Multifactor ANOVA

Summary

The **Multifactor ANOVA** procedure is designed to construct a statistical model describing the impact of two or more categorical factors X_j on a dependent variable Y. Tests are run to determine whether or not there are significant differences between the means of Y at the different levels of the factors and whether or not there are interactions between the factors. In addition, the data may be displayed graphically in various ways, including a multiple scatterplot, a means plot, and an interaction plot.

This procedure is designed for relatively simple experiments, such as factorial experiments with fixed effects. The *General Linear Models* procedure should be used for more complicated situations.

Sample StatFolio: *anova.sgp*

Sample Data:

The file *stresstest.sgd* contains data from a stress test of *n* = 36 individuals, reported by Kutner et al. (1996). In the study, each subject exercised on a treadmill and the number of minutes required to reach a predefined level of stress was recorded. The table below shows a partial list of the data in that file:

Subject	Body fat	Gender	Smoking	Minutes
1	low	male	none	34
2	low	male	none	32
3	low	male	none	31
4	low	male	light	27
5	low	male	light	24
6	low	male	light	23
7	low	male	heavy	20
8	low	male	heavy	21
9	low	male	heavy	24
10	low	female	none	25
11	low	female	none	35
12	low	female	none	26
13	low	female	light	22
14	low	female	light	22
15	low	female	light	20
16	low	female	heavy	15
17	low	female	heavy	10
18	low	female	heavy	13
19	high	male	none	21
20	high	male	none	20

3 individuals were selected for each of 12 combinations of the following factors:
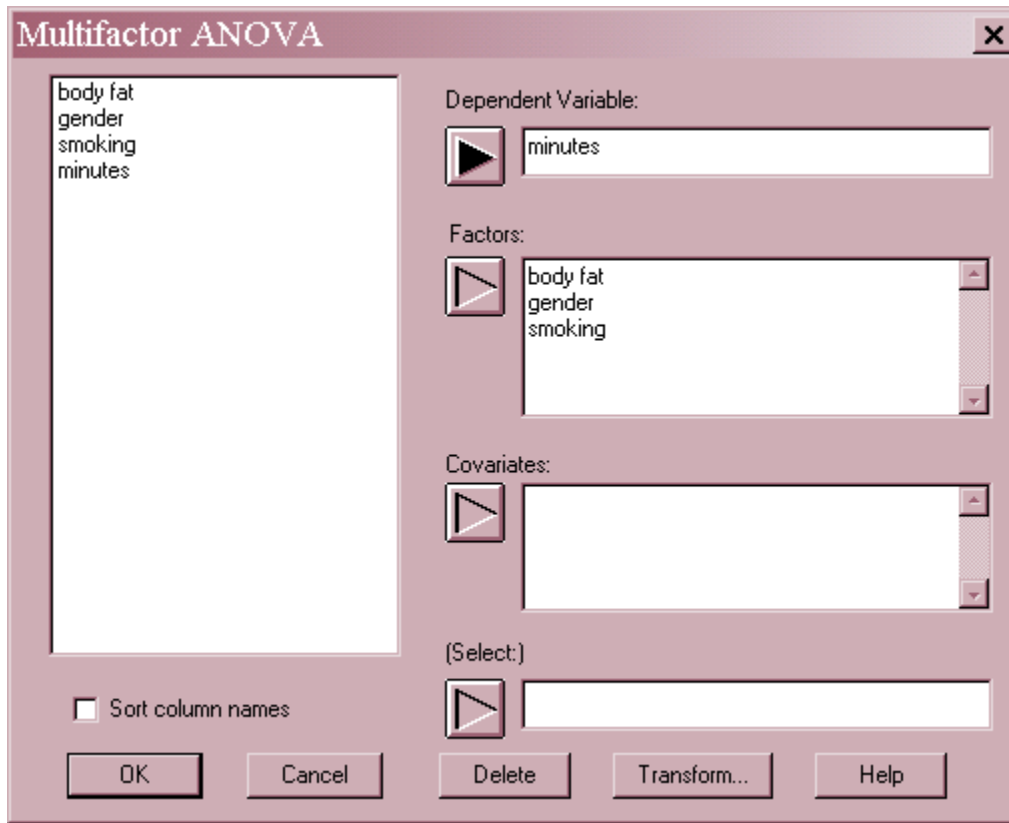
   Amount of body fat: low or high

Gender: male or female
Smoking history: none, light, or heavy

The study is a replicated 2 by 2 by 3 factorial design.

