

Item Reliability Analysis



Revised: 10/11/2017



Summary	1
Data Input.....	4
Analysis Options	5
Tables and Graphs.....	5
Analysis Summary	6
Matrix Plot	8
Alpha Plot	10
Correlation Matrix	11
Covariances.....	11

Summary

The **Item Reliability Analysis** is designed to estimate the reliability or consistency of a set of variables. It is commonly used to assess how well a set of questions in a survey, each of which is designed to illicit information about the same characteristic, give consistent results.

The major output of the procedure is Cronbach's alpha. Alpha may be calculated directly from the input variables, or the variables may first be standardized so that they have equal variances. The effect on alpha when each variable is separately omitted is also estimated, so that unreliable questions can be identified.

Sample StatFolio: *itemanalysis.sgp*

Sample Data

The file *itemanalysis.sgd* contains the results of a fictitious survey designed to determine how useful users have found a statistical software program to be. A total of $n = 15$ users were asked to rate their agreement with 6 statements on a scale of 1 to 5, where 1 implies strong disagreement and 5 implies strong agreement. The statements were:

Q1: I use the program frequently.

Q2: Using the program has made me more productive.

Q3: I have obtained important information from my data by using the program.

Q4: The program is more useful than other comparable programs that I have used.

Q5: I am looking forward eagerly to the next version.

Q6: I would recommend the program to a friend or colleague.

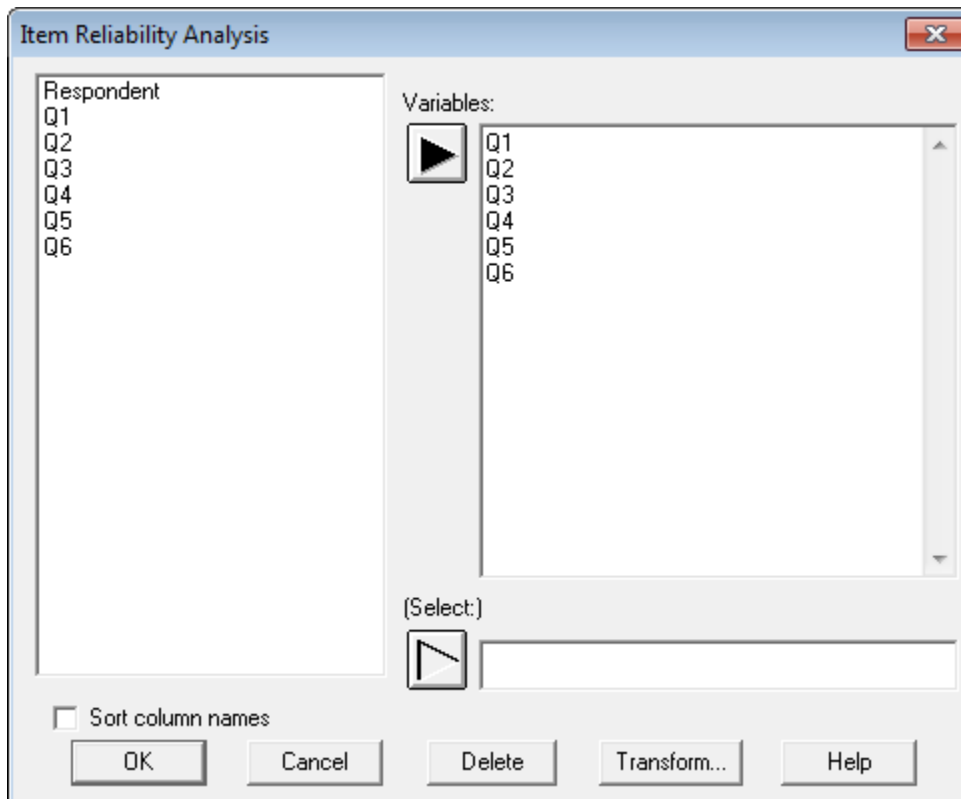
The data are shown below:

