

One-Way ANOVA

Summary

The **One-Way ANOVA** procedure is designed to construct a statistical model describing the impact of a single categorical factor X on a dependent variable Y. Tests are run to determine whether or not there are significant differences between the means, variances, and/or medians of Y at the different levels of X. In addition, the data may be displayed graphically in various ways, including a multiple scatterplot, a means plot, an ANOM plot, and a medians plot.

In this procedure, it is assumed that the data will be placed in two columns, one for the dependent variable Y and a second identifying the levels of X. A one-way analysis of variance can also be performed using the *Multiple Samples Comparison* procedure, which should be used if the data have been placed into separate columns for each level of X.

Sample StatFolio: *oneway.sgp*

Sample Data:

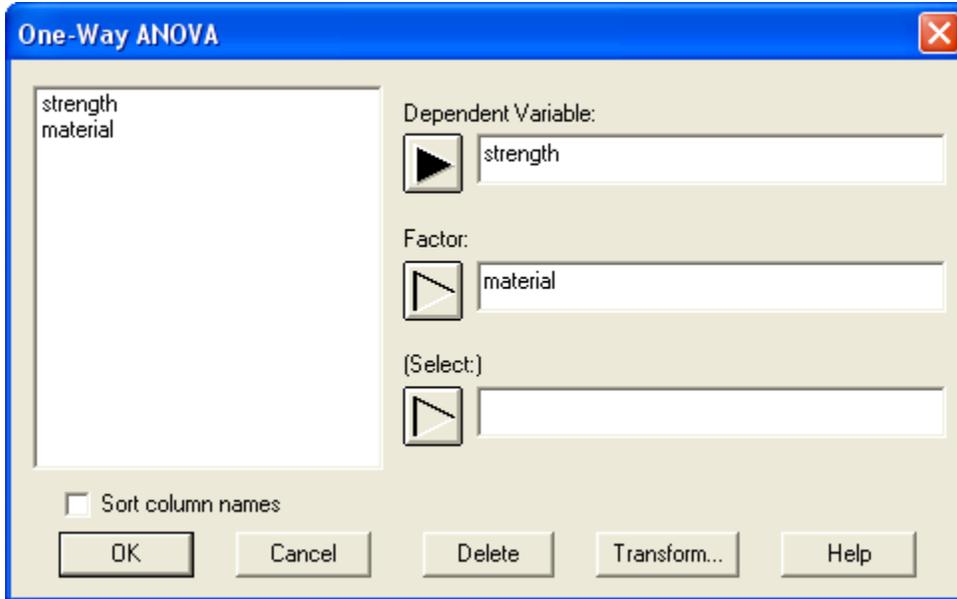
The file *breaking.sgd* contains the results of an experiment in which the breaking strength of widgets was compared amongst 4 different materials. A portion of the data is shown below:

| <i>Strength</i> | <i>Material</i> |
|-----------------|-----------------|
| 39 | A |
| 57 | A |
| 42 | A |
| 32 | A |
| 43 | A |
| 50 | A |
| 31 | A |
| 51 | A |
| 27 | B |
| 43 | B |
| 25 | B |
| 11 | B |
| 39 | B |
| ... | ... |

The entire data set consists of $n = 32$ widgets, 8 of which were made from each of $q = 4$ different materials.

Data Input

The data input dialog expects requests the names of the columns containing the measurements Y and the levels of the factor X :



- **Data:** numeric column containing the n observations for Y .
- **Factor:** numeric or non-numeric column containing an identifier for the levels of the factor X .
- **Select:** subset selection.

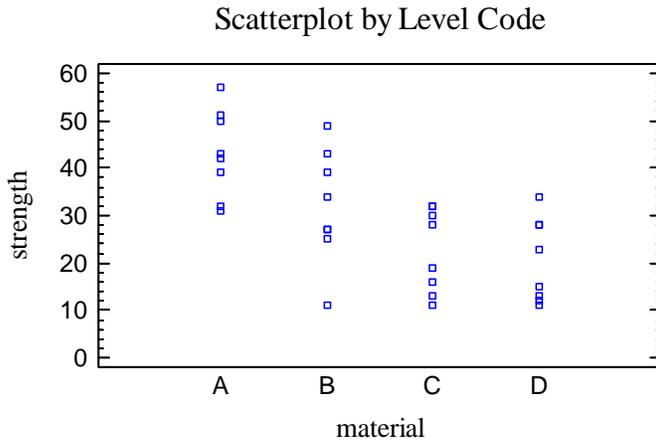
Analysis Summary

The *Analysis Summary* shows the number of levels of X and the total number of observations n .

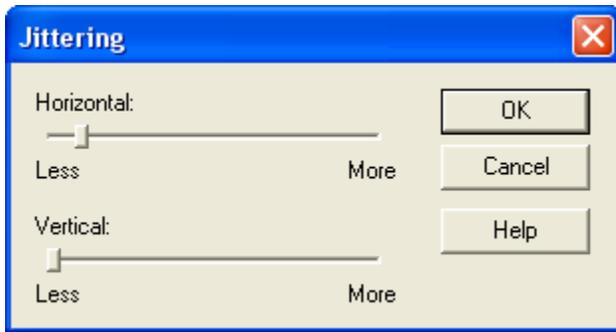
| |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><u>One-Way ANOVA - strength by material</u> Dependent variable: strength Factor: material Number of observations: 32 Number of levels: 4</p> |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Scatterplot

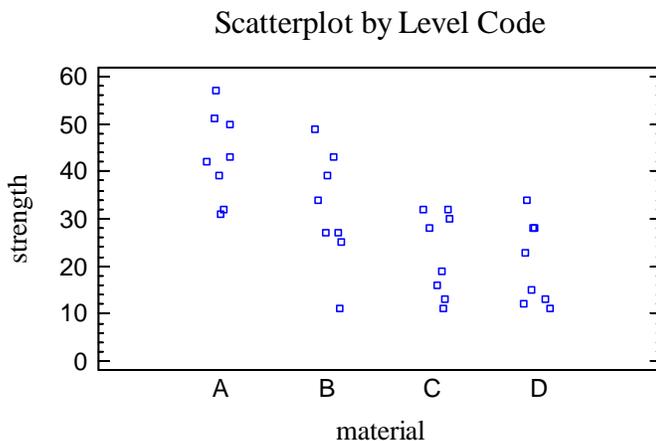
The *Scatterplot* pane plots the data by levels of the factor X.



