alpha_f / (2(k - 1))$. The significance test for any two means would be as follows, where $t_{\alpha, v}$ is a Student's t

$$\frac{|\bar{y}_i - \bar{y}_j|}{\sqrt{s^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}} \geq t_{\alpha, v}$$

Comparison

This is a planned (a priori) significance test for a specific comparison that would also have a comparisonwise error rate. If you wanted to make several planned comparisons contained within one of the possible comparison options (the standard set of comparisons, the set of orthogonal polynomials, the set of comparisons with each treatment with the first treatment, the set with each treatment with the last treatment, or a set of no more than three customized contrasts), you could adjust the comparisonwise error rate to achieve a specific overall error rate. This test, distributed as a Student's $t_{\alpha/2, v}$, would be as follows, where a_j are the comparison coefficients:

$$\frac{|\sum_j a_j \bar{y}_j|}{s \sqrt{\sum_j \frac{a_j^2}{n_j}}} \geq t_{\alpha/2, v}$$

One-Way Analysis of Variance

Duncan's

This MCP looks at all pairwise comparisons among k means, but the error rate is neither on an experimentwise nor on a comparisonwise basis. The error rate is based on the number of steps apart, r , the two means are when they are ordered. The probability of falsely rejecting the equality of two population means when the sample means are r steps apart is $1 - (1 - \alpha)^{r-1}$. The significance test is based on the Studentized range, $q_{\alpha,r,v}$:

$$\frac{|\bar{y}_i - \bar{y}_j|}{s / \sqrt{\frac{2}{\frac{1}{n_i} + \frac{1}{n_j}}}} \geq q_{\alpha,r,v}$$

Dunnett's One and Two-Tailed Tests Versus a Control

If one of the treatments is a control group and you want to compare all of the other means to the mean of this control group, there are $k - 1$ comparisons. Dunnett's multiple comparison procedure (see Hsu(1996)) gives an experimentwise error rate of α . The significance test for any two means would be as follows, where $q_{\alpha,v}$ is calculated using Dunnett's formulas.

$$\frac{|\bar{y}_i - \bar{y}_j|}{\sqrt{s^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}} \geq q_{\alpha,v}$$

Often, it is of interest to find only those treatments that are better (or worse) than the control, so both one and two sided versions of this test are provided.

A set of two-sided, simultaneous confidence intervals are also provided for the difference between each treatment and the control.

Dwass-Steel-Critchlow-Fligner Test

The Dwass-Steel-Critchlow-Fligner test procedure is referenced in Hollander, Wolfe, and Chicken (2014). It is a two-sided, nonparametric procedure that provides family-wise error rate protection. It is most often used after significance is found from the Kruskal-Wallis test.

The Dwass-Steel-Critchlow-Fligner test procedure is to calculate a Wilcoxon rank sum test statistic, W , on each pair of groups. If there are ties, average ranks are used in the individual rank sum statistics. The p -values for these tests are calculated using the distribution of the range of independent, standard-normal variables. This is different from the usual calculation based on the Mann-Whitney U statistic. This provides the family-wise error rate protection.

We refer you to the *Two-Sample T-Test* chapter for the details of calculating the Wilcoxon rank sum statistic W .

One-Way Analysis of Variance

Fisher's LSD

Fisher's least significant difference (FSD) is a special version of the least significant difference (LSD). The difference between LSD and FSD is that FSD is only used when the F-test for the term is significant. LSD and FSD are used for pairwise comparisons

The error rate for each comparison is comparisonwise. This test has no control of the experimentwise error rate. The significance test is as follows, where $\gamma = \alpha/2$ for the LSD and $\gamma = \alpha/c$ for the FSD.

$$\frac{|\bar{y}_i - \bar{y}_j|}{\sqrt{s^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}} \geq t_{\gamma, v}$$

Hsu's Tests Versus the Best

Hsu (1996) chapter 4 provides a procedure for testing each group versus the best. This procedure is useful when you want to determine which of treatments is the best. Note that because of sampling variability, the sample best may not necessarily be the true best. Hsu's constrained multiple comparison with the best procedure allows the candidates for the best to be compared.

The method uses Dunnett's one-sided critical values, q_i , to provide simultaneous confidence intervals for

$$\mu_i - \max_{j \neq i}(\mu_j), i = 1, \dots, k$$

which are constrained to include 0. The constraints were suggested by John W. Tukey because a confidence interval for the above quantity whose lower limit is 0 indicates that the i^{th} treatment is the best. Likewise, a confidence interval for the above quantity whose upper limit is 0 indicates that the i^{th} treatment is not the best.