Respondent	Q1	Q2	Q3	Q4	Q5	Q6
1	5	4	4	5	5	5
2	4	4	5	4	1	5
3	4	4	4	5	3	4
4	3	3	2	4	3	4
5	5	3	4	4	5	5
6	1	2	2	3	5	3
7	2	2	4	3	3	3
8	1	1	2	3	5	2
9	4	5	4	5	3	5
10	3	4	4	4	2	4
11	4	4	5	5	3	4
12	5	5	4	5	3	5
13	4	3	4	4	2	4
14	3	2	2	4	2	2
15	3	4	3	3	1	1

The primary question of interest is how reliably the questions elicit the user's attitude toward the usefulness of the software. If any questions appear to be unreliable, they might be removed from future surveys.

Data Input

To analyze the data, enter the names of the variables into the following data input dialog box:

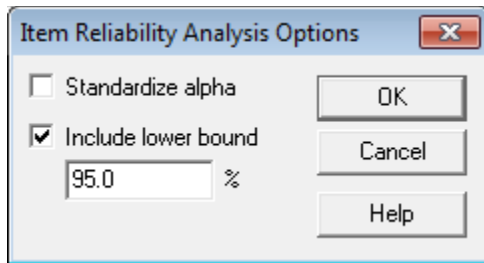


- **Variables:** 2 or more numeric columns containing the data to be analyzed.
- **Select:** subset selection.

In the discussion below, $X_{i,j}$ represents the i -th value of the j -th variable, for $i = 1, \dots, n$ and $j = 1, \dots, k$.

Analysis Options

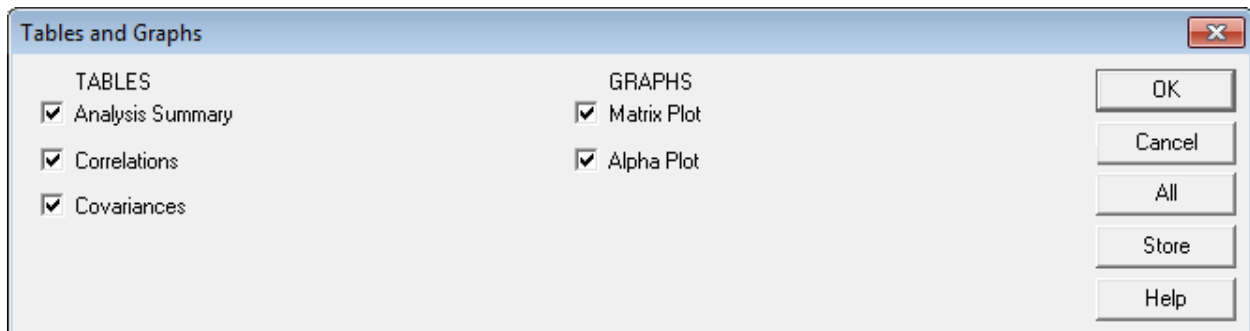
The *Analysis Options* dialog box allows the user to request that the variables be standardized:



Standardization affects the calculated alphas and the item-total correlations. The dialog box also governs whether or not a lower confidence bound for alpha is displayed.

Tables and Graphs

The following tables and graphs are available:



Analysis Summary

The *Analysis Summary* displays important information calculated from the data:

Item Reliability Analysis

Variable	Count	Sample Mean	Std. Deviation
Q1	15	3.4	1.29835
Q2	15	3.33333	1.17514
Q3	15	3.53333	1.0601
Q4	15	4.06667	0.798809
Q5	15	3.06667	1.38701
Q6	15	3.73333	1.27988
Sum	15	21.1333	4.85308

Cronbach's alpha = 0.772503 (lower 95% confidence bound = 0.582388)