If there are many common values, you may wish to add a small amount of horizontal jitter to the plot by pressing the *Jitter* button on the analysis toolbar:



This offsets each point randomly in the horizontal direction so that identical values do not plot on top of each other:



The above plot suggests that there are differences in breaking strength amongst the four materials.

Summary Statistics

The *Summary Statistics* pane calculates a number of different statistics that are commonly used to summarize a sample of variable data:

| Summary Statistics for strength | | | | | | |
|---------------------------------|-------|---------|--------------------|-------|---------------|---------------|
| material | Count | Average | Standard deviation | Range | Std. skewness | Std. kurtosis |
| A | 8 | 43.125 | 9.18753 | 26.0 | 0.0718672 | -0.59635 |
| B | 8 | 31.875 | 11.9695 | 38.0 | -0.369758 | 0.0369302 |
| C | 8 | 22.625 | 8.81456 | 21.0 | -0.191636 | -1.24894 |
| D | 8 | 20.5 | 8.86405 | 23.0 | 0.397909 | -1.0005 |
| Total | 32 | 29.5313 | 13.0062 | 46.0 | 0.530301 | -0.827139 |

Most of the statistics fall into one of three categories:

1. measures of *central tendency* – statistics that characterize the “center” of the data.
2. measure of *dispersion* – statistics that measure the spread of the data.
3. measures of *shape* – statistics that measure the shape of the data relative to a normal distribution.

The statistics included in the table by default are controlled by the settings on the *Stats* pane of the *Preferences* dialog box. These selections may be changed using *Pane Options*. For a detailed description of each statistic, see the *One Variable Analysis* documentation.

Of particular interest are:

1. *Sample means* \bar{Y}_j : the average breaking strength for each material.
2. *Sample standard deviations* s_j : the standard deviations for each material.
3. *Standardized skewness and kurtosis*: These statistics should be between -2 and $+2$ if the data come from normal distributions.

For the widgets, the average breaking strength was highest for material A, followed by material B. All of the standardized skewness and kurtosis statistics are within the range expected for data from normal distributions.

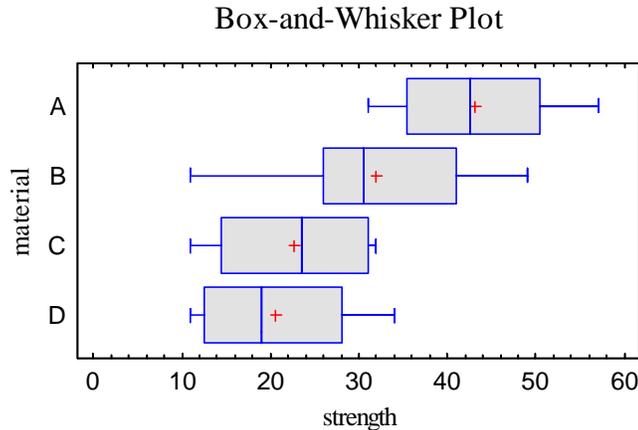
Pane Options



Select the desired statistics.

Box-and-Whisker Plot

This pane displays a box-and-whisker plot for Y at each level of X.

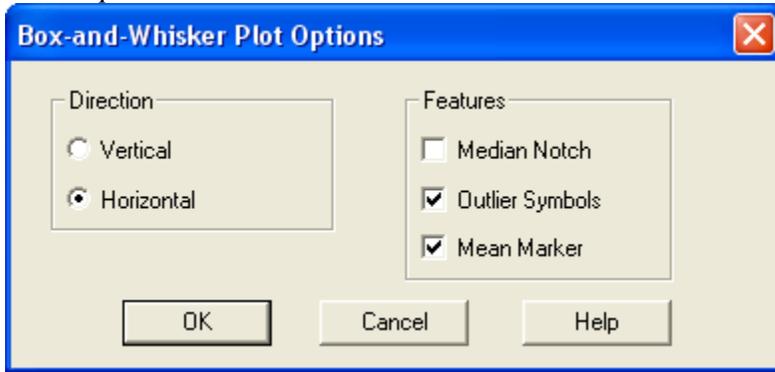


Box-and-whisker plots are constructed in the following manner:

- A box is drawn extending from the *lower quartile* of the sample to the *upper quartile*. This is the interval covered by the middle 50% of the data values when sorted from smallest to largest.
- A vertical line is drawn at the *median* (the middle value).
- If requested, a plus sign is placed at the location of the sample mean.
- Whiskers are drawn from the edges of the box to the largest and smallest data values, unless there are values unusually far away from the box (which Tukey calls *outside points*). Outside points, which are points more than 1.5 times the interquartile range (box width) above or below the box, are indicated by point symbols. Any points more than 3 times the interquartile range above or below the box are called *far outside points*, and are indicated by point symbols with plus signs superimposed on top of them. If outside points are present, the whiskers are drawn to the largest and smallest data values which are not outside points.

In the sample data, the variability appears to be similar within each material, although the locations show some differences. There are no outside points.