Data Input

The data consist of a single column containing the measurements and multiple columns indicating the levels of the experimental factors.

- **Dependent variable:** numeric column containing the observations.

- **Factors**: numeric or non-numeric columns containing levels identifying each factor.

- **Covariates:** optional numeric columns containing the values of quantitative variables that vary together with the response and whose effects should be adjusted for before comparing levels of the categorical factors.

- **Select**: subset selection.

Analysis Summary

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

Multifactor ANOVA - minutes
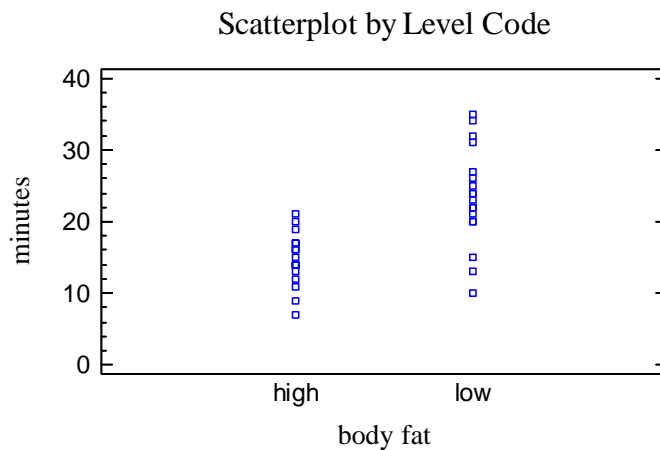Dependent variable: minutes
Factors:
     body fat
   gender
     smoking

Number of complete cases: 36

Scatterplot

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

Scatterplot by Level Code

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:
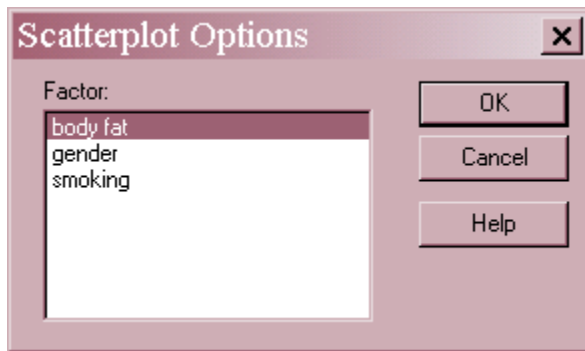

This offsets each point randomly in the horizontal direction so that identical values do not plot on top of each other:

Scatterplot by Level Code
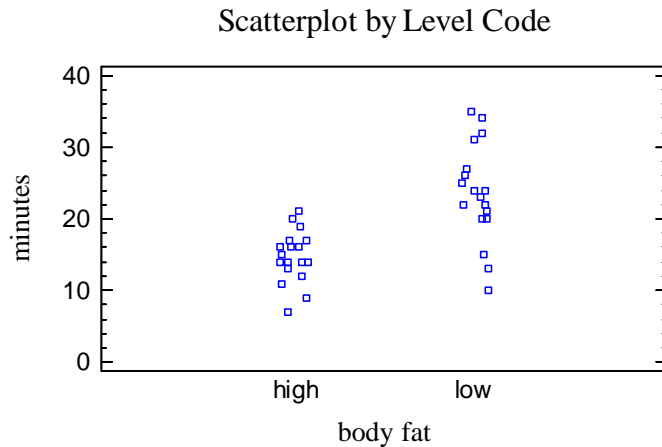
The above plot suggests that there are differences between individuals with high body fat and those with low body fat.

Pane Options

- **Factor**: factor to be plotted on the horizontal axis.

ANOVA Table

In order to determine whether or not the factors have a significant effect on the dependent variable, an analysis of variance is performed. The results are displayed in the *ANOVA Table*:

Analysis of Variance for minutes - Type III Sums of Squares					
Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
MAIN EFFECTS					
A:body fat	702.25	1	702.25	79.10	0.0000
B:gender	210.25	1	210.25	23.68	0.0000
C:smoking	343.056	2	171.528	19.32	0.0000
INTERACTIONS					
AB	2.25	1	2.25	0.25	0.6189
AC	204.167	2	102.083	11.50	0.0003
BC	21.5	2	10.75	1.21	0.3142
RESIDUAL	230.833	26	8.87821		
TOTAL (CORRECTED)	1714.31	35			
All F-ratios are based on the residual mean square error.					