Omitted Item Statistics

Omitted Variable	Adj. Sum Mean	Adj. Sum Std. Deviation	Item-Total Correlation	Squared Multiple R	Alpha if Omitted
Q1	17.7333	3.76955	0.782267	0.742626	0.660188
Q2	17.8	4.03909	0.616996	0.728303	0.712931
Q3	17.6	4.13694	0.605884	0.595351	0.719254
Q4	17.0667	4.18273	0.810943	0.738456	0.696108
Q5	18.0667	4.84719	-0.138824	0.451907	0.905959
Q6	17.4	3.73784	0.83015	0.704872	0.645876

The top table shows the number of observations for each variable together with its sample mean and standard deviation. (Note: any rows which do not have complete information on all variables are excluded from the analysis.) Also displayed are statistics for the sum of the variables, defined by:

$$Sum_j = \sum_{i=1}^n X_{i,j}, j = 1, \dots, k \quad (1)$$

Cronbach's Alpha

The output also displays the value of Cronbach's alpha, defined by

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{j=1}^k \sigma_j^2}{\sigma_{sum}^2} \right) \quad (2)$$

where σ_j^2 is the variance of variable j and σ_{sum}^2 is the variance of the sum. Alpha normally ranges between 0 and 1, although in certain cases it can be negative. The larger the value of alpha, the more reliable the set of variables is considered to be. A general rule of thumb requires α to be 0.7 or greater before an instrument will be considered to be reliable. If requested on the

Analysis Options dialog box, a lower confidence bound for alpha is also displayed, calculated from

$$\alpha \geq 1 - \left[(1 - \alpha) F_{p, n-1, (n-1)(k-1)} \right] \quad (3)$$

using the p -th quantile of the indicated F distribution.

An alternative formula for Cronbach's alpha is

$$\alpha = \frac{k\bar{c}}{\bar{v} + (k-1)\bar{c}} \quad (4)$$

where \bar{v} is the average variance of the k variables and \bar{c} is the average of the covariances. If desired, a standardized Cronbach's alpha may be calculated from

$$\alpha_{std} = \frac{k\bar{r}}{1 + (k-1)\bar{r}} \quad (5)$$

where \bar{r} is the average of the correlations amongst all pairs of variables. This may be helpful if there are large differences amongst the variances of the k variables.

In general terms, alpha represents the ratio of the true-score variance to the total-score variance, where the total-score is assumed to equal the true-score plus error.

Omitted Item Statistics

The lower section of the output shows statistics calculated by removing each variable one at a time, leaving all other variables in the analysis. For each variable, it shows:

