Transformation

Menu:	QC.Expert	Transformation

This module is helpful when dealing with skewed (asymmetrically distributed) data, for which a departure from normality was detected, using *the Basic data analysis* module. The *Transformation* module uses two strategies to find the best transformation. Mean, confidence interval for the mean and the table of selected quantiles are produced for the transformed data. Computed characteristics are backtransformed to the original scale. The *Exponential* transformation is based on skewness minimization, while the *Box-Cox* transformation is based on the maximum likelihood method. The transformation approach takes into account asymmetry of the data distribution and often yields better mean and quantile estimates. It is intended for systematically skewed data, but not for situations where skewness is caused by several outliers. The transformation technique is used also for asymmetric control chart limits construction.

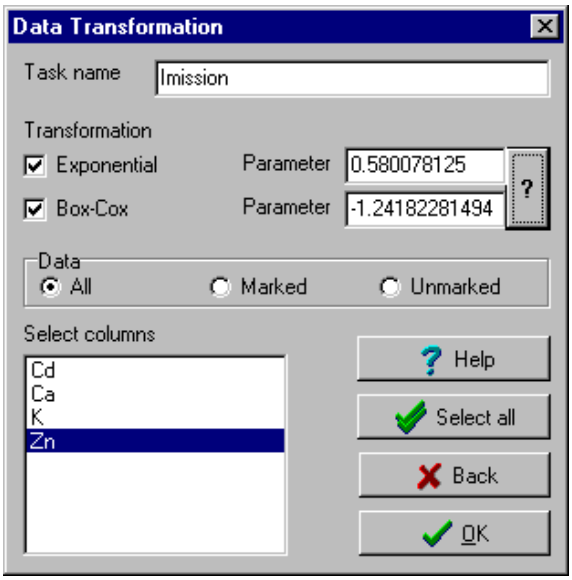


Fig. 1 Transformation dialog panel

Data and parameters

Data are in columns. Columns can be selected in the *Columns* field of the *Transformation* dialog panel. When the transformation parameter *r* is known, it can be entered in the *Parameter* window. The optimal *r* value is computed upon pressing the "?" button. Computations are started by the OK button. When several columns are selected, computations are done as if all data came from a single column.

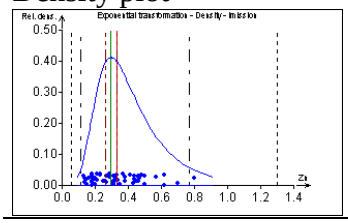
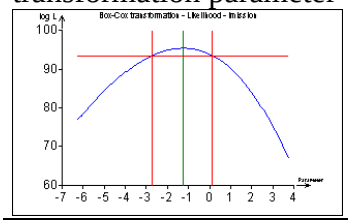
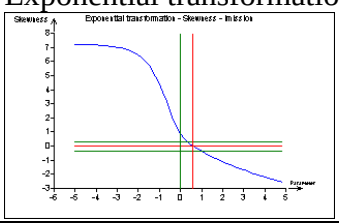
Protocol

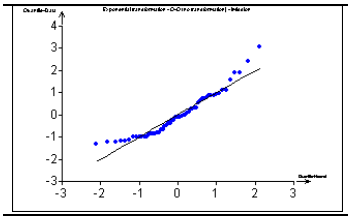
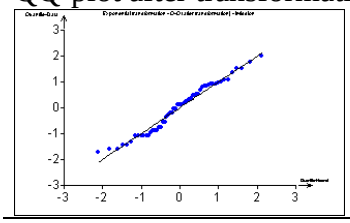
Optimized parameter	The best transformation parameter *r* value, found by likelihood maximization.
Lower and upper confidence limit	Confidence interval for the transformation parameter *r*. The interval becomes narrower as the sample size increases. When the interval contains 1, it is not advisable to transform the data. This situation occurs either when the data are already normal or when the sample size is not large enough to rule out the possibility that they are. If the interval contains 0, the data might be considered lognormal.
Likelihood for r=1	Logarithm of the likelihood function evaluated at r=1 (corresponds to no transformation).
Likelihood maximum	Maximum of the log likelihood function , i.e. the log likelihood evaluated at the optimal *r* value.

Conclusion	Recommendation about transformation. NO means that the transformation does not lead to a substantial improvement. YES means that the transformation is recommended. It is recommended when the *r* parameter is significantly different from 1 at 95% significance level.
Significance	Result of the *r*=1 hypothesis test. When the result is significant at 95% level, transformation is recommended, see the previous item.
Inputted parameter	Value of the transformation parameter, entered by the user in the Transformation dialog panel. Other than optimal value can be entered.
Likelihood	Log likelihood evaluated at the selected parameter value.
Corrected mean	Arithmetic average, computed after the optimal transformation was selected by the Box-Cox transformation approach. It might be a better estimate of the mean than the simple arithmetic average, computed in the Basic data analysis module when the data come from a highly skewed distribution.
LCL	Suggested value for the Shewhart X-bar control chart lower limit. The number of data columns corresponds to the subgroup size.
UCL	Suggested value for the Shewhart X-bar control chart lower limit. The number of data columns corresponds to the subgroup size.
LWL	Suggested value for the lower warning limit.
UWL	Suggested value for the upper warning limit.

Corrected percentiles	Percentiles estimates which take asymmetry into account. In case of numerical problems, some of the values are not reported.
p	Percentile specification.
Lower, upper	Percentiles, corresponding to p% (lower) or to (100-p)% (upper)

Graphs

Density plot 	Density plot shows the data distribution shape as well as corrected mean (green vertical line), confidence interval for mean value (red line), ±2σ and ±3σ lines (capturing 99.73% of data).
Likelihood for the Box-Cox transformation parameter 	Log likelihood is plotted on y-axis, while the parameter *r* is plotted on x-axis. Maximum is achieved at the optimal *r* value. Horizontal line gives 95% confidence limit for the log likelihood maximum. Vertical lines give the 95% confidence interval for the parameter *r*. If the confidence interval contains 1, no transformation is necessary and the estimates from the Basic data analysis module can be used, or the Transformation module computations can be done with *r*=1.
Skewness plot for the Exponential transformation 	Skewness of the data plotted as a function of the transformation parameter *r*. Zero skewness corresponds to the optimal *r* value. Meaning of this plot is similar to the previous plot meaning. It helps to find the optimum transformation parameter value and to construct its confidence interval. When the vertical green line crosses the curve outside the confidence interval for skewness, (marked by the horizontal line), a transformation is recommended.
QQ-plot before transformation	The same QQ-plot as in the Basic data analysis module. Transformation is a useful remedy mainly for data showing

	systematically bended shape of the QQ-plot, see the left panel. The plot is useful when checking whether any nonlinearity (i.e. non-normality) detected is caused by just a few points or by a general tendency shared by all data.
QQ-plot after transformation 	The plot should be more linear, when the transformation was successful. Statistics given in the protocol should be checked to judge the transformation more properly.