ANOVA 2 factors

Menu:	QCExpert	Anova-2 factors

Anova for 2 factors is an extension of the 1-factor ANOVA described above. In 2-factor ANOVA we analyze influence of two factors on a numeric response. Observed response Z_{ij} at n_X different levels of the factor X and at n_Y different levels of the factor Y may be described by Anova model for 2 factors X, Y :

$$Z_{ij} = Z_0 + \alpha_i X_i + \beta_j Y_j + \lambda_{ij} + \varepsilon_{ij}, \quad i = 1,..,n_X,\ j = 1,..,n_Y,$$

where Z_0 is absolute term (overall mean), α a β are contributions of the individual levels, elements λ_{ij} of a matrix Λ are called interactions and ε_{ij} is the random error with normal distribution and zero mean, $\varepsilon \sim N(0,\sigma^2)$. Further we define $\Sigma\alpha_i = 0$, $\Sigma\beta_j = 0$, $\Sigma\lambda_{i(j)} = 0$, $\Sigma\lambda_{j(i)} = 0$.

Data and parameters

The module expects data in 3 columns. Two columns contain, levels of the two factors, in one column there are the corresponding observed response values. The combination of factors may be in arbitrary order. Factor levels are entered in form of any text string, such as RED, BLUE, GREEN, different string means different factor level. Both factors may have two or more levels. Each possible combination must be represented by al least one row.

The following table gives an example of data. First factor is the plant species with 3 levels: *Brazil*, *Longleaf*, *Cassablanca*, second factor is the fertilizer used, with 2 levels: *Nitrate* and *Phosphate*. The response is the observed increment in plant weight (*Yield*). There is 1 observation for each combination of levels, this experimental plan is called plan without replications. If we had the same number $N > 1$ of observations for each combination, we would have a balanced plan with N replications. If number of replications N_{ij} is not the same for all combinations, we have an unbalanced experimental plan. Each different combination of factor levels is called *a cell*. We can distinguish 3 types of 2-factor Anova:

Species	Fertilizer	Yield
Brazil	Phosphate	14.6
Brazil	Nitrate	17.4
Longleaf	Phosphate	13.3
Longleaf	Nitrate	12.6
Casablanca	Phosphate	17.5
Casablanca	Nitrate	14.9

1 observations in each cell – *Balanced ANOVA without replications*
Equal number $n_0 > 1$ of observations in each cell – *Balanced ANOVA with n_0 replications*
Unequal number $n_{ij} > 0$ of observations in each cell – *Unbalanced ANOVA*

In the dialog window (Fig. 1), select columns with both factors and the response. Clicking OK will run the analysis and results will be written in Protocol and Graphs windows.

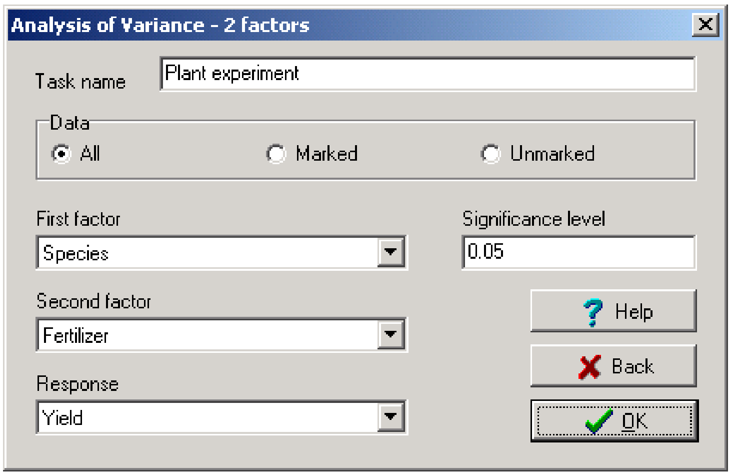
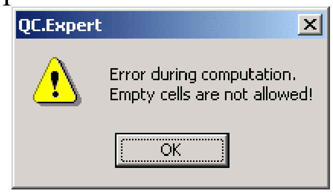


Fig. 1 ANOVA 2 factors – Dialog window

From the data the module automatically recognizes which of the three Anova types should be used. If there is one or more cells (or combinations of factor levels) which has no observations, the message „Empty cells are not allowed“ appears.