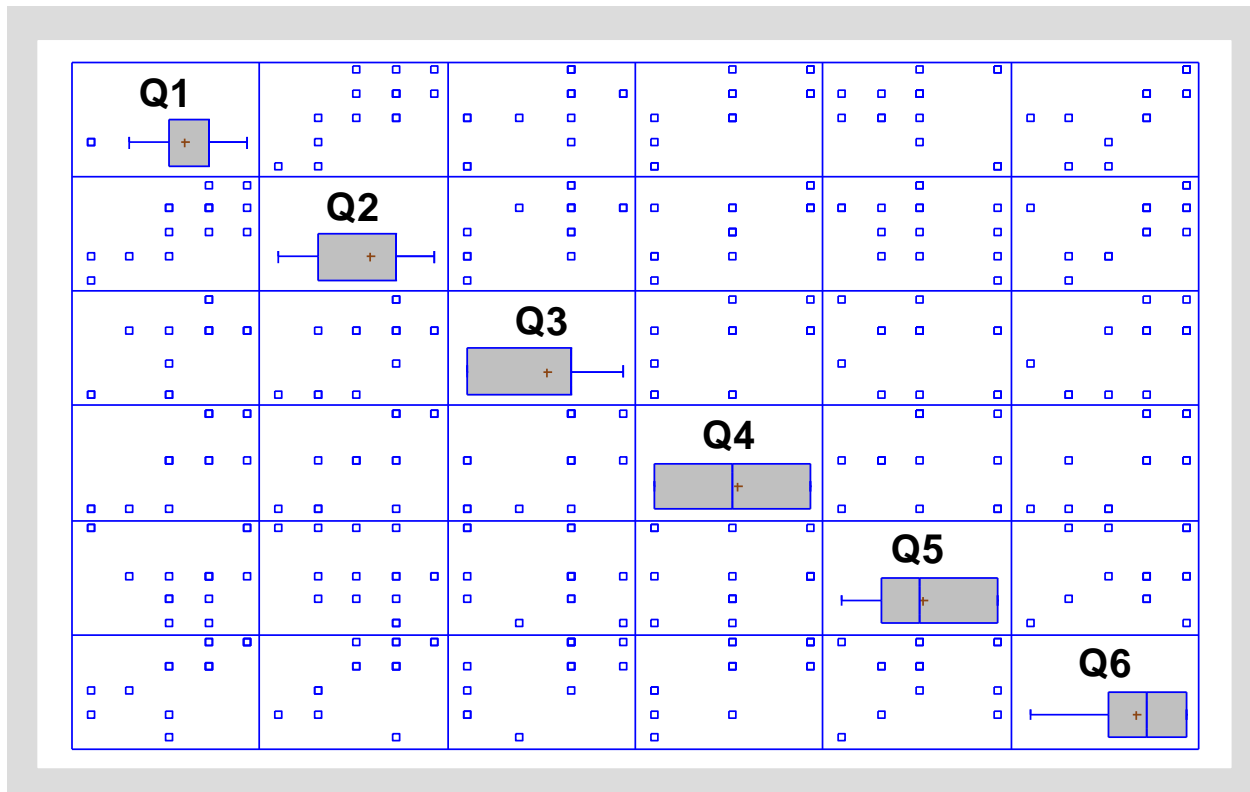
- *Adj. Sum Mean* – the mean of the *Sum* without the omitted variable.
- *Adj. Sum Std. Deviation* – the standard deviation of the *Sum* without the omitted variable.
- *Item – Total Correlation* – the correlation between values of the omitted variable and the *Sum* calculated without the omitted variable.
- *Squared Multiple R* – the squared multiple correlation between the omitted variable and the set of non-omitted variables.
- *Alpha if Omitted* – the value of Cronbach's alpha for the set of variables without the omitted variable.

Note that question *Q5* has a larger *Adj. Sum Std. Deviation* than any of the others and a negative *Item - Total Correlation*. If *Q5* is omitted, Cronbach's alpha jumps from approximately 0.77 to

over 0.9. This is strong evidence that $Q5$ is not providing reliable information about the usefulness of the software.