Hsu's confidence intervals are given by

$$[D_i^-, D_i^+], i = 1, \dots, k$$

where

$$D_i^+ = \max \left[0, \min_{j \neq i} \left((\hat{\mu}_i - \hat{\mu}_j) + sq_i \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \right) \right]$$

$$G = \{i: D_i^+ > 0\}$$

$$D_i^- = \begin{cases} 0 & \text{if } G = \{i\} \\ \min_{j \in G, j \neq i} \left((\hat{\mu}_i - \hat{\mu}_j) - sq_j \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \right) & \text{otherwise} \end{cases}$$

One-Way Analysis of Variance

Kruskal-Wallis Z (Dunn's Test)

This test is attributed to Dunn (1964) and is referenced in Daniel (1990), pages 240 - 242. This MCP is a distribution-free multiple comparison, meaning that the assumption of normality is not necessary. It is to be used for testing pairs of medians following the Kruskal-Wallis test. The test needs sample sizes of at least five (but preferably larger) for each treatment. The error rate is adjusted on a comparisonwise basis to give the experimentwise error rate, α_f . Instead of using means, this MCP uses average ranks, as the following formula indicates, with $\alpha = \alpha_f / (k(k - 1))$:

$$\frac{|\bar{R}_i - \bar{R}_j|}{\sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}} \geq z_\alpha$$

Adjusted for ties the inequality becomes

$$\frac{|\bar{R}_i - \bar{R}_j|}{\sqrt{\frac{[N(N^2 - 1) - (\sum t^3 - \sum t)] \left[\frac{1}{n_i} + \frac{1}{n_j} \right]}{12(N - 1)}}} \geq z_\alpha$$

In these inequalities, N is the total sample size and t is the number of values in the combined sample that are tied at a given rank.

Newman-Keuls

The Newman-Keuls MCP relies on the number of ordered steps r , where r ranges from 2 to k , between two sample means. The error rate is neither experimentwise nor comparisonwise. Instead, it is defined for sample means which are the same number of ordered steps apart. This test relies on the Studentized range distribution.

$$\frac{|\bar{y}_i - \bar{y}_j|}{s / \sqrt{\frac{2}{\frac{1}{n_i} + \frac{1}{n_j}}}}} \geq q_{\alpha, r, v}$$

Scheffe's

This MCP can be used to examine all possible comparisons among k means or just to look at all pairs as done here. It controls the overall or experimentwise error rate and is less sensitive than the Tukey-Kramer MCP. The significance test for pairs is as follows:

$$\frac{|\bar{y}_i - \bar{y}_j|}{\sqrt{s^2 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}} \geq \sqrt{(k - 1) F_{\alpha, k-1, v}}$$

One-Way Analysis of Variance

Tukey-Kramer

This test can be used to examine all pairs of treatment means. The error rate is experimentwise, and this test uses the Studentized range distribution. This test is conservative, which means that the two averages must be very different. The significance test follows:

$$\frac{|\bar{y}_i - \bar{y}_j|}{\sqrt{\frac{s^2}{2} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}} \geq q_{\alpha, k, v}$$

Tukey-Kramer Simultaneous Confidence Interval and P-Values

The Tukey-Kramer procedure can be used to produce simultaneous confidence intervals for all differences based on the Studentized range distribution. The report will also give the multiple comparison p-values of these differences. These p-values are the significance level of the multiple comparison p-value at which the difference becomes significant.

Recommendations

These recommendations assume that normality and equal variance are valid. If normality is not valid for each treatment, then use the Kruskal-Wallis Z MCP.

1. Planned all-possible pairs. If you are interested in paired comparisons only and you know this in advance, use either the Bonferroni for pairs or the Tukey-Kramer MCP.
2. Unplanned all-possible pairs. Use Scheffe's MCP.
3. Each versus a control. Use Dunnett's test.
4. Selected but planned. Use Comparison and adjust the alpha level accordingly.
5. Comparison with the best. Use Hsu's procedure.

Data Structure

The data may be entered in two formats, as shown in the examples below. The examples give the yield of corn for three types of fertilizer. The first format, shown in the first table below, puts the responses for each group in separate variables; that is, each variable contains all responses for a single group.