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

1. A component attributable to the **Main Effect** of each factor, which measures the variability amongst the mean responses at each level of the factor.

2. A component attributable to the **Interaction** between different factors. An interaction occurs if the effect of one factor depends on the level of another factor.

3. If **Covariates** are present, a component attributable to each covariate.

4. A **Residual** component, which measures the variability amongst subjects at identical levels of the factors

Of particular importance are F-ratios and their associated P-Values. Small P-Values (less than 0.05 if operating at the 5% significance level) correspond to significant effects.

In the current example, all of the main effects are statistically significant as is the interaction between factors A and C (*Body fat* and *Smoking*).

Pane Options
The *Pane Options* dialog box controls how the F-tests are calculated:
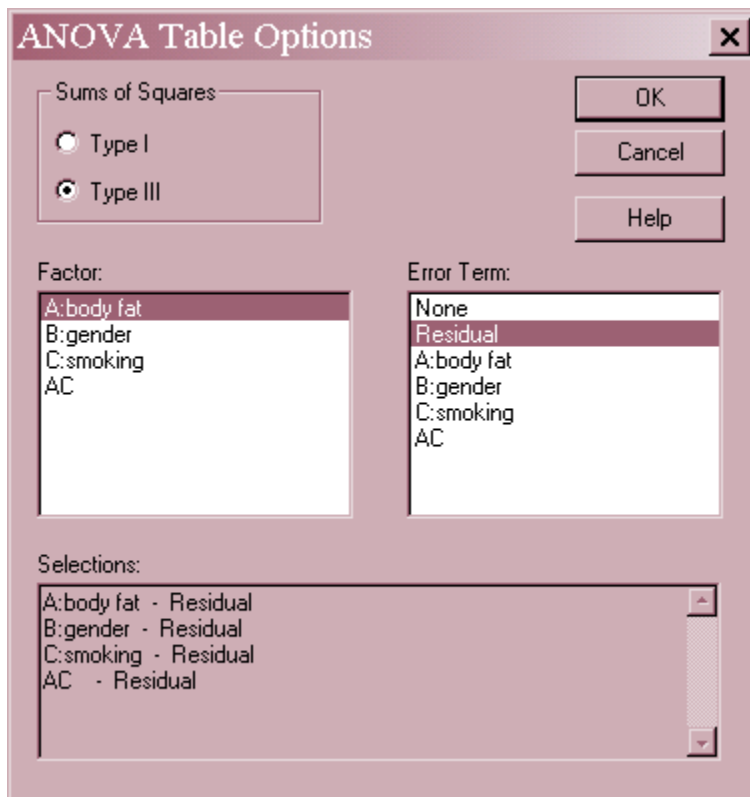
- **Sums of Squares:** the type of decomposition used to calculate the sums of squares in the ANOVA table. The default selection is Type III, which quantifies the increase in the error sum of squares that would occur if each effect were removed from the analysis, given that all of the other effects remain. In contrast, Type I sums of squares represent the reduction in the error sum of squares that occurs as each variable is added to the model, in the order shown in the ANOVA table. In a *balanced* experiment (an experiment with equal numbers of

observations at all combinations of the factors) such as the current example, both types of sums of squares yield identical results. In unbalanced cases, there will be a difference. Type III is the default since it quantifies the marginal contribution of each effect given that all of the other effects have been accounted for.

- **Error Term**: the mean square to be used as the denominator of the F-test when testing the significance of each effect. In a design in which all factors are crossed and non-random, the default selection of *Residual* is correct. For more complicated types of designs, the analyst may wish to specify another denominator for certain effects. Note: the *General Linear Models* procedure automatically determines the proper denominator for many types of models involving random and nested factors and should normally be used to analyze those types of experimental designs.

Analysis Options

The *Analysis Options* dialog box specifies the interactions to be included in the analysis.
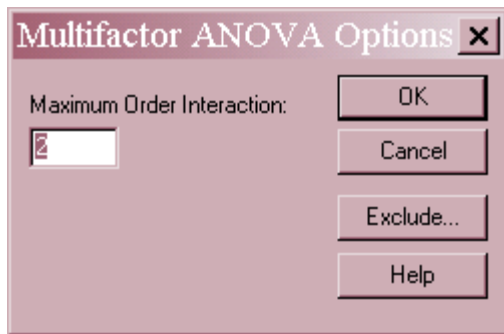
- **Maximum Order Interaction**: maximum number of factors for which an interaction will be estimated.

