Variance Components Analysis

Summary

The **Variance Components Analysis** procedure is designed to estimate the contribution of multiple factors to the variability of a dependent variable Y. It is designed to analyze a nested experiment, in which the factors are structured in a hierarchical manner. In such a study, samples of each factor are taken from within samples of the factor immediately above it.

For example, *b* batches might be taken from a process. Then *s* samples might be taken from each batch. Finally, *t* tests might be run on each sample. The final data set would have a total of *n* = *bst* measurements.

This procedure is designed for an experiment in which factors are structured in a strictly hierarchical manner, and in which all effects are assumed to be random. The *General Linear Models* procedure should be used for more complicated situations.

Sample StatFolio: *varcomp.sgp*

Sample Data:

The file *pigment.sgd* contains data from an experiment described by Box, Hunter and Hunter (1978). In that experiment, *b* = 15 batches of pigment paste were selected. From each batch, *s* = 2 samples were taken, and *t* = 2 tests were run on each sample to measure the moisture content. A total of *n* = 60 measurements are contained in the file, a portion of which is shown below:

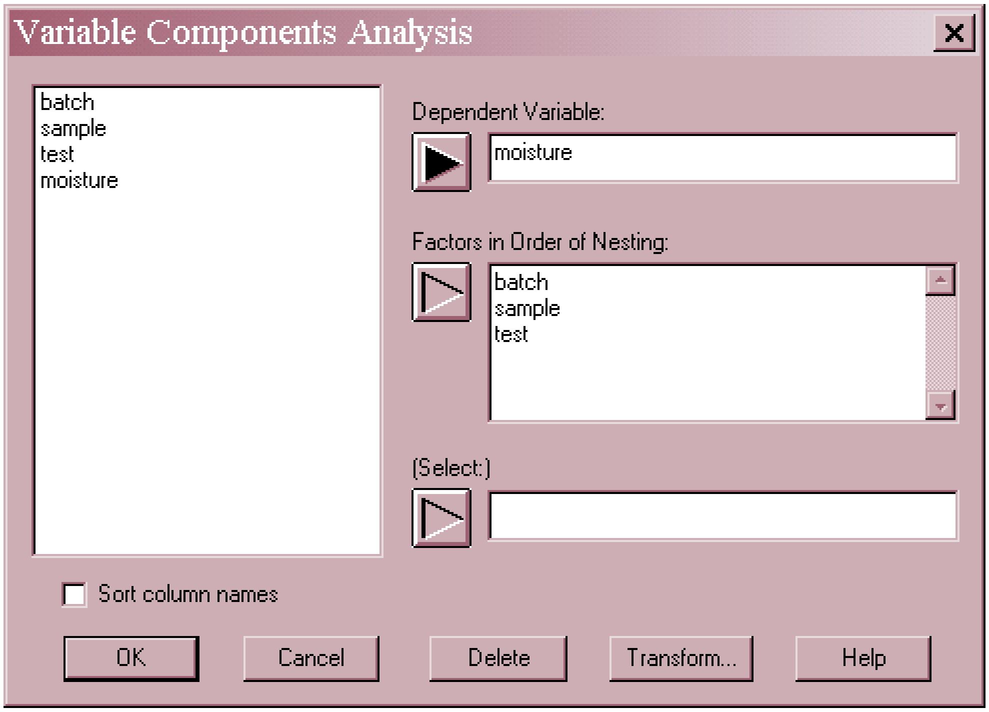
Batch	*Sample*	*Test*	*Moisture*
1	1	1	40
1	1	2	39
1	2	1	30
1	2	2	30
2	3	1	26
2	3	2	28
2	4	1	25
2	4	2	26
3	5	1	29
3	5	2	28
3	6	1	14
3	6	2	15

The batches are numbered from 1 to *b* = 15. The samples are numbered from 1 to *bs* = 30, although they could have been labeled from 1 to *s* = 2 within each batch. The tests are numbered from 1 to *t* = 2 within each sample, although they could have been numbered from 1 to *bst* = 60. Each numbering scheme will give identical results.

© 2009 by StatPoint Technologies, Inc.

Data Input

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



Dependent Variable: moisture

Factors in Order of Nesting: batch, sample, test

(Select:)

□ Sort column names

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

- **Factors in Order of Nesting**: numeric or non-numeric columns containing levels identifying each factor. The factors must be entered from the top down, i.e., each factor is assumed to be nested in the factor immediately above it in the list. This is one of the few STATGRAPHICS procedures in which the order of the factors affects the analysis.

- **Select**: subset selection.