The second format, shown in the second table below, arranges the data so that all responses are entered in a single variable. A second variable, the Grouping Variable, contains an index that gives the group (A, B, or C) to which that row of data belongs.

In most cases, the second format is more flexible. Unless there is some special reason to use the first format, we recommend that you use the second.

Three Response Variables

Yield A	Yield B	Yield C
452	546	785
874	547	458
554	774	886
447	465	536
356	459	
754	665	669
558	467	857
574	365	821
664	589	772
682	534	732
	456	689
547	651	654
	654	
435	665	297
	546	830
245	537	827

One Grouping and One Response Variable

Row	Fertilizer	Yield
1	B	546
2	B	547
3	B	774
4	B	465
5	B	459
6	B	665
7	B	467
8	B	365
9	B	589
10	B	534
11	B	456
12	B	651
13	B	654
14	B	665
15	B	546
16	B	537
17	A	452
18	A	874
19	A	554
20	A	447
21	A	356
22	A	754
23	A	558
24	A	574
25	A	664
26	A	682
27	A	547
28	A	435
29	A	245
30	C	785
31	C	458
32	C	886
33	C	536
34	C	669
35	C	857
36	C	821
37	C	772
38	C	732
39	C	689
40	C	654
41	C	297
		356

Example 1 – Running a One-Way ANOVA

This section presents an example of how to run a one-way analysis of variance. We will use the corn yield data contained in the Corn Yield dataset. These data are contained in the variables labeled YldA, YldB, and YldC.

Setup

To run this example, complete the following steps:

1 Open the Corn Yield example dataset

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

2 Specify the One-Way Analysis of Variance procedure options

- Find and open the **One-Way 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

Input Type.....**Multiple Response Variables, One Variable per Factor Category**
 Response Variables.....**YldA-YldC**
 Planned Comparisons**Custom**
 Custom 1**-2 1 1**

Reports Tab

Compute Power.....**Checked**
 Duncan's Test.....**Checked**
 Kruskal-Wallis Z-Value Test (Dunn's Test)**Checked**
 Show Notes**Checked**

Report Options (*in the Toolbar*)

Variable Labels.....**Column Names**

3 Run the procedure

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

Residual Normality Assumption Tests

Residual Normality Assumption Tests

Test Name	Test Statistic	P-Value	Reject Normality at $\alpha = 0.2?$
Shapiro-Wilk	0.9875	0.91421	No
Anderson-Darling	0.2667	0.68883	No
D'Agostino Skewness	-0.1787	0.85814	No
D'Agostino Kurtosis	0.4200	0.67447	No
D'Agostino Omnibus (Skewness and Kurtosis)	0.2084	0.90106	No

This section presents the results of tests validating the normality assumption. The ANOVA assumes residuals (deviations for group means) are normal. Hence, these normality tests are performed on the combined set of residuals from all groups to test the normality assumption. Other assumptions concerning independence and random sampling are not tested here. You must justify those assumptions by considering your experiment procedure.

When using this report, all you need to do is scan down the column labeled *Reject Normality?* If none of the tests are marked as 'Yes', there is no evidence that the assumptions are invalid. (Of course, the power of these tests is influenced by your sample size. If you have a small sample size, say less than 25 per group, the power of the normality tests will be questionable, and you will have to rely on other means to justify your assumptions.)

Two aspects of normality are tested for, skewness and kurtosis. If the normality of residuals fails because of skewness, it might be possible to use the square root or logarithmic transformation to *normalize* your data.

If the data are found to be non-normal, one of the nonparametric tests is recommended. The basic assumptions of independent samples, continuous random variables, and a measurement scale of at least ordinal scale hold for these tests. The Kruskal-Wallis test has the additional assumption that the distributions for the groups are identical (although not necessarily normal) in form and shape (i.e., same variance) but differ only in location (i.e., in medians).

Finally, you should back up the results of these numerical tests by considering the box plots of the groups. As explained below, they let you visually determine if the assumptions of normality and equal variance are justified.