- **Exclude**: Press this button to remove one or more interactions from the analysis.

Example – Removing Insignificant Interactions
To remove one or more interactions from the analysis, press the *Exclude* button on the *Analysis Options* dialog box. This displays a second dialog box:
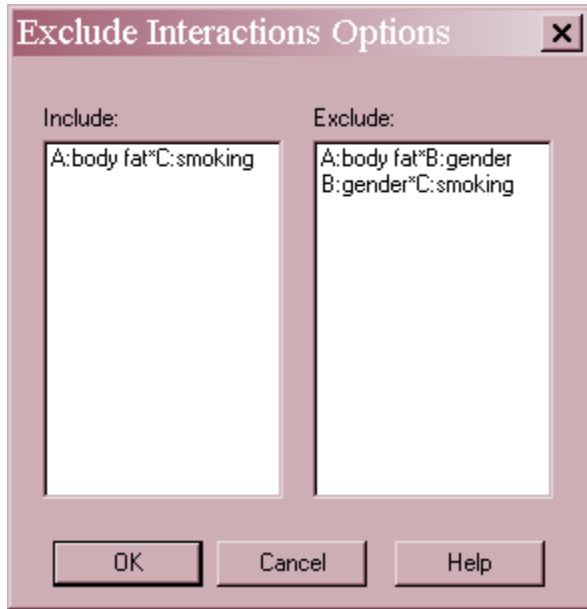
By double-clicking on any interaction, it can be moved from the left to the right or back again. Any interactions specified in the *Exclude* field will *not* be estimated.

After removing the two insignificant effects from the stress test data, the ANOVA table shows the remaining effects:

| Analysis of Variance for minutes - Type III Sums of Squares					
Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
MAIN EFFECTS					
A:body fat	702.25	1	702.25	79.99	0.0000
B:gender	210.25	1	210.25	23.95	0.0000
C:smoking	343.056	2	171.528	19.54	0.0000
INTERACTIONS					
AC	204.167	2	102.083	11.63	0.0002
RESIDUAL	254.583	29	8.77874		
TOTAL (CORRECTED)	1714.31	35			
All F-ratios are based on the residual mean square error.					

Graphical ANOVA

The *Graphical ANOVA* plot, developed by Hunter (2005), is a technique for displaying graphically the importance of each factor in the analysis. It is a plot of the scaled effects of each factor, where the "effect" of a factor equals the difference between the least squares mean for a level of that factor and the estimated grand mean. Each of the effects is multiplied by a scaling factor

$$\sqrt{\frac{v_R n_i}{v_T \bar{n}}} \tag{1}$$

where v_R is the residual degrees of freedom, v_T is the degrees of freedom for the main effect of 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:

Graphical ANOVA for minutes
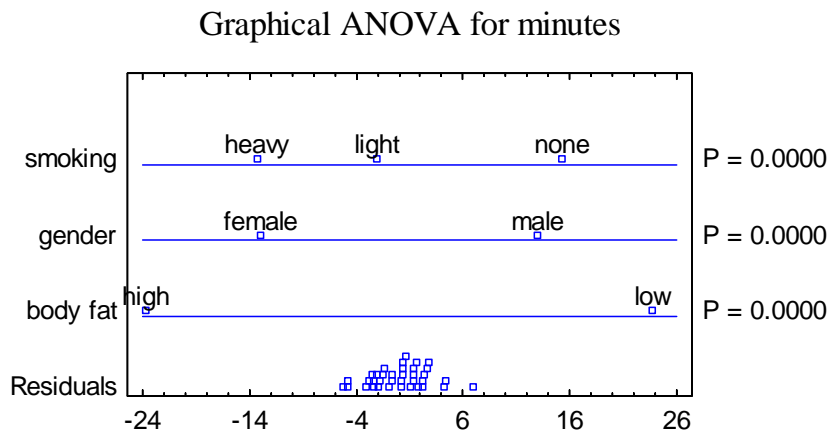
Along the right-hand side of the display are the P-values for the main effects, taken from the ANOVA table.

By comparing the variability amongst the treatment effects in the above plot to that of the residuals, it is easy to see that all of the factors show differences 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

For factor that shows significant P-Values in the ANOVA table and which do not interact with other factors, a further analysis can be performed by selecting the *Multiple Range Tests*.