Note: the final factor *test* can be omitted from the list of factors in the dialog box. If so, its effect will be included as a "Residual" term in the ANOVA table.

Statistical Model

The relevant statistical model for the sample data is

$$Y_{bst} = \mu + \varepsilon_b + \varepsilon_s + \varepsilon_t \qquad (1)$$

where

μ = process mean

ε_b = deviation of the mean of batch *b* from the process mean μ

ε_s = deviation of the mean of sample *s* from the mean of batch *b*

ε_t = deviation of the test *t* measurement from the mean of sample *s*

The deviations are usually assumed to be random samples from normal distributions with standard deviations:

σ_b = standard deviation amongst the batches

σ_s = standard deviation amongst the samples within the batches

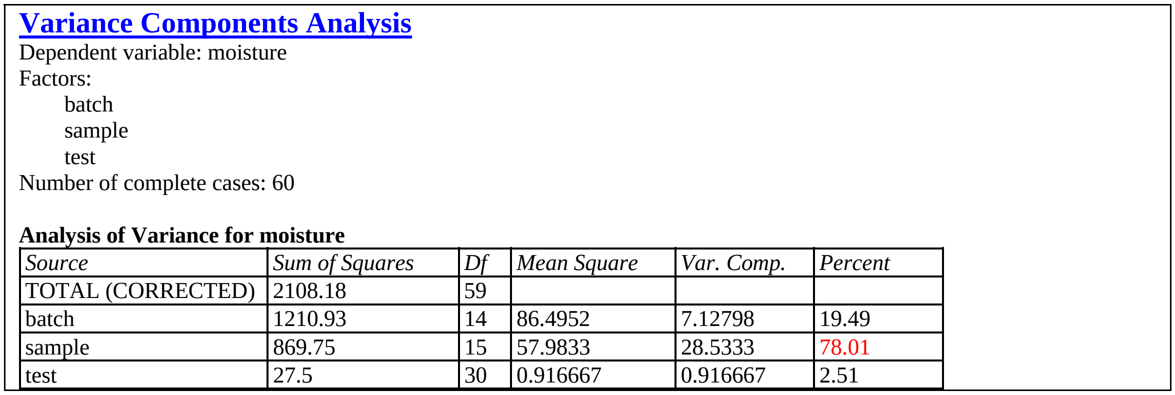
σ_t = standard deviation amongst the test results within each sample

Assuming that the various error components are independent, the overall process variability is then the sum of the variability due to the different components, i.e.,

$$\sigma^2 = \sigma_b^2 + \sigma_s^2 + \sigma_t^2 \qquad (2)$$

Analysis Summary

The *Analysis Summary* shows the number of observations *n* and an analysis of variance table.



Variance Components Analysis

Dependent variable: moisture
Factors:
    batch
    sample
    test
Number of complete cases: 60

Analysis of Variance for moisture

Source	*Sum of Squares*	*Df*	*Mean Square*	*Var. Comp.*	*Percent*
TOTAL (CORRECTED)	2108.18	59			
batch	1210.93	14	86.4952	7.12798	19.49
sample	869.75	15	57.9833	28.5333	78.01
test	27.5	30	0.916667	0.916667	2.51

The table shows:

- **Sums of Squares**: a decomposition of the sum of squared deviations around the grand mean.

- **Df**: the degrees of freedom associated with each sum of squares.

- **Mean Square:** the sums of squares divided by their degrees of freedom.

- **Var. Comp.:** the estimated variance components, which are the estimated variances of each factor within the factor it is nested in. The variance components are estimated by setting the mean squares in the ANOVA table equal to their expected values and solving the resultant equations.

- **Percent:** the percentage of the total process variance represented by each component.

In the sample data, the variance component estimates are:

$$\hat{\sigma}_b^2 = 7.128 \quad \hat{\sigma}_s^2 = 28.53 \quad \hat{\sigma}_t^2 = 0.9167.$$

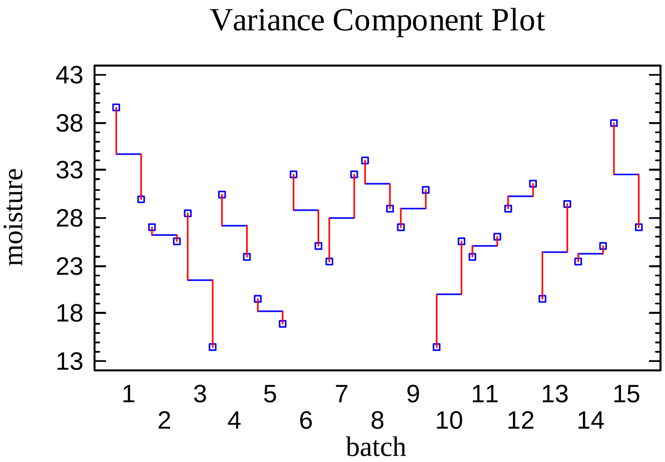
The estimate of the total process variability is

$$\hat{\sigma}^2 = \hat{\sigma}_b^2 + \hat{\sigma}_s^2 + \hat{\sigma}_t^2 = 36.63$$

Note that the variability among samples within the same batch represents over 78% of the total variability, indicating a problem with the homogeneity within the batches.