Equality of Group Variances Assumption Tests

Equality of Group Variances Assumption Tests

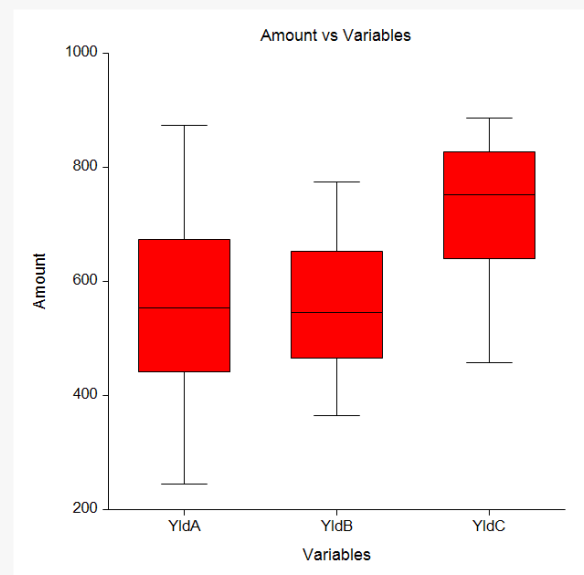
Test Name	Test Statistic	P-Value	Reject Equal Variances at $\alpha = 0.2?$
Brown-Forsythe (Data - Medians)	1.0866	0.34711	No
Levene (Data - Means)	1.0789	0.34964	No
Conover (Ranks of Deviations)	3.0747	0.21495	No
Bartlett (Likelihood Ratio)	3.0198	0.22093	No

This section presents the results of tests validating the equal variance assumption. Conover et al. (1981) did extensive simulation involving different distributions, sample sizes, means, and variances. They found that the Brown-Forsythe test is the most robust and powerful for testing the equality of variance in most situations.

The details of these tests are presented above in the Technical Details section.

Box Plot

Box Plot



Box plots are useful for assessing symmetry, presence of outliers, general equality of location, and equality of variation.

Expected Mean Squares

Expected Mean Squares

Model Term	DF	Term Fixed?	Denominator Term	Expected Mean Square
A (...)	2	Yes	σ^2	$\sigma^2 + sA$
Error	40	No		σ^2

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

The Expected Mean Square 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. In this procedure the factor variable is always fixed.

Denominator Term

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

Expected Mean Square

This is the expected value of the mean square for the term in the ANOVA model assuming balanced data (equal group counts). σ^2 represents the expected value of the mean square error. The uppercase letters represent the adjusted sum of squared treatment means. The lowercase letter represents the number of levels for that factor, and "s" represents the number of replications of the whole experimental layout.

Analysis of Variance Table

Analysis of Variance Table

Model Term	DF	Sum of Squares	Mean Square	F-Ratio	P-Value	Reject Equal Means at $\alpha = 0.05$?	Power*
Between	2	268532.4	134266.2	7.4740	0.00175	Yes	0.92528
Within (Error)	40	718574.3	17964.36				
Adjusted Total	42	987106.6					
Total	43						

* Power was calculated using the observed F-Ratio as the population effect size with a significance level of $\alpha = 0.05$.

The formulas used in this report were given in the Technical Details section above.

One-Way Analysis of Variance

Model Term

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 level 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.

Reject Equal Means at $\alpha = 0.05$?

This column indicates whether the null hypothesis of mean equality can be rejected at the indicated significance level. If the result is a ‘No,’ it might be because the means are about the same or it might be because the sample size is too small.

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.

One-Way Analysis of Variance

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 α level. Perhaps you should be testing at $\alpha = 0.05$ instead of $\alpha = 0.01$. Increasing the alpha level will increase the power.
2. Increasing your sample size 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.

Welch's Test of Means Allowing for Unequal Variances

Welch's Test of Means Allowing for Unequal Variances

Model Term	DF		F-Ratio	P-Value	Reject Equal Means at $\alpha = 0.05$?
	Numerator	Denominator			
Between Groups	2	24.27	8.0236	0.00211	Yes

The formulas used in this report were given in the Technical Details section above. This test is recommended when there is an indication that the group variance are unequal. It is the direct unequal-variance analog of the F-test.

Model Term

The term in the model.

Numerator DF

The degrees of freedom associated with the between group term. This is one less than the number of groups.

Denominator DF

The degrees of freedom associated with error (within group) term. This value is calculated to make the approximation of the F-distribution to the distribution of the F-Ratio as close as possible.

F-Ratio

A test statistic that has an approximate F distribution with the degrees of freedom indicated to the left.

P-Value

The significance level of the F-ratio to the left. The probability of an F-ratio larger than that obtained by this analysis. For example, to test at an alpha level of 0.05, this probability would have to be less than 0.05 to make the F-ratio significant.