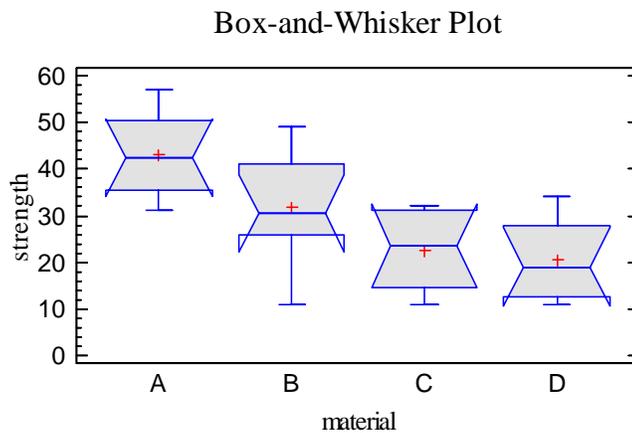
Pane Options



- **Direction:** the orientation of the plot, corresponding to the direction of the whiskers.
- **Median Notch:** If selected, a notch will be added to the plot showing the estimation error associated with each median. The notches are scaled in such a way that, for samples of equal size, if they do not overlap, the two medians are significantly different at the default system confidence level (set on the *General* tab of the *Preferences* dialog box on the *Edit* menu).
- **Outlier Symbols:** if selected, indicates the location of outside points.
- **Mean Marker:** if selected, shows the location of the sample mean as well as the median.

Example – Notched Box-and-Whisker Plot

The following plot adds median notches at the 95% confidence level.



Each notch covers the interval

$$\tilde{x}_j \pm \frac{z_{\alpha/2}}{2} \frac{1.25(IQR_j)}{1.35\sqrt{n_j}} \left(1 + \frac{1}{\sqrt{2}}\right) \tag{1}$$

where \tilde{x}_j is the median of the j -th sample, IQR_j is the sample interquartile range, n_j is the sample size, and $z_{\alpha/2}$ is the upper $(\alpha/2)\%$ critical value of a standard normal distribution. In cases where the sample size is small, the notch may extend beyond the box, resulting in a folding back appearance.

In the sample data, the notch for material A does not overlap the notch for either materials C or D, indicating that the median is significantly higher for A than C and D. However, it is not significantly higher than material B since the notches for A and B overlap.

ANOVA Table

In order to determine whether or not the means of the q groups are significantly different from each other, a oneway analysis of variance can be performed. The results are displayed in the *ANOVA Table*:

| ANOVA Table for strength by material | | | | | |
|--------------------------------------|----------------|----|-------------|---------|---------|
| Source | Sum of Squares | Df | Mean Square | F-Ratio | P-Value |
| Between groups | 2556.34 | 3 | 852.115 | 8.88 | 0.0003 |
| Within groups | 2687.63 | 28 | 95.9866 | | |
| Total (Corr.) | 5243.97 | 31 | | | |

The table divides the overall variability among the n measurements into two components:

1. A “within groups” component, which measures the variability amongst breaking strengths of widgets made of the same material.
2. A “between groups” component, which measures the variability amongst widgets made of different materials.

Of particular importance is the F-ratio, which tests the hypothesis that the mean response for all levels of X is the same. Formally, it tests the null hypothesis

$$H_0: \mu_1 = \mu_2 = \dots = \mu_q$$

versus the alternative hypothesis

$$H_A: \text{not all } \mu_j \text{ equal}$$

If F is sufficiently large, the null hypothesis is rejected.

The statistical significance of the F-ratio is most easily judged by its P-value. If the P-value is less than 0.05, the null hypothesis of equal means is rejected at the 5% significance level, as in the current example. This does not imply that every mean is significantly different from every other mean. It simply implies that the means are not all the same. Determining which means are significantly different from which others requires additional tests, as discussed below.

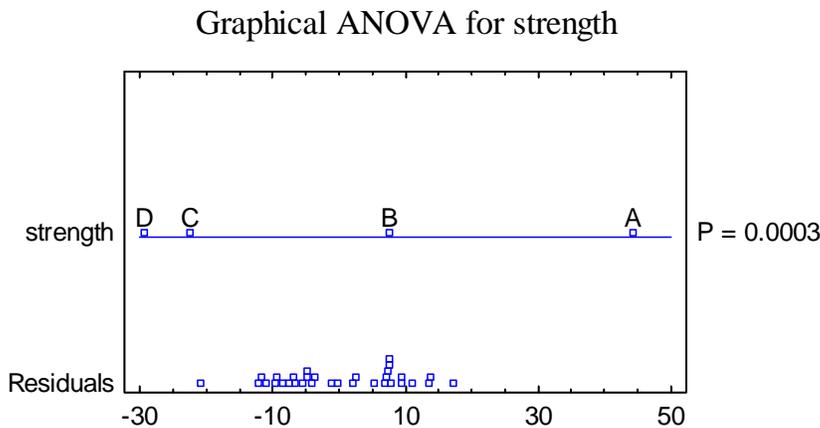
Graphical ANOVA

The *Graphical ANOVA* plot, developed by Hunter (2005), is a technique for displaying graphically the importance of the differences between levels of the experimental factor. It is a plot of the scaled effects of the factor, where the “effect” equals the difference between the mean for a level of that factor and the estimated grand mean. Each of the effects is multiplied by a scaling factor

$$\sqrt{\frac{\nu_R n_i}{\nu_T \bar{n}}} \tag{2}$$

where ν_R is the residual degrees of freedom, ν_T is the degrees of freedom for the factor, n_i equals the number of observations in the i -th level of the factor, and \bar{n} is the average number of observations at all levels of the factor. This scales the effects so that the natural variance of the points in the diagram is comparable to that of the residuals, which are displayed at the bottom of the plot.

The plot for the sample data is shown below:



Along the right-hand side of the display is the P-Value for between group differences, taken from the ANOVA table.

By comparing the variability amongst the factor effects in the above plot to that of the residuals, it is easy to see that the differences are of a greater magnitude than could be accounted for solely by experimental error. Depending upon the relative location of the effects, it may also be possible in some cases to visually identify which levels are significantly different from which other levels, which is done formally by the *Multiple Range Tests* described below.