Multiple Range Tests for minutes by gender

Method: 95.0 percent LSD

gender	Count	LS Mean	LS Sigma	Homogeneous Groups
female	18	16.7222	0.698361	X
male	18	21.5556	0.698361	X

Contrast	Sig.	Difference	+/- Limits
female - male	*	-4.83333	2.01994

* denotes a statistically significant difference.

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

- **Count** - the number of observations at the specified level of the factor.

- **LS Mean** - the estimated least squares mean. In the case of a balanced design, the least squares mean is equivalent to the average of all observations at the indicated factor level.

In unbalanced designs, the least squares mean is the predicted value of the dependent variable when the specified factor is set to a particular level while all other factors are set equal to their mean levels. The least squares means adjust for any imbalance in the data by making predictions at a common level of all the factors.

- **LS Sigma** – the estimated standard error of the least squares mean.

- **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. In the example, there are 2 columns, each containing a single X, indicating that the two genders fall into significantly different groups.

The second half of the table displays a comparison between each pair of level means.

- **Difference** - the difference between the two least squares means.

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

- **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
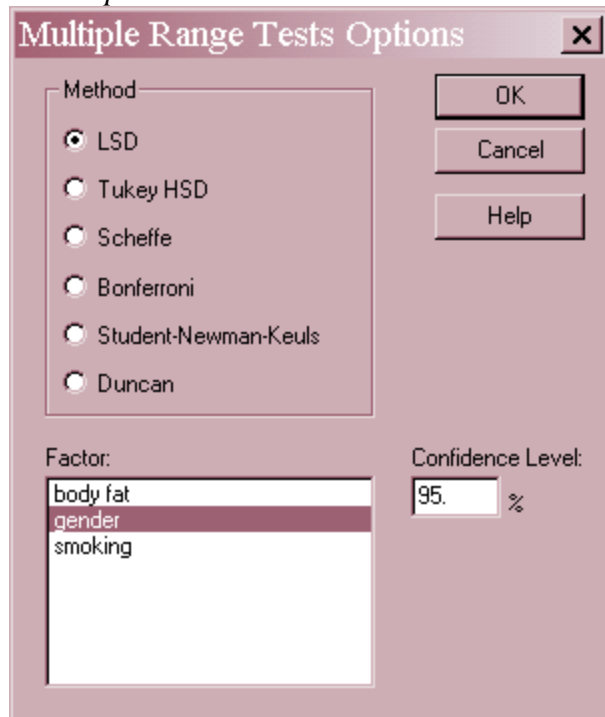
- **Method**: the method used to make the multiple comparisons.

- **Factor**: the factor to be displayed.

- **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 Student's t distribution. 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 allows for multiple comparisons amongst all pairs of means using Tukey's T. 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). 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. 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).
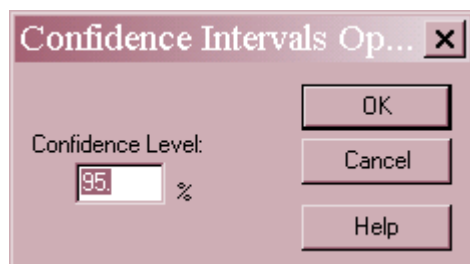
The choice between the LSD procedure and a multiple comparisons procedure such as Tukey's HSD should depend on the relative cost of making a Type I error (calling a pair of means different when they're really not) versus the cost of making a Type II error (not calling a pair of means different when they 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 the least squares mean for each level of the factors and for pairs of levels for any included two-factor interactions. Each mean is shown together its estimated standard error and a confidence interval:

Level	Count	Mean	Stnd. Error	Lower Limit	Upper Limit
GRAND MEAN	36	19.1389			
body fat					
high	18	14.7222	0.698361	13.2939	16.1505
low	18	23.5556	0.698361	22.1272	24.9839
gender					
female	18	16.7222	0.698361	15.2939	18.1505
male	18	21.5556	0.698361	20.1272	22.9839
smoking					
heavy	12	15.6667	0.855314	13.9174	17.416
light	12	18.5833	0.855314	16.834	20.3326
none	12	23.1667	0.855314	21.4174	24.916
body fat by smoking					
high,heavy	6	14.1667	1.2096	11.6928	16.6406
high,light	6	14.1667	1.2096	11.6928	16.6406
high,none	6	15.8333	1.2096	13.3594	18.3072
low,heavy	6	17.1667	1.2096	14.6928	19.6406
low,light	6	23.0	1.2096	20.5261	25.4739
low,none	6	30.5	1.2096	28.0261	32.9739