One-Way Analysis of Variance

Reject Equal Means at $\alpha = 0.05$?

This column indicates whether the null hypothesis of mean equality can be rejected at the indicated significance level. If the result is a 'No,' it might be because the means are about the same or it might be because the sample size is too small.

Kruskal-Wallis One-Way ANOVA on Ranks**Kruskal-Wallis One-Way ANOVA on Ranks**

Hypotheses: H0: All medians are equal.
H1: At least two medians are different.

Test Results

Method	DF	Chi ² (H)	P-Value	Reject H0 at $\alpha = 0.05$?
Not Corrected for Ties	2	11.2674	0.00358	Yes
Corrected for Ties	2	11.2708	0.00357	Yes
Number of Sets of Ties	4			
Multiplicity Factor	24			

Group Details

Group	Count	Ranks		Z-Value	Median
		Sum	Mean		
YldA	13	229.5	17.65	-1.4941	554
YldB	16	279.0	17.44	-1.8342	546
YldC	14	437.5	31.25	3.3564	752

This test is a nonparametric substitute for the one-way ANOVA when the assumption of normality is not valid. When in doubt about normality, play it safe and use this test. The assumptions for this test were given in the "Assumptions" section at the beginning of this chapter. Two key assumptions that we remind you of is that the distributions are at least ordinal in nature and that they are identical, except for location. This means that ties (repeated values) are not acceptable. When ties are present in your data, you should use the corrected version of this test. We next present the individual definitions of items on this report.

Hypotheses

The null hypothesis is that the medians are equal versus the alternative that at least one median is different from the rest.

Method

The results of two tests are presented. The first line gives the Kruskal-Wallis test with no correction for ties. The second line reports a modified Kruskal-Wallis test that has been modified to adjust for ties. If there are no ties, the results are identical.

DF

The degrees of freedom of the large sample chi-squared approximation to the Kruskal-Wallis test distribution. Note that the degrees of freedom are equal to the number of groups minus one.

One-Way Analysis of Variance

Chi² (H)

The value of H, the uncorrected (for ties) Kruskal-Wallis test statistic.

P-Value

The significance level of H assuming a chi-squared distribution. The probability of an H larger than that obtained by this analysis. For example, to test at an alpha level of 0.05, this probability would have to be less than 0.05 to make H significant.

Reject H0 at $\alpha = 0.05$?

States whether or not the null hypothesis is rejected at the specified alpha level.

Number of Sets of Ties

This is the number of sets of tied values. If there are no ties, this number is zero. A set of ties is two or more observations with the same value.

Multiplicity Factor

This is the tie portion of the correction factor for H.

$$\sum_{j=1} (t_j^3 - t_j)$$

Normal Scores Tests

Normal Scores Tests

Hypotheses: H0: All group data distributions are the same.
H1: At least one group has observations that tend to be greater than those of the other groups.

Test	DF	Chi ² (H)	P-Value	Reject H0 at $\alpha = 0.05$?
Terry-Hoeffding - Expected Normal Scores	2	10.9026	0.00429	Yes
Van der Waerden - Normal Quantiles	2	10.9819	0.00412	Yes

These normal score tests are similar to the Kruskal-Wallis test, except that instead of being calculated on the ranks of the data, they are calculated on the normal scores. The Terry-Hoeffding (or Fisher-Yates-Terry-Hoeffding) test replaces the data values with the expected value of the normal order statistics. The van der Waerden test replaces the data values with the inverse normal scores (the $k/(n+1)$ th quantile). These tests are equivalent in large samples.

The amazing feature of these tests is that, unlike the Kruskal Wallis test, their asymptotic relative efficiency is one when the underlying data distribution is normal. That is, their power is as good as that of the F-test when the data are normal. When they data are not normal, their efficiency is often much better than the F-test.

One-Way Analysis of Variance

Hypotheses

The null hypothesis is that the group distributions are equal versus the alternative that at least one group tends to have greater data values.

Method

This is the normal scores method that is used.

DF

The degrees of freedom of the large sample Chi-squared approximation to the test distribution. Note that the degrees of freedom are equal to the number of groups minus one.

Chi² (H)

The chi-squared value. This is the test statistic.

P-Value

The significance level of H assuming a chi-squared distribution. The probability of an H larger than that obtained by this analysis. For example, to test at an alpha level of 0.05, this probability would have to be less than 0.05 to make H significant.