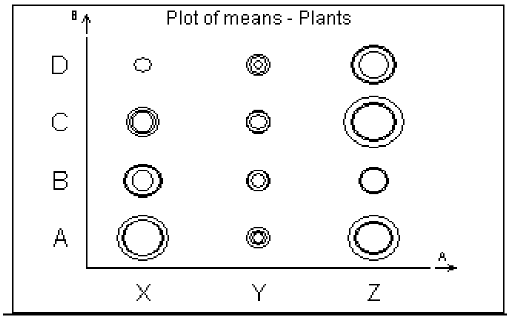
Protocol

Analysis of variance	Name of the module
Task name :	Task name from the dialog window
Type of model	Automatically recognized plan: - balanced without replications, or - balanced with replications, or - unbalanced
Factors levels	Names of the factors and their levels
No of replications	In case of balanced ANOVA: number of replications in cell n_0 In case of unbalanced ANOVA: a table of numbers of replications in each cell, n_{ij}
Table of means	Table of arithmetical averages in each cell
Means for factor	Total averages for each level of one factor regardless the levels of the other
Overall mean	Total average of all responses. (This would be the estimate of mean response if no factor had any significant influence.)
Model parameters	Estimates of parameters α and β, the contributions of each level.
ANOVA Table	Summarized analysis of variance structure

Source of variability	Identification of the source (or cause) of variability
First factor	Variability caused by the first factor
Second factor	Variability caused by the second factor
Interaction	Variability caused by the interaction of factors
Residuals	Residual variability
Total	Total variability
Sum of squares	Sum of squares caused (or explained) by the source
Mean square	Mean square caused (or explained) by the source
Degrees of freedom	Degrees of freedom of the source
Std deviation	Std deviation caused (or explained) by the source
F-statistic	Computed F-statistic value for the given source
Critical quantile	Critical quantile of the F-distribution, If Critical quantile is less than F-statistic, then the parameter is a statistically significant source o variability.
Conclusion	Verbal conclusion of the significance test
p-value	The *p*-value for each test

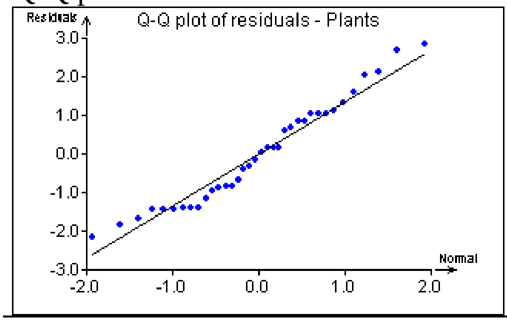
Graphs

Plot of means



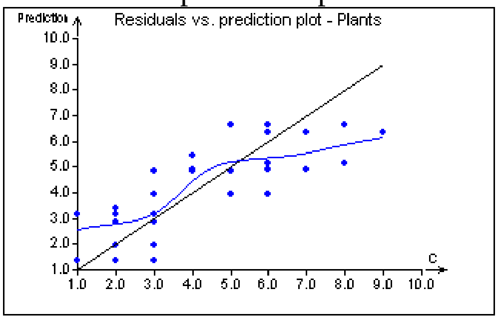
This plot visually compares the response values for each combination of factor levels (or each cell). Each circle stands for one observation. Diameter of the circles are roughly proportional to the response. Differences between cells, between levels and interactions can be seen on this plot.

Q-Q plot of residuals



Q-Q plot of residuals is to explore normality of distribution of the residuals. If the points lie roughly on the line, then the distribution is close to normal. Recall that normality of residuals is one of important assumptions of the Anova model.

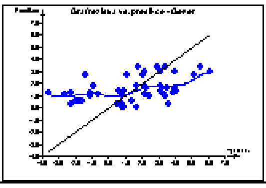
Residuals vs. prediction plot



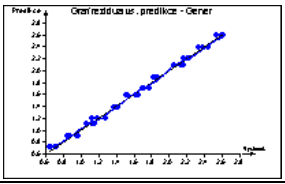
Residuals vs. prediction plot shows the quality of fit of the model. If the model is not significant (cannot explain much of the variability of response), the points lie on a horizontal line (illustration A).
The closer are the points to the line $y = x$, the more significant the model is.