Multiple Range Tests

To determine which sample means are significantly different from which others, the *Multiple Range Tests* can be performed:

| Multiple Range Tests for strength by material | | | |
|-----------------------------------------------|-------|--------|--------------------|
| Method: 95.0 percent LSD | | | |
| material | Count | Mean | Homogeneous Groups |
| D | 8 | 20.5 | X |
| C | 8 | 22.625 | XX |
| B | 8 | 31.875 | X |
| A | 8 | 43.125 | X |

| Contrast | Sig. | Difference | +/- Limits |
|----------|------|------------|------------|
| A - B | * | 11.25 | 10.0344 |
| A - C | * | 20.5 | 10.0344 |
| A - D | * | 22.625 | 10.0344 |
| B - C | | 9.25 | 10.0344 |
| B - D | * | 11.375 | 10.0344 |
| C - D | | 2.125 | 10.0344 |

* denotes a statistically significant difference.

The top half of the table displays each of the estimated sample means in increasing order of magnitude. It shows:

- **Count** - the number of observations n_j .
- **Mean** - the estimated sample mean \bar{Y}_j .
- **Homogeneous groups** - a graphical illustration of which means are significantly different from which others, based on the contrasts displayed in the second half of the table. Each column of X's indicates a group of means within which there are no statistically significant differences. For example, the first column in the above table contains an X for materials C and D, indicating that their means are not significantly different. Likewise, materials B and C show no significant differences. The mean of material A, on the other hand, is significantly greater than the mean of any other material.
- **Difference** - the difference between the two sample means

$$\hat{\Delta}_{j_1j_2} = \bar{Y}_{j_1} - \bar{Y}_{j_2} \tag{3}$$

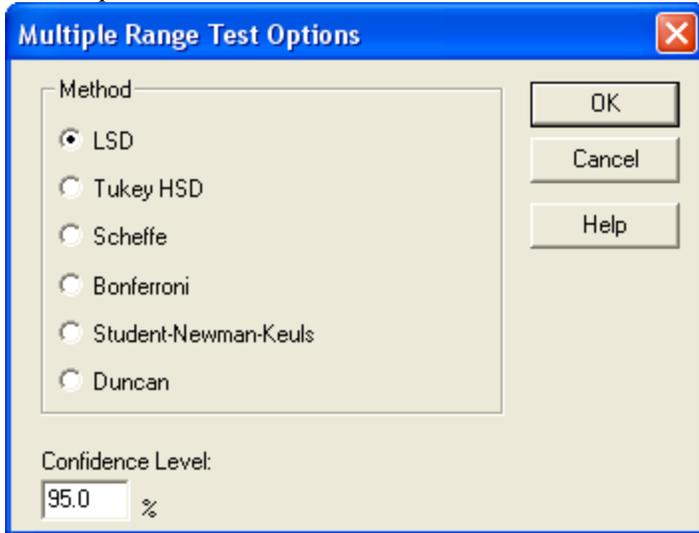
- **Limits** - an interval estimate of that difference, using the currently selected multiple comparisons procedure:

$$\hat{\Delta}_{j_1j_2} \pm M \sqrt{MS_{within} \left(\frac{1}{n_{j_1}} + \frac{1}{n_{j_2}} \right)} \tag{4}$$

where M is a constant that depends upon the procedure selected.

- **Sig.** - An asterisk is placed next to any difference that is statistically significantly different from 0 at the currently selected significance level, i.e., any interval that does not contain 0.

Pane Options



- **Method:** the method used to make the multiple comparisons.
- **Confidence Level:** the level of confidence used by the selected multiple comparison procedure.

The available methods are:

- **LSD** - forms a confidence interval for each pair of means at the selected confidence level using:

$$M = t_{\alpha/2, n-q} \tag{5}$$

where t represents the value of Student's t distribution with $n - q$ degrees of freedom leaving an area of $\alpha/2$ in the upper tail of the curve. This procedure is due to Fisher and is called the *Least Significant Difference* procedure, since the magnitude of the limits indicates the smallest difference between any two means that can be declared to represent a statistically significant difference. It should only be used when the F-test in the ANOVA table indicates significant differences amongst the sample means. The probability of making a Type I error α applies to each pair of means separately. If making more than one comparison, the overall probability of calling at least one pair of means significantly different when they are not may be considerably larger than α .

- **Tukey HSD** - widens the intervals to allow for multiple comparisons amongst all pairs of means, using:

$$M = T_{\alpha/2, q, n-q} \tag{6}$$

which uses Tukey's T instead of Student's t. Tukey's T is equal to $(1/\sqrt{2})$ times the Studentized range distribution, which is tabulated in books such as Neter et al. (1996). Tukey called his procedure the *Honestly Significant Difference* procedure since it controls the experiment-wide error rate at α . If all of the means are equal, the probability of declaring *any* of the pairs to be significantly different in the entire experiment equals α . Tukey's procedure is more conservative than Fisher's LSD procedure, since it makes it harder to declare any particular pair of means to be significantly different.

- **Scheffe** - designed to permit the estimation of all possible contrasts amongst the sample means (not just pairwise comparisons). It uses a multiple related to the F distribution:

$$M = \sqrt{(q-1)F_{\alpha, q-1, n-q}} \quad (7)$$

In the current instance, this procedure is likely to be very conservative, since only pairs are being estimated.

- **Bonferroni** - designed to permit the estimation of any preselected number of contrasts. In this case, it uses a multiple equal to

$$M = t_{\alpha/(q(q-1)), n-q} \quad (8)$$

since $q(q-1)/2$ pairwise differences are being estimated. These limits are usually wider than Tukey's limits when all pairwise comparisons are being made.

- **Student-Newman-Keuls** - Unlike the previous methods, this method does not create intervals for the pairwise differences. Instead, it sorts the means in increasing order and then begins to separate them into groups according to values of the Studentized range distribution. Eventually, the means are separated into homogeneous groups within which there are no significant differences.
- **Duncan** - similar to the Student-Newman-Keuls procedure, except that it uses a different critical value of the Studentized range distribution when defining the homogeneous groups. A detailed discussion of the Duncan and Student-Newman-Keuls procedures is given by Milliken and Johnson (1992).