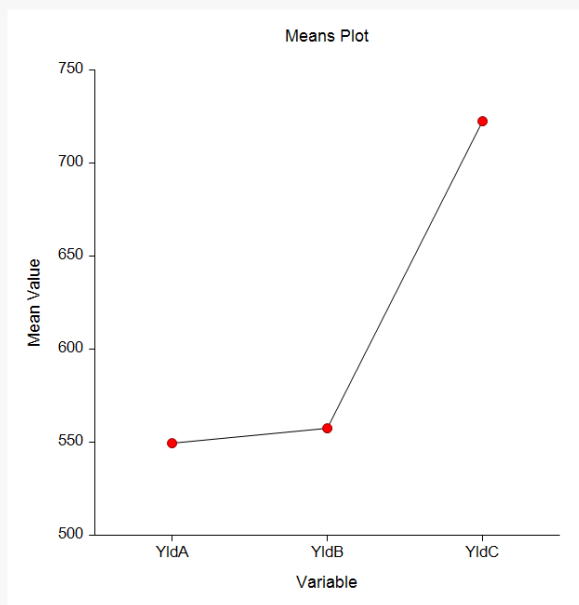
Reject H_0 at $\alpha = 0.05$?

The decision about the null hypothesis based on this test.

Descriptive Statistics and Means Plot
Descriptive Statistics

Group	Count (ni)	Mean	Effect	Median	Standard Deviation	Standard Error $\sqrt{(MSE/ni)}$
All	43	608.7209	609.7473			
YldA	13	549.3846	-60.36264	554	168.7629	37.17356
YldB	16	557.5	-52.24725	546	104.6219	33.50779
YldC	14	722.3571	112.6099	752	127.8873	35.82134

One-Way Analysis of Variance

Means Plot**Group**

The label for this line of the report.

Count (ni)

The number of observations in the mean.

Mean

The value of the sample mean.

Effect

The component that this term contributes to the mean. For example, the mean of the first group is equal to the sum of the overall effect (from the "All" line) plus the effect of the first term.

Median

The value of the sample median.

Standard Deviation

The sample standard deviation of the data in this group.

Standard Error

The standard error of the mean used in the multiple comparison tests. Note that the standard errors are simply the square root of the mean square error divided by the count. These standard errors are not the same as the simple standard errors calculated separately for each group.

Means Plot

This plot displays the means for the data analyzed. Note how easily you can see patterns in the plot.

Multiple Comparison Reports

Duncan's Multiple Comparison Test

Comparison Variables: YldA, YldB, YldC

Group	Count	Mean	Different From Groups
YldA	13	549.3846	YldC
YldB	16	557.5	YldC
YldC	14	722.3571	YldA, YldB

$\alpha = 0.05$, Error Term = S(A), DF = 40, MSE = 17964.36

Notes:

This report provides multiple comparison tests for all pairwise differences between the means. According to Hsu(1996, page 130), the specified family-wise error rate (alpha) is overstated and the Tukey-Kramer method is recommended instead.

This section presents the results of the multiple-comparison procedures selected. These reports all use a uniform format that will be described by considering Duncan's Multiple-Comparison Test. The reports for the other procedures are similar. For more information on the interpretation of the various multiple-comparison procedures, turn to the section by that name.

We next present the individual definitions of items on this report.

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.

One-Way Analysis of Variance

DF

The degrees of freedom of the error term.

MSE

The value of the mean square error.

Planned Comparison Reports

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

Planned Comparison: Custom Comparison 1

Comparison Variables: YldA, YldB, YldC

Test Results

Comparison Value	Standard Error	T-Statistic	Two-Sided P-Value	Reject H0 at $\alpha = 0.05$?	95% Confidence Interval Limits for the Comparison Value	
					Lower	Upper
181.0879	89.06983	2.0331	0.04872	Yes	1.071063	361.1048

$\alpha = 0.05$, Error Term = S(A), DF = 40, MSE = 17964.36

Group Details

Group	Comparison Coefficient	Count	Mean
YldA	-2	13	549.3846
YldB	1	16	557.5
YldC	1	14	722.3571

Notes:
This section presents the results of the custom planned comparisons.

Comparison Value

The value of the comparison. This is formed by the multiplying the Comparison Coefficient times 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).

One-Way Analysis of Variance

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 i^{th} group, M_i is the mean of the i^{th} group, and n_i is the sample size of the i^{th} 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 $\alpha = 0.05$?