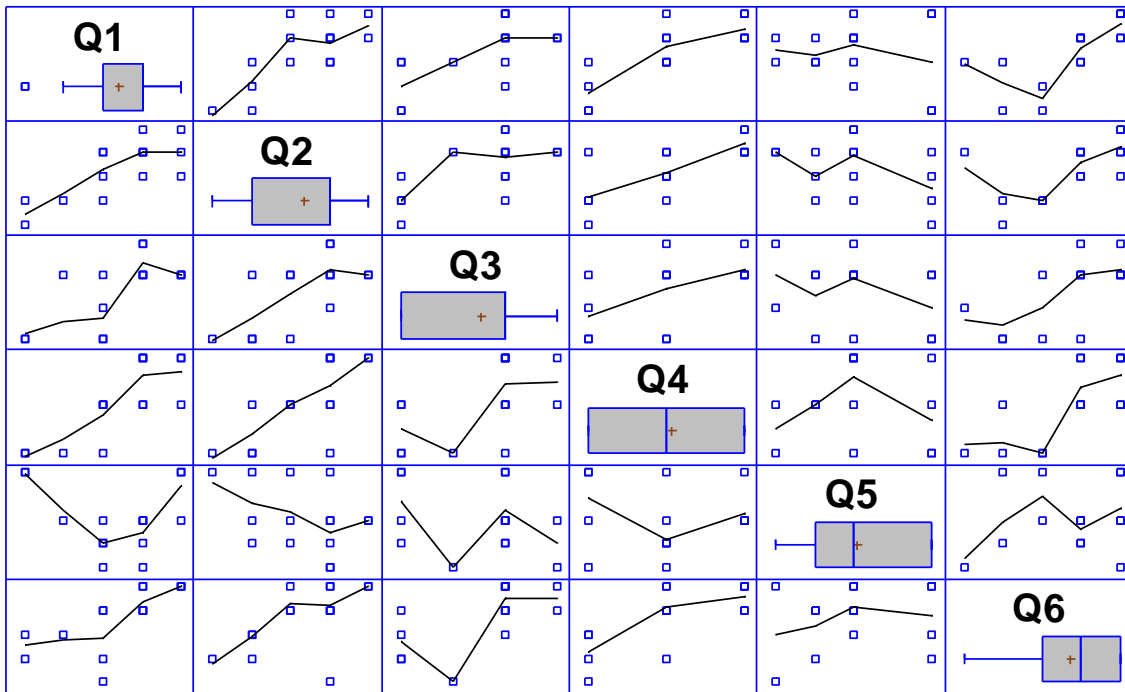
Matrix Plot

The *Matrix Plot* creates a matrix of two variable scatterplots for all pairs of variables.



The scatterplot in row i , column j displays variable i on the vertical axis and variable j on the horizontal axis. In the matrix, every pair of variables is plotted twice, once with the first variable on the X axis and once with that variable on the Y axis. The plot can often be used to identify those variables which are highly correlated, as well as occasional outliers.

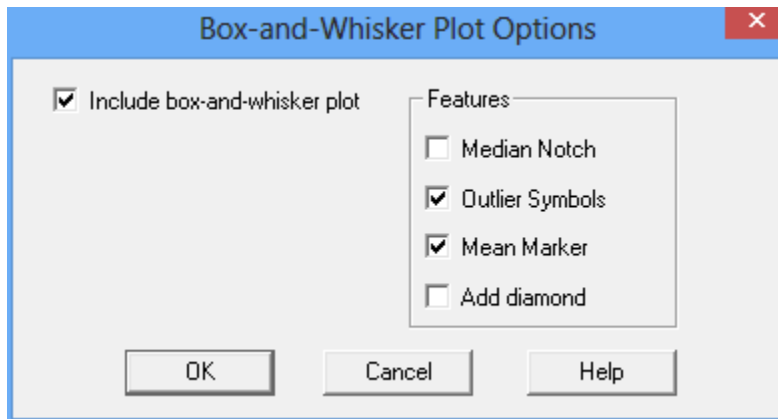
It is sometimes helpful to smooth the scatterplots by pushing the *Smooth/Rotate* button on the analysis toolbar. The plot below uses the default *Robust LOWESS* smoother:



It is now easier to judge the relationships that exist amongst the variables.