The choice between the LSD procedure and a multiple comparisons procedure such as Tukey's HSD should depend on the relative cost of making a Type I error (calling a pair of means different when they're really not) versus the cost of making a Type II error (not calling a pair of means different when they really are). In early stages of an investigation, one may not want to be as conservative as when final verifications are being made.

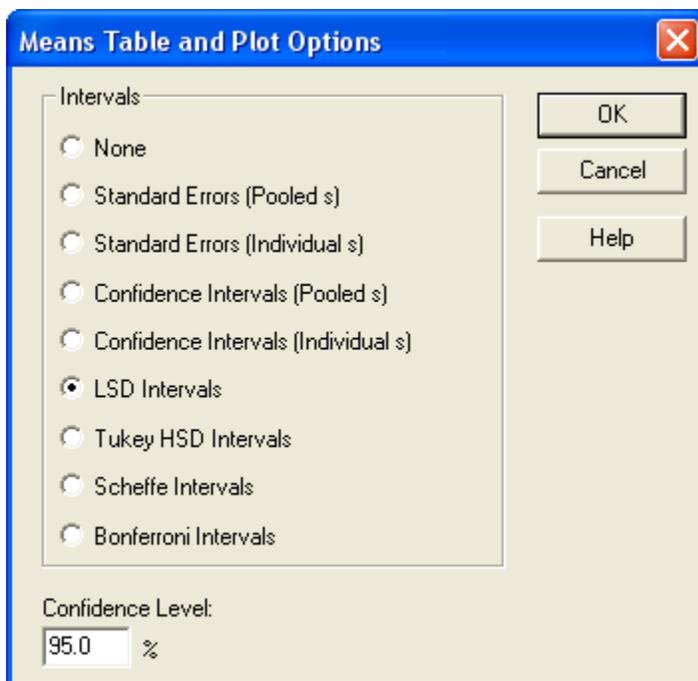
Table of Means

This table displays each level mean together with an uncertainty interval:

| | | | <i>Smd. error</i> | | |
|-----------------|--------------|-------------|-------------------|--------------------|--------------------|
| <i>material</i> | <i>Count</i> | <i>Mean</i> | <i>(pooled s)</i> | <i>Lower limit</i> | <i>Upper limit</i> |
| A | 8 | 43.125 | 3.46386 | 38.1078 | 48.1422 |
| B | 8 | 31.875 | 3.46386 | 26.8578 | 36.8922 |
| C | 8 | 22.625 | 3.46386 | 17.6078 | 27.6422 |
| D | 8 | 20.5 | 3.46386 | 15.4828 | 25.5172 |
| Total | 32 | 29.5313 | | | |

The type of interval displayed depends on *Pane Options*.

Pane Options



- **Intervals:** the method used to construct the intervals.
- **Confidence Level:** the level of confidence associated with each interval.

The type of intervals that may be selected are:

- **None** - no intervals are displayed.
- **Standard errors (pooled s)** - displays the standard errors using the pooled within-sample standard deviation:

$$\bar{Y}_j \pm \sqrt{\frac{MS_{within}}{n_j}} \tag{9}$$

- **Standard errors (individual s)** - displays the standard errors using the standard deviation of each sample separately:

$$\bar{Y}_j \pm \sqrt{\frac{s_j^2}{n_j}} \quad (10)$$

- **Confidence intervals (pooled s)** - displays confidence intervals for the group means using the pooled within-group standard deviation:

$$\bar{Y}_j \pm t_{\alpha/2, n-q} \sqrt{\frac{MS_{within}}{n_j}} \quad (11)$$

- **Confidence intervals (individual s)** - displays confidence intervals for the sample means using the standard deviation of each group separately:

$$\bar{Y}_j \pm t_{\alpha/2, n_j-1} \sqrt{\frac{s_j^2}{n_j}} \quad (12)$$

- **LSD intervals** - designed to compare any pair of means with the stated confidence level. The intervals are given by

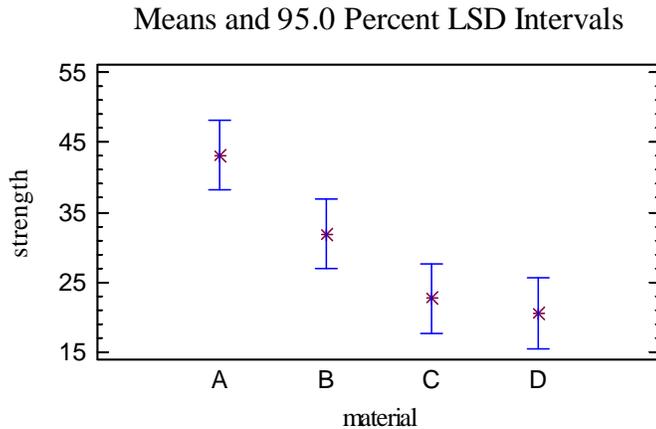
$$\bar{Y}_j \pm \frac{\sqrt{2}M}{2} \sqrt{\frac{MS_{within}}{n_j}} \quad (13)$$

where M is defined as in the *Multiple Range Tests*. This formula also applies to the three selections below.

- **Tukey HSD Intervals** - designed for comparing all pairs of means. The stated confidence level applies to the entire family of pairwise comparisons.
- **Scheffe Intervals** - designed for comparing all contrasts. Not usually relevant here.
- **Bonferroni Intervals** - designed for comparing a selected number of contrasts. Tukey's intervals are usually tighter.

Means Plot

The level means may be plotted together with the uncertainty intervals:



The types of intervals that may be used are the same as for the *Means Table* above.

Provided all of the sample sizes are the same (or close), the analyst can determine which means are significantly different from which others using the LSD, Tukey, Scheffe, or Bonferroni procedure simply by looking at whether or not a pair of intervals overlap in the vertical direction. A pair of intervals that do not overlap indicates a statistically significant difference between the means at the selected confidence level. In this case, note that the interval for material A does not overlap the intervals for any of the other materials.