Caption (above table): **Table of Least Squares Means for minutes with 95.0 Percent Confidence Intervals**
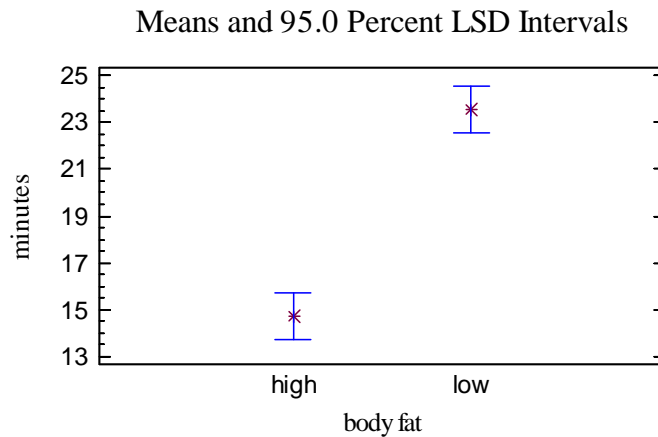
Pane Options

- **Confidence Level:** the level of confidence associated with each interval.

Means Plot

The level means may be plotted together with uncertainty intervals:

Means and 95.0 Percent LSD Intervals

The type of interval displayed depends on the settings in *Pane Options*.

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 high body fat does not overlap the interval for low body fat, indicating a statistically significant difference between the means at those two levels.

Pane Options
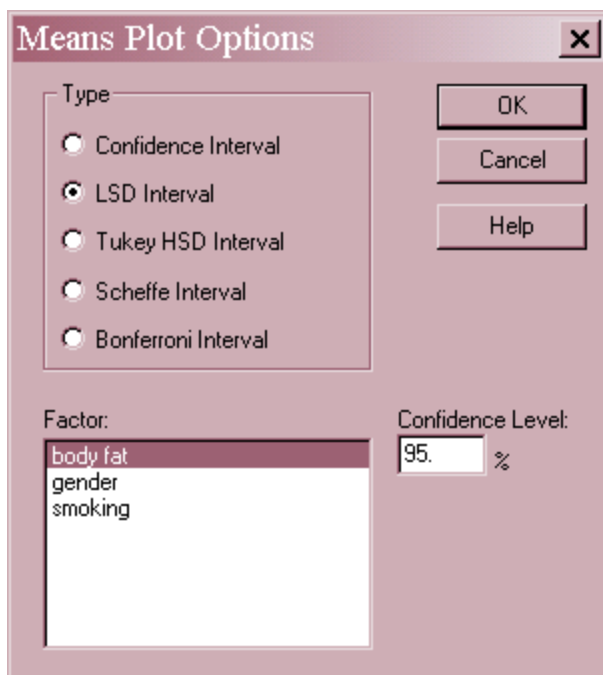
- **Intervals**: the method used to construct the intervals.

- **Factor**: the factor to be plotted.

- **Confidence Level:** the level of confidence associated with each interval.

The type of intervals that may be selected are:

- **Confidence intervals** - displays confidence intervals for the level means using the mean squared error from the ANOVA table.

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

- **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.

Interaction Plot

When one or more significant interactions exist, they should be examined together using the *Interaction Plot*.
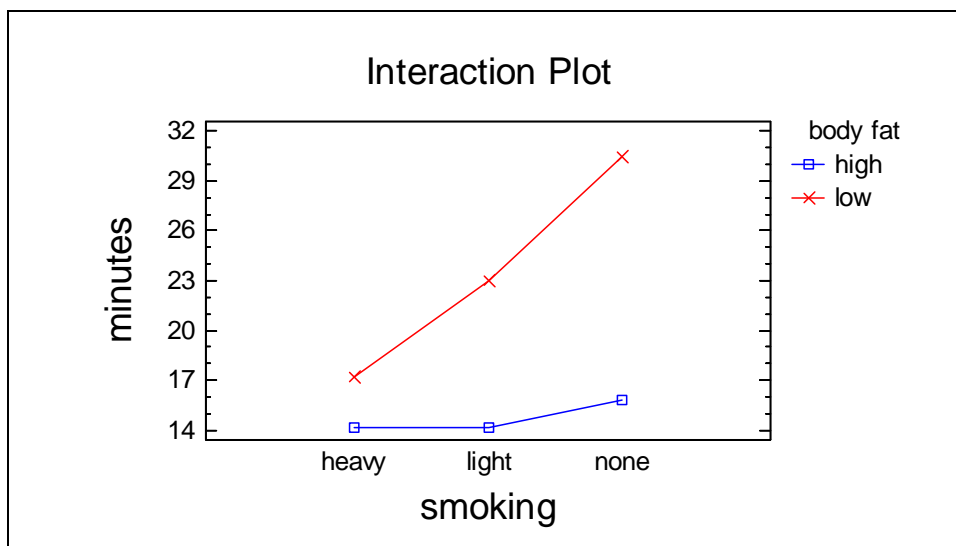
The interaction plot displays the least squares means at all combinations of two factors. If the factors do not interact, the lines on the plot should be approximately parallel. If they are not, then the effect of one factor depends upon the level of the other, which is the definition of an interaction.