Illustration:

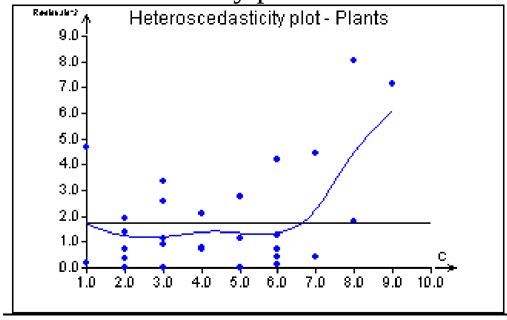
A



B

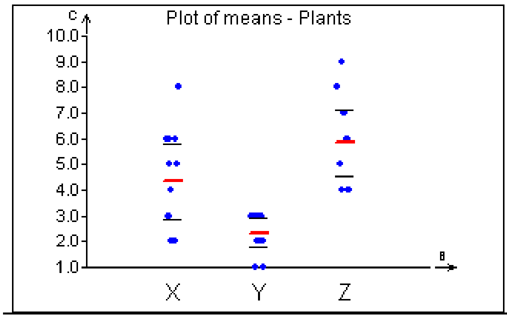


Heteroscedasticity plot



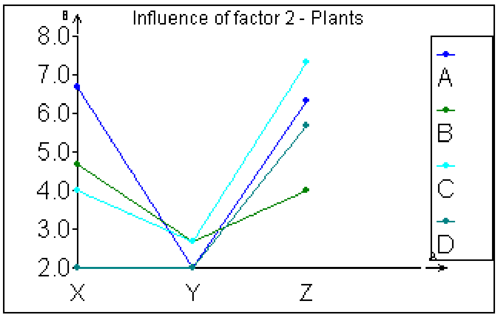
Sometimes it shows, that the variance of the response is higher for bigger values of the response. Heteroscedasticity plot shows the dependence of variance on the response value. As this Anova models assumes constant variance of the response, heteroscedasticity may affect its performance. The blue curve on the plot should not decline from horizontal line.

Plot of means



This plot shows the differences between means for individual levels of the first and second factor. Dots are the measured responses, short thick lines are the means, short black lines are the confidence intervals of the means. If the factor is statistically significant, the mean lines are red.

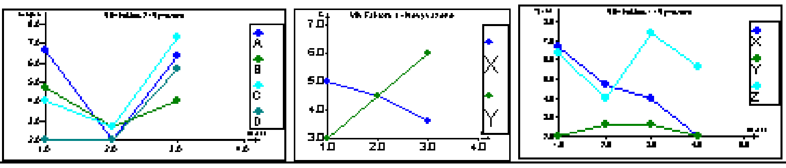
Influence of factor



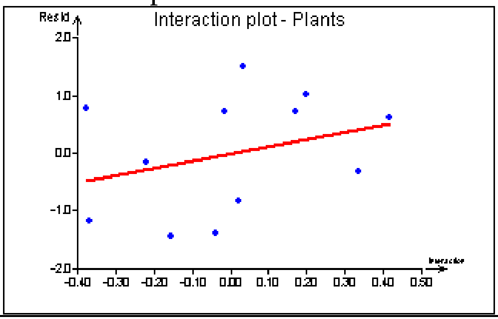
Plot of the influence of a given factor at separate levels of the other factor. If the lines for factor *A* are similar for each level of the factor *B*, then the factor A is probably significant (illustration A). If the lines have rather opposite direction, then there is probably strong interaction between the factors (illustration B). If the lines are shaped rather randomly, then the influence and significance of the respective factor is probably low, illustration C. The plot is only qualitative visual tool an cannot fully replace the F-test.

Illustration:

A **B** **C**



Interaction plot



Interaction plot visualizes the significance of the interaction term in the Anova model. If there is a significant slope in the plot, then the interaction between factors is significant. The statistical significance of the interaction is shown by red color of the line.