Variance Check

One of the assumptions underlying the analysis of variance is that the variances of the populations from which the samples come are the same. The *Variance Check* pane performs any of several tests to verify this assumption:

| Variance Check | | | | |
|----------------|----------|----------|----------|---------|
| | Test | P-Value | | |
| Levene's | 0.316515 | 0.813302 | | |
| Comparison | Sigma1 | Sigma2 | F-Ratio | P-Value |
| A / B | 9.18753 | 11.9695 | 0.589181 | 0.5018 |
| A / C | 9.18753 | 8.81456 | 1.08642 | 0.9157 |
| A / D | 9.18753 | 8.86405 | 1.07432 | 0.9271 |
| B / C | 11.9695 | 8.81456 | 1.84394 | 0.4381 |
| B / D | 11.9695 | 8.86405 | 1.82341 | 0.4464 |
| C / D | 8.81456 | 8.86405 | 0.988864 | 0.9886 |

The hypotheses to be tested are:

Null Hypothesis: all σ_j are equal

Alt. Hypothesis: not all σ_j are equal

The four tests are:

1. *Cochran's test*: compares the maximum within-sample variance to the average within-sample variance. A P-value less than 0.05 indicates a significant difference amongst the within-sample standard deviations at the 5% significance level. The test is appropriate only if all group sizes are equal.
2. *Bartlett's test*: compares a weighted average of the within-sample variances to their geometric mean. A P-value less than 0.05 indicates a significant difference amongst the within-sample standard deviations at the 5% significance level. The test is appropriate for both equal and unequal group sizes.
3. *Hartley's test*: computes the ratio of the largest sample variance to the smallest sample variance. This statistic must be compared to a table of critical values, such as the one contained in Neter et al. (1996). For 6 samples and 62 degrees of freedom for experimental error, H would have to exceed approximately 2.1 to be statistically significant at the 5% significance level. Note: this test is only appropriate if the number of observations within each treatment level is the same.
4. *Levene's test*: performs a one-way analysis of variance on the variables

$$Z_{ij} = |y_{ij} - \bar{y}_j| \quad (14)$$

The tabulated statistic is the F statistic from the ANOVA table.

Using Levene's test for the widgets, there is no reason to reject the assumption that the standard deviations are the same for all materials, since the P-value is greater than 0.05. Any apparent differences among the sample standard deviations are not statistically significant at the 5% significance level.

The table also displays the results of a set of two-sample F-tests that compare the standard deviations for each pair of levels. Any pair with a small P-Value would be a pair whose standard deviations were significantly different. In the current example, there are no significant pairs. NOTE: It is recommended that the F-tests be ignored if the initial overall test does not show significant differences amongst the sigmas.

Residual Plots

As with all statistical models, it is good practice to examine the residuals. In a oneway analysis of variance, the residuals are defined by:

$$e_{ij} = y_{ij} - \bar{y}_j \quad (15)$$

i.e., the residuals are the differences between the observed data values and their respective level means.

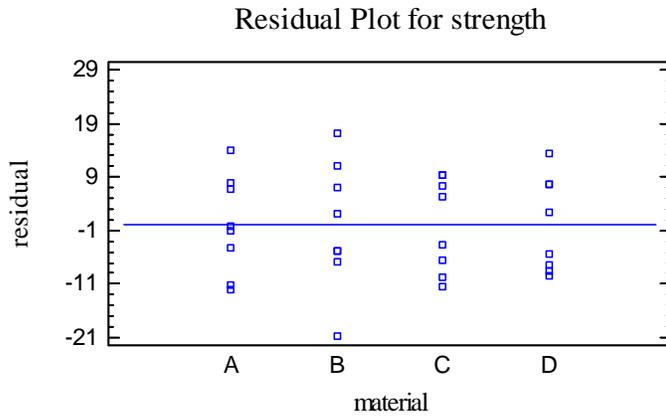
The *Multiple Sample Comparison* procedure creates 3 residual plots:

1. versus factor level.

2. versus predicted value.
3. versus row number.

Residuals versus Factor Level

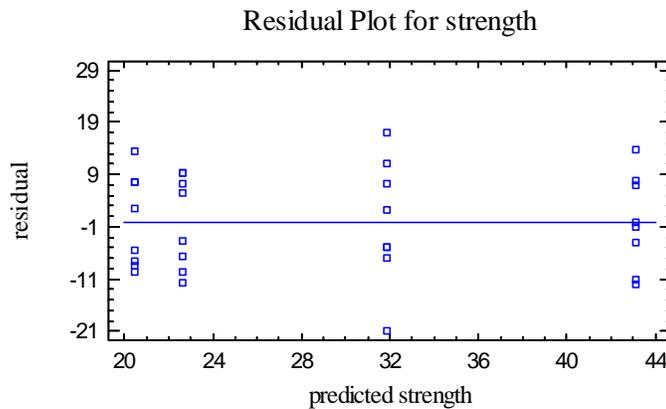
This plot is helpful in visualizing any differences in variability amongst the levels.



The average residual at each level equals 0.

Residuals versus Predicted

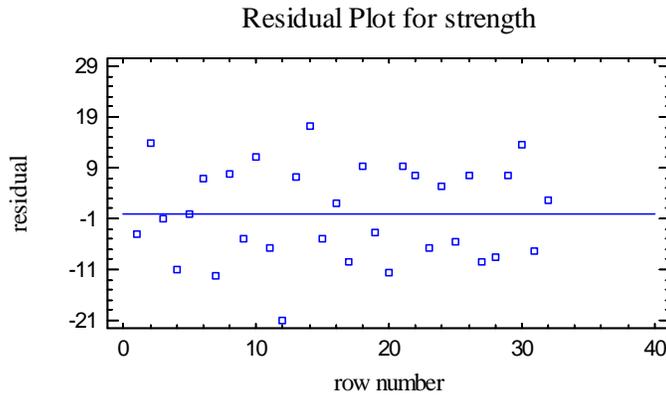
This plot is helpful in detecting any heteroscedasticity in the data.