Notice that the effect of smoking is much greater on individuals with low body fat than on those with high body fat.

Pane Options


- **Interval** - the type of interval (if any) to be placed around each mean.

- **Confidence Level** - the confidence level for the interval.

- **Interaction** - the interaction to be plotted. A point will be displayed showing the predicted mean value for each combination of the factors in the selected interaction.

- **Plot on Axis** - the factor within the selected interaction that will be used to define the horizontal axis. Separate lines will be drawn for each level of the other factor.

Example – Interaction Plot with Tukey Intervals
Adding Tukey HSD intervals permits a pairwise comparison amongst the 6 combinations of *Smoking* and *Body Fat*:

Interactions and 95.0 Percent Tukey HSD Intervals
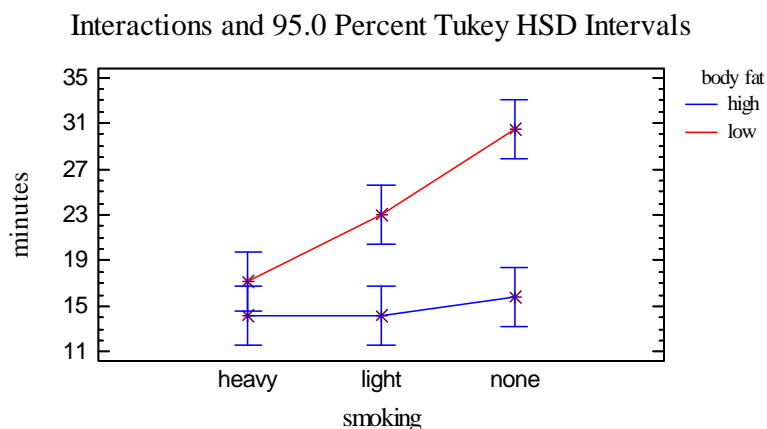
Examining the overlap of the intervals, three homogenous groups are identifiable:

Group 1: non-smoking, low body fat individuals, whose time in the test is significantly longer than all other individuals.

Group 2: light-smoking, low body fat individuals, whose time in the test is less than Group 1 but significantly longer than everyone else.

Group 3: everyone else. Note that their intervals all overlap, indicating no statistically significant differences amongst any of the remaining individuals.
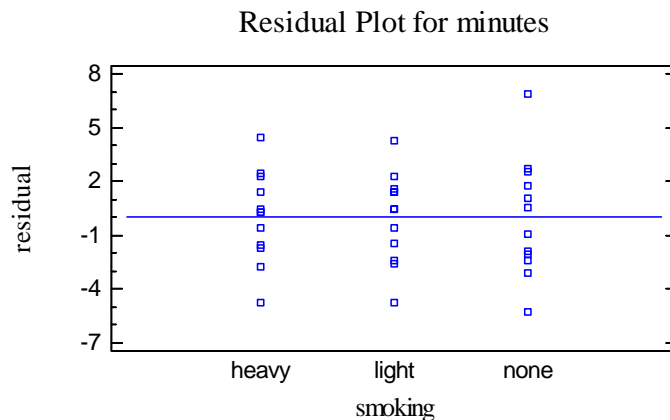
Residual Plots

As with all statistical models, it is good practice to examine the residuals. The residuals are equal to the observed data values minus the values predicted by the underlying statistical model.