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.

Kruskal-Wallis Multiple Comparison Z-Value Test (Dunn's Test)

Kruskal-Wallis Multiple Comparison Z-Value Test (Dunn's Test)

Variable	YIdA	YIdB	YIdC
YIdA	0.0000	0.0462	2.8117
YIdB	0.0462	0.0000	3.0063
YIdC	2.8117	3.0063	0.0000

Regular Test: Medians significantly different if z-value > 1.9600.
 Bonferroni Test: Medians significantly different if z-value > 2.3940.

$\alpha = 0.05$

Z-Values

The values in the table are appropriate for testing whether the medians of any two groups are significantly different. The formula for z_{ij} (comparing group i to group j) is

$$z_{ij} = \frac{\left| \frac{R_i}{n_i} - \frac{R_j}{n_j} \right|}{\sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}}$$

In the presence of ties, the adjusted formula is

$$z_{ij} = \frac{\left| \frac{R_i}{n_i} - \frac{R_j}{n_j} \right|}{\sqrt{\frac{[N(N^2 - 1) - (\sum t^3 - \sum t)] \left[\frac{1}{n_i} + \frac{1}{n_j} \right]}{12(N - 1)}}}$$

where N is the total sample size, n_i is the sample size of the i^{th} group, t is the number of values in the combined sample that are tied at a given rank, and R_i is the sum of the ranks of the i^{th} group.

The distribution of z_{ij} is normal with mean equal to zero and variance equal to one. If you are only making one or two tests, you would compare the value in the table to the Regular Test value, $z_{\alpha/2}$. If the computed z_{ij} is greater than this value, the two groups are significantly different.

However, if you are using all the tests from the table, you should use the Bonferroni Test value. This is a z-value that has been adjusted for multiple tests by dividing $\alpha/2$ by $k(k - 1)/2$, making it $z_{\alpha/(k(k-1))}$. Note the $k(k - 1)/2$ is the number of possible pairs of k groups.

If you are making a specific number of tests, say m , that is less than all-possible pairs, you will have to manually make the correct adjustment by dividing $\alpha/2$ by m . This might happen if you are comparing each treatment group with a control group, in which case you would have $k - 1$ tests.

One-Way ANOVA Checklist

This checklist, prepared by a professional statistician, is a flowchart of the steps you should complete to conduct a valid one-way ANOVA (or its nonparametric counterpart). You should complete these tasks in order.

Step 1 – Data Preparation

Introduction

This step involves scanning your data for anomalies, data entry errors, typos, and so on.

Sample Size

The sample size (number of nonmissing rows) has a lot of ramifications. The one-way ANOVA was developed under the assumption that the sample sizes in each group are equal. In practice, this seldom happens, but the closer you can get to equal sample sizes the better.

With regard to the combined sample size, the ANOVA may be performed on very small samples, such as 4 or 5 observations per group. However, in order to test assumptions and obtain reliable estimates of variation, you should attempt to obtain at least 30 individuals per group.

Missing Values

The number and pattern of missing values are always issues to consider. Usually, we assume that missing values occur at random throughout your data. If this is not true, your results will be biased since a particular segment of the population is underrepresented. If you have a lot of missing values, some researchers recommend comparing other variables with respect to missing versus nonmissing. If you find large differences in other variables, you should begin to worry about whether the missing values are cause for a systematic bias in your results.

Type of Data

The mathematical basis of the F-test assumes that the data are continuous. Because of the rounding that occurs when data are recorded, all data are technically discrete. The validity of the assumption of the continuity of the data then comes down to determining when we have too much rounding. For example, most statisticians would not worry about human-age data that were rounded to the nearest year. However, if these data were rounded to the nearest ten years or further to only three groups (young, adolescent, and adult), most statisticians question the validity of the probability statements. Some studies have shown that the F-test is reasonably accurate when the data have only five possible values (most would call this discrete data). If your data contain fewer than five unique values, any probability statements made are tenuous.

Outliers

Generally, outliers cause distortion in F-tests. You must scan your data for outliers (the box plot is an excellent tool for doing this). If you have outliers, you have to decide if they are one-time occurrences or if they would occur in another sample. If they are one-time occurrences, you can remove them and proceed. If you know they represent a certain segment of the population, you have to decide between biasing your results (by removing them) or using a nonparametric test that can deal with them. Most would choose the nonparametric test.