Heteroscedasticity occurs when the variability of the data changes as the mean changes, and might necessitate transforming the data before performing the ANOVA. It is usually evidenced by a funnel-shaped pattern in the residual plot.

Residuals versus Observation

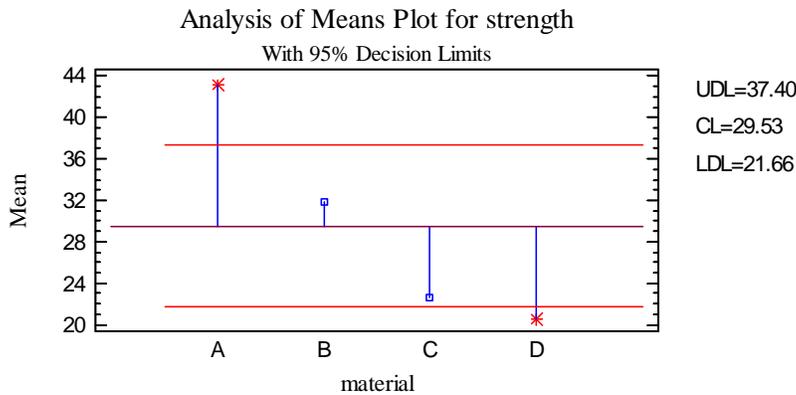
This plot shows the residuals versus row number in the datasheet:



If the data are arranged in chronological order, any pattern in the data might indicate an outside influence. No such pattern is evident in the above plot.

Analysis of Means (ANOM) Plot

If the number of samples is between 3 and 20, a somewhat different approach to the comparison of level means is presented in the *Analysis of Means* or *ANOM Plot*:



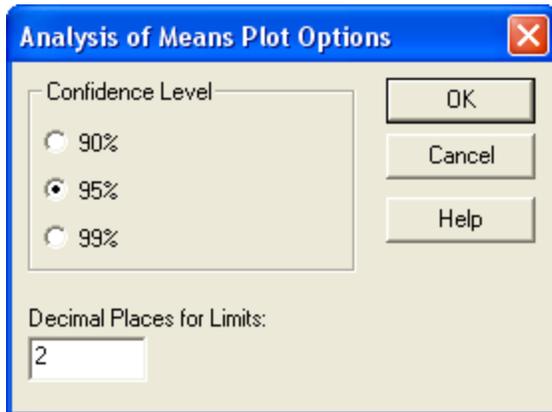
This plot constructs a chart similar to a standard control chart, where each level mean is plotted together with a centerline and upper and lower decision limits. The centerline is located at the grand average of all of the observations \bar{Y} . The decision limits are located at

$$\bar{Y} \pm h_{n-q, 1-\alpha} \sqrt{\frac{MS_{within}}{n_j} \left(\frac{q-1}{q} \right)} \tag{16}$$

where h is a critical value obtained from a table of the multivariate t distribution. The chart tests the null hypothesis that all of the level means are equal to the grand mean. Any means that fall outside the decision limits indicate that the corresponding level differs significantly from that overall mean.

The advantage of the ANOM plot is that it shows at a glance which means are significantly different than the average of all the levels. It also does so using a type of chart with which many engineers and operators are quite familiar. It is easy to see from the above chart that Material A has a significantly higher mean breaking strength than average, while material D has a significantly lower breaking strength. The procedure is exact if all sample sizes are equal and approximate if they don't differ too much.

Pane Options



- **Confidence Level:** level used to position the decision limits.
- **Decimal Places for Limits:** number of decimal places shown when displaying the decision limits.

Kruskal-Wallis Test

An alternative to the standard analysis of variance that compares level *medians* instead of means is the *Kruskal-Wallis Test*. This test is much less sensitive to the presence of outliers than a standard oneway ANOVA and should be used whenever the assumption of normality within levels is not reasonable. It tests the hypotheses:

Null Hypothesis: all level medians are equal

Alt. Hypothesis: not all level medians are equal

The test is conducted by:

1. Sorting all of the n data values from smallest to largest and ranking them, assigning a rank of 1 to the smallest and n to the largest. If any observations are exactly equal, then the tied observations are given a rank equal to the average of the positions at which the tie occurs.
2. Computing the average ranks of the observations at each level \bar{R}_j .
3. Calculating a test statistic to compare the differences amongst the average ranks.

4. Calculating a P-value to test the hypotheses.

The output is shown below:

| Kruskal-Wallis Test for strength by material | | |
|----------------------------------------------|-------------|--------------|
| material | Sample Size | Average Rank |
| A | 8 | 26.125 |
| B | 8 | 17.6875 |
| C | 8 | 12.0625 |
| D | 8 | 10.125 |

Test statistic = 14.0787 P-Value = 0.00279989

Small P-Values (less than 0.05 is operating at the 5% significance level) indicate that there are significant differences amongst the level medians, as in the example above.

Mood's Median Test

Mood's Median Test is another method of determining whether or not the medians of all q materials are equal. It is less sensitive to outliers than the Kruskal-Wallace test, but is also less powerful when the data come from distributions such as the normal. The output is shown below.

| Mood's Median Test for strength by material | | | | | | |
|---------------------------------------------|-------------|-----|----|--------|--------------|--------------|
| Total n = 32 | | | | | | |
| Grand median = 29.0 | | | | | | |
| material | Sample Size | n<= | n> | Median | 95% lower CL | 95% upper CL |
| A | 8 | 0 | 8 | 42.5 | | |
| B | 8 | 4 | 4 | 30.5 | | |
| C | 8 | 5 | 3 | 23.5 | | |
| D | 8 | 7 | 1 | 19.0 | | |

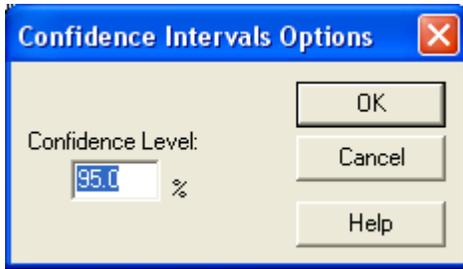
Test statistic = 13.0 P-Value = 0.00463639

Displayed at the top of the table are the total number of observations n and the overall median. For each level of X, the table shows

1. *Sample Size*: the number of observations n_j .
2. $n<=$: of the observations at that level, how many are less than or equal to the overall median.
3. $n>$: of the observations at that level, how many are greater than the overall median.
4. *Median*: the level median.
5. *CL*: the lower and upper confidence limits for the median of the population from which the data came. If the sample sizes are too small, as in the current example, it may not be possible to obtain confidence limits.