The *Multifactor ANOVA* procedure creates 3 residual plots:

1. versus factor level.
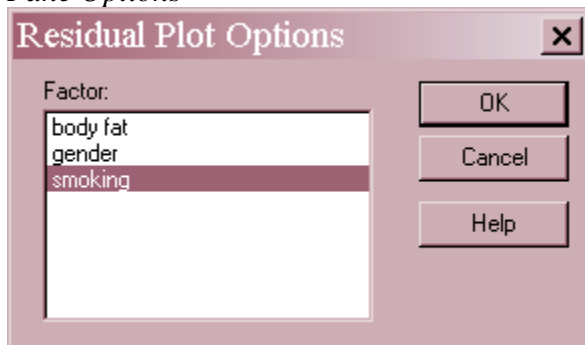2. versus predicted value.
3. versus row number.

Residuals versus Factor Level
This plot is helpful in visualizing any differences in variability at various levels of a factor.

Residual Plot for minutes
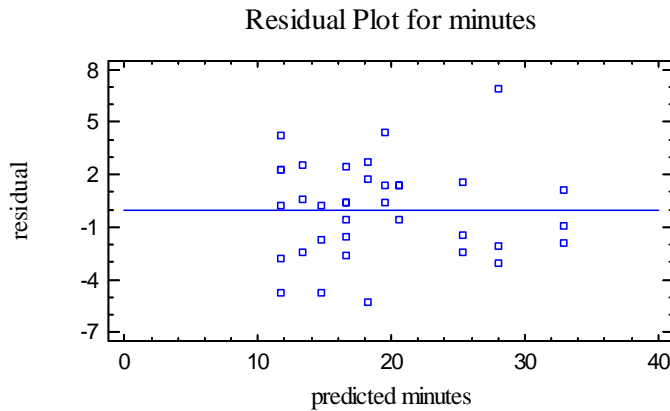
The average residual at each level equals 0.

Pane Options

- **Factor**: factor to display on the horizontal axis.

Residuals versus Predicted
This plot is helpful in detecting any heteroscedasticity in the data.

Residual Plot for minutes

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:

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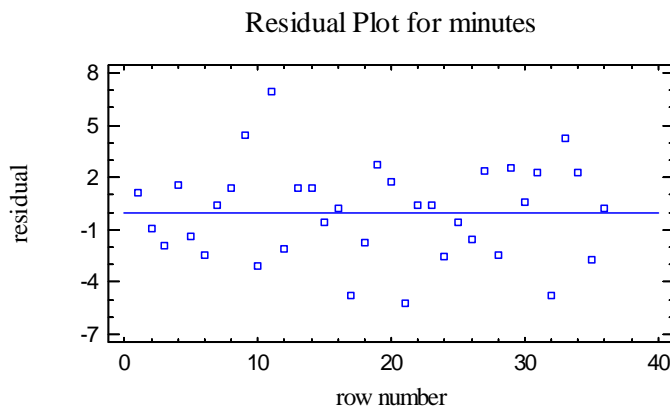
If the data are arranged in chronological order, any pattern in the data might indicate an outside influence. No such pattern is evident in the above plot.

Save Results

The following results can be saved to the datasheet:

1. *Level Counts* – the number of observations at each level of the factors and each pair of factors.
2. *Level Means* – the mean response at each level of the factors and each pair of factors.
3. *Level Standard Errors* – the standard error at each level of the factors.
4. *Least Squares Mean* – the least square mean at each level of the factors.
5. *Residuals* – the *n* residuals.

<u>Calculations</u>

Statistical Model

In order to fit a model to the data, STATGRAPHICS constructs an *n* by *p* matrix of independent variables *X*. The matrix includes:

- A column of 1's to represent a constant.

- Indicator variables for each factor. For a factor with *k* levels, *k* – 1 indicator variables are constructed. The j-th indicator variable for a factor contains the value 1 for each observation equal to the j-th level of the factor, –1 for each observation equal to the k-th level of the factor, and 0 otherwise.

- A column containing the values of each covariate, if any.

- Cross-products of the indicator variables and covariate columns to represent any interactions.

A model is then fit using linear least squares estimation:

$$\hat{\beta} = (X'X)^{-1} X'Y \tag{2}$$

Least Squares Means

The least squares means are the predicted response

$$\hat{Y}_p = X'_p (X'X)^{-1} X'Y \tag{3}$$

where X_p is the vector of independent variables in which each indicator variable corresponding to factors not included in the specified effect is set to 0 and each covariate is set to its observed mean level.

Standard Errors
Estimated from

$$s.e._p = \sqrt{X'_p MSE(X'X)^{-1} X_p} \tag{4}$$

where *MSE* equals the mean squared error in the ANOVA table.

Multiple Comparison Procedures
See the *Oneway ANOVA* documentation.