Step 2 – Setup and Run the Panel

Introduction

Now comes the fun part: running the program. **NCSS** is designed to be simple to operate, but it can still seem complicated. Before you run a procedure such as this for the first time, take a few minutes to read through the chapter again and familiarize yourself with the issues involved.

Enter Variables

The **NCSS** procedures were set with ready-to-run defaults. About all you have to do is select the appropriate variables (columns of data).

Select All Plots

As a rule, you should select all diagnostic plots, even though they may take a few extra seconds to generate. They add a great deal to your analysis of the data.

Specify Alpha

Most beginners at statistics forget this important step and let the alpha value default to the standard 0.05. You should make a conscious decision as to what value of alpha is appropriate for your study. The 0.05 default came about during the dark ages when people had to rely on printed probability tables and there were only two values available: 0.05 or 0.01. Now you can set the value to whatever is appropriate.

A special note on setting the Multiple Comparison alpha. We suggest that you set this at 0.10 so that the individual tests are made at a more reasonable significance level.

Step 3 – Check Assumptions

Introduction

Once the output is displayed, you will be tempted to go directly to the probability of the F-test, determine if you have a significant result, and proceed to something else. However, it is very important that you proceed through the output in an orderly fashion. The first task is to determine if the assumptions are met by your data.

Sometimes, when the data are nonnormal for all samples, a data transformation (like square roots or logs) might normalize the data. Frequently, when one sample is normal and the other is not, this transformation, or re-expression, approach works well.

Random Sample

The validity of this assumption depends upon the method used to select the sample. If the method used assures that each individual in the population of interest has an equal probability of being selected for this sample, you have a random sample. Unfortunately, you cannot tell if a sample is random by looking at it or statistics from it.

Sample Independence

The samples must be independent. For example, if you randomly divide a sample of individuals into two groups, you have met this requirement. However, if your population consists of cars and you assign the left tire to one group and the right tire to the other, you do not have independence. Here again, you cannot tell if the samples are independent by looking at them. You must consider the sampling methodology.

Check Means Report

You should check the Means and Effects Section first to determine if the Counts and the Means are reasonable. If you have selected the wrong variable, these values will alert you.

Normality

To validate this assumption, you should first look at the plots. Outliers will show up on the box plots and the probability plots. No data will be perfectly normal. After considering the plots, look at the Tests of Assumptions section to get numerical confirmation of what you see in the plots. Remember that the power of these normality tests is directly related to the sample size, so when the normality assumption is accepted, double check that your sample is large enough to give conclusive results.

Equal Variance

The equal variance assumption is important in determining which statistical test to use. Check the box plots for boxes with about the same widths. Confirm your conclusion by looking at the Equal-Variance Test (Modified Levene) line.

Step 4 – Choose the Appropriate Statistical Test

Introduction

You are now ready to determine which statistical procedures will be valid.

Normal Data with Equal Variances

Use the Analysis of Variance Section for hypothesis testing.

Normal Data with Unequal Variances

You can use Welch's test of try variance stabilizing transformations like the log or square root. The Kruskal-Wallis tests assumes that the variances are equal, so it cannot be used.

Nonnormal Data with Equal Variances

Use the Kruskal-Wallis, van der Waerden, or Terry-Hoeffding test for hypothesis testing.

Nonnormal Data with Unequal Variances

If you cannot find a variance-stabilizing transformation, you might test each pair of groups using the Kolmogorov-Smirnov test. Of course, the Kolmogorov-Smirnov test tests both the mean and variance. Since you already know that the variances are different from the Levene test, it is questionable whether this test will add new information. If you decide to make several Kolmogorov-Smirnov tests, you should make appropriate adjustments to your significance level to avoid the multiplicity problem discussed in the Multiple Comparison section.

Step 5 – Interpret Findings

Hypothesis Testing

The interpretation of an analysis of variance table is rather easy. You simply look at the *Prob>F* value. If this value is less than your chosen significance level (say 0.05), you can declare that at least two of the means are significantly different. You then determine which means are different using planned comparisons or an appropriate paired-comparison procedure. With a list of significantly different means, you can view the plot of the means and discuss the meaning of your results.

Step 6 – Record Your Results

Finally, as you finish a test, take a moment to jot down what decisions you made and what you have found. Explain what you did, why you did it, what conclusions you reached, which outliers you deleted, areas for further investigation, and so on.