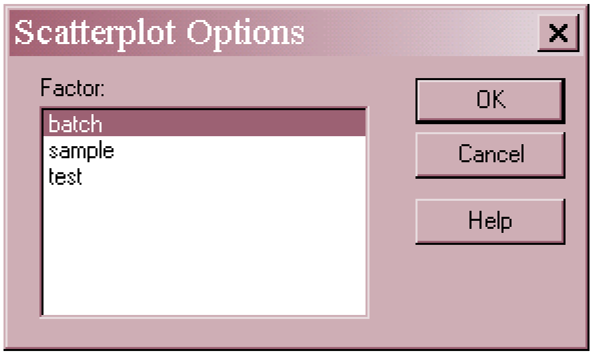
Scatterplot

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



The above plot shows horizontal lines at each of the 15 batch means. Each point represents the mean of one sample within a batch.

Pane Options

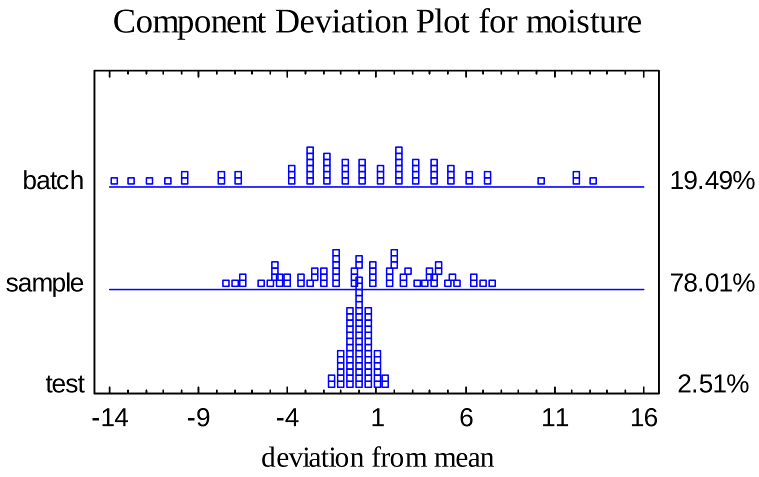


- **Factor**: factor to be plotted on the horizontal axis.

© 2009 by StatPoint Technologies, Inc.

Component Deviation Plot

The *Component Deviation Plot* displays the deviation of each observation from the mean of all observations at the same level of a selected factor:



Each section of the plot contains a point corresponding to each observation. In each section, a different mean has been subtracted from the data value.

Topmost section (batch): shows the deviation of each observation from the overall mean of all the observations.

Center section (sample): shows the deviation of each observation from the mean of the batch from which it was taken.

Lower section (test): shows the deviation of each observation from the mean of the sample from which it was taken.

The effect is to show, from bottom up, the additional contribution of each component. Variability in the lower section is due solely to the testing process. Variability in the center section includes both testing variability and variability between samples within the same batch. Variability in the topmost section comes from all three components.

In the above plot, it is clear that a substantial amount of variability is introduced at the level of samples within batches.

Summary Statistics

The Su*mmary Statistics* table shows the sample sizes, means, and standard deviations at each level of the factors. A portion of the table is shown below:

Summary Statistics for moisture

			Standard
Level	*Count*	*Mean*	*Deviation*
GRAND MEAN	60	26.7833	5.97762
batch			
1	4	34.75	5.5
2	4	26.25	1.25831
3	4	21.5	8.1035
4	4	27.25	3.77492
5	4	18.25	1.5
6	4	28.75	4.42531
7	4	28.0	5.22813
8	4	31.5	2.88675
9	4	29.0	2.3094
10	4	20.0	6.58281
11	4	25.0	1.63299
12	4	30.25	1.5
13	4	24.5	5.8023
14	4	24.25	0.957427
15	4	32.5	6.45497
sample			
1	2	39.5	0.707107
2	2	30.0	0.0
3	2	27.0	1.41421
4	2	25.5	0.707107
5	2	28.5	0.707107
6	2	14.5	0.707107
7	2	30.5	0.707107
8	2	24.0	0.0

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.