Displayed at the bottom of the screen is a test statistic and P-Value. Treating the $n<=$ and the $n>$ columns as columns of a two-way contingency table, a chi-squared test statistic is calculated. Small P-Values (less than 0.05 if operating at the 5% significance level) lead to the conclusion that the medians are not all equal, as in the current example.

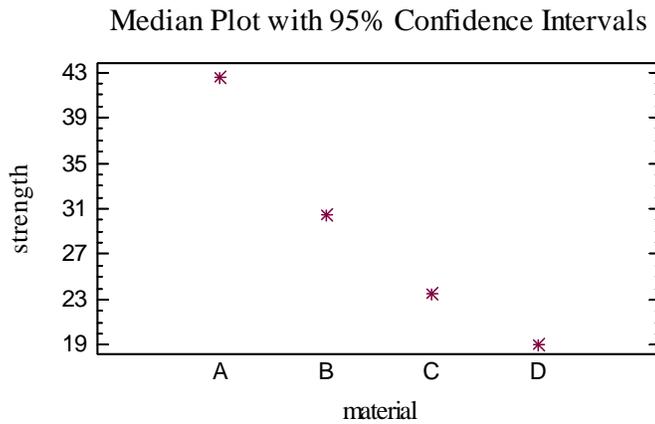
Pane Options



- **Confidence Level:** level used for the confidence limits.

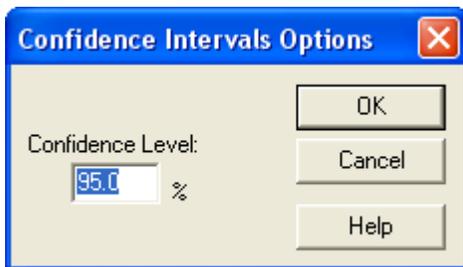
Medians Plot

The *Medians Plot* displays the confidence intervals for the medians displayed by the *Mood's Median Test* pane.



In the current case, the sample sizes are too small to permit the estimation of confidence intervals.

Pane Options



- **Confidence Level:** level used for the confidence limits.

Save Results

The following results can be saved to the datasheet:

1. *Level Counts* – the q sample sizes n_j .
2. *Level Means* – the q level means.
3. *Level Medians* – the q level medians.
4. *Level Standard Deviations* – the q level standard deviations.
5. *Level Standard Errors* – the standard errors of each level mean, $\sqrt{MS_{within} / n_j}$.
6. *Level Labels* – q labels, one for each level.
7. *Level Indicators* – n level indicators, identifying each residual.
8. *Residuals* – the n residuals.
9. *Level Ranges* – the q level ranges.

Calculations

Analysis of Variance

| Source | Sum of Squares | D.F. | Mean Square | F-Ratio |
|----------------|----------------------------------------------------------------------|----------------------------------------|----------------------------------------------------|----------------------------------------|
| Between groups | $SS_{between} = \sum_{j=1}^q n_j (\bar{Y}_j - \bar{Y})^2$ | $df_{between} = q - 1$ | $MS_{between} = \frac{SS_{between}}{df_{between}}$ | $F = \frac{MS_{between}}{MS_{within}}$ |
| Within groups | $SS_{within} = \sum_{j=1}^q \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)^2$ | $df_{within} = \sum_{j=1}^q (n_j - 1)$ | $MS_{within} = \frac{SS_{within}}{df_{within}}$ | |
| Total | $SS_{total} = \sum_{j=1}^q \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y})^2$ | n-1 | | |

Cochran's Test

The statistic displayed is calculated by

$$A = \frac{\max(s_j^2)}{\sum_{j=1}^q s_j^2} \tag{17}$$

To test for statistical significance,

$$C = (q - 1) \left(\frac{A}{1 - A} \right) \tag{18}$$

is compared to an F distribution with $(n/q - 1)$ and $(n/q - 1)(q - 1)$ degrees of freedom.

Bartlett's Test

The statistic displayed is calculated by

$$B = \frac{1}{C} \left[(dfe) \ln(MSE) - \sum_{j=1}^q (n_j - 1) \ln(s_j^2) \right] \tag{19}$$

where

$$C = 1 + \frac{1}{3(q-1)} \left[\left(\sum_{j=1}^q (n_j - 1)^{-1} \right) - \frac{1}{dfe} \right] \quad (20)$$

$$MSE = \frac{1}{dfe} \sum_{j=1}^q (n_j - 1) s_j^2 \quad (21)$$

$$dfe = \sum_{j=1}^q (n_j - 1) \quad (22)$$

B is compared to the chi-squared distribution with $(q-1)$ degrees of freedom.

Hartley's Test

$$H = \frac{\max(s_j^2)}{\min(s_j^2)} \quad (23)$$

Median Confidence Limits

The limits displayed are a nonlinear interpolation of the confidence intervals at the nearest confidence levels above and below the level requested. After sorting the observations, the interval extending from the d -th smallest observation in the sample to the d -th largest observation forms a confidence interval for the median with confidence level $1 - 2 P_B(d-1)$, where P_B represents the cumulative binomial distribution with $p = 0.5$ and $n = n_j$.