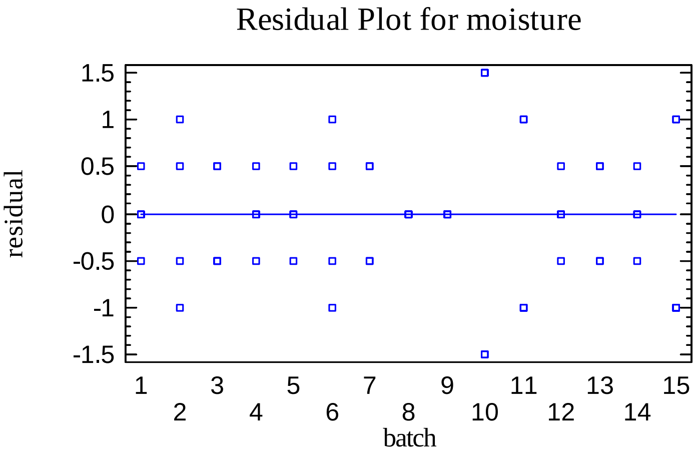
The *Variance Components* procedure creates 3 residual plots:

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

Note: In the sample data, factors have been specified for each level of experimental error, so that the residuals are all equal to 0. If *test* is removed as a factor, then its effect will be reflected in a residual term. The plots below reflect such as analysis.

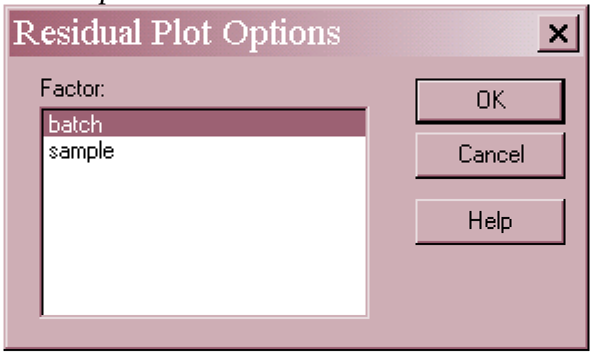
Residuals versus Factor Level

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



The average residual at each level equals 0.

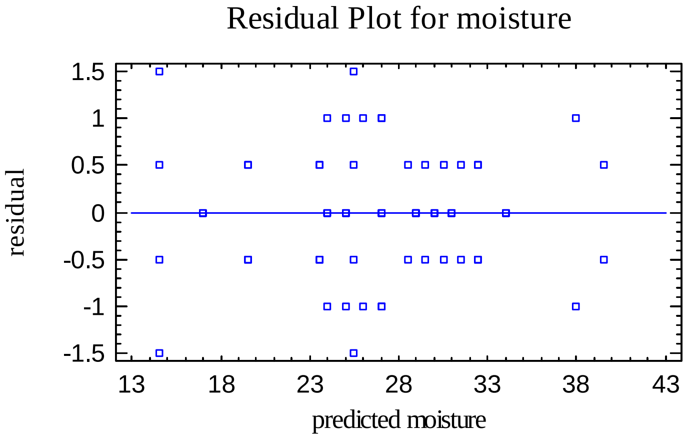
Pane Options



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

Residuals versus Predicted

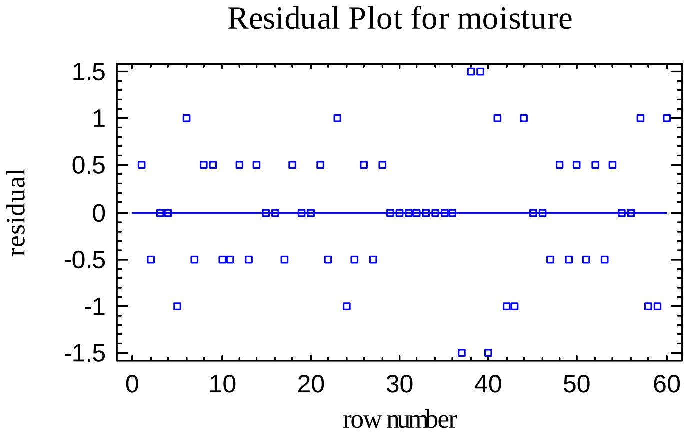
This plot is helpful in detecting any heteroscedasticity in the data.



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.

Save Results

The following results can be saved to the datasheet:

1. *Variance Components* – the estimated variance components.
2. *Residuals* – the *n* residuals.

Calculations

The estimation of variance components follows the procedure described in the *General Linear Models* documentation.

© 2009 by StatPoint Technologies, Inc.