Pane Options

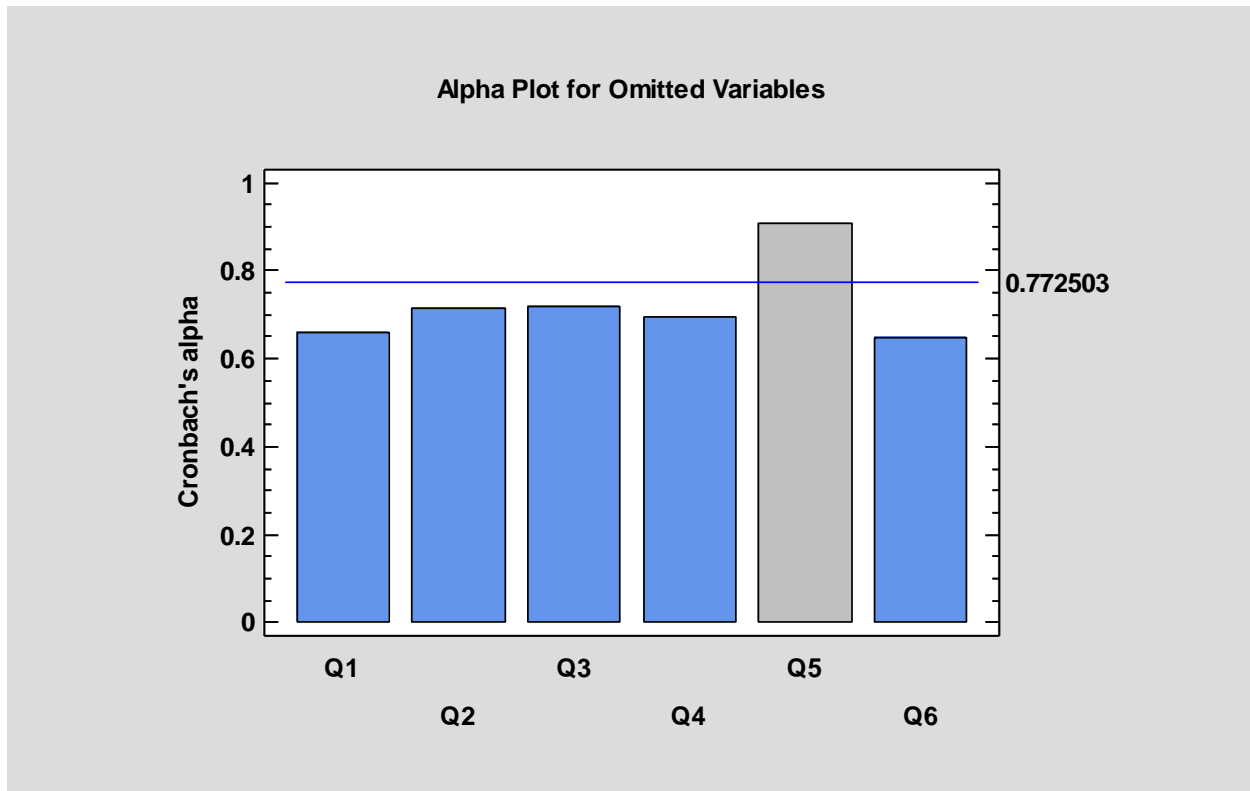
If desired, box-and-whisker plots may be added to the diagonal locations on the plot. Properties of the plots are controlled by the following dialog box:



- **Include box-and-whisker plot:** whether to include box-and-whisker plots in the display.
- **Features:** the box-and-whisker plots may include a notch to indicate a 95% confidence interval for the median, outlier symbols to indicate the presence of outside points, a plus sign to indicate the location of the sample mean, and/or a diamond covering the 95% confidence interval for the mean.

Alpha Plot

The *Alpha Plot* displays the calculated values of Cronbach's alpha:



The horizontal line shows alpha for all of the variables. The heights of the bars show the values that would be calculated if each variable was omitted while all others were retained. Bars for variables that would increase alpha if omitted are colored differently than those that would reduce alpha.

The graph shows clearly the unreliability of Q5.

Correlation Matrix

Correlation coefficients measure the strength of the linear relationship between two variables on a scale of -1 to $+1$. The larger the absolute value of the correlation, the stronger the linear relationship between the two variables. STATGRAPHICS presents correlation coefficients as a matrix:

Correlations					
	Q1	Q2	Q3	Q4	Q5
Q2	0.7490				
Q3	0.6643	0.6498			
Q4	0.7989	0.7356	0.5455		
Q5	-0.1349	-0.3652	-0.2688	-0.0043	
Q6	0.7135	0.5857	0.6388	0.7173	0.1717

Only non-redundant coefficients are displayed.

Covariances

Covariances provide a measure of the extent to which two variables vary together.

Covariances						
	Q1	Q2	Q3	Q4	Q5	Q6
Q1	1.68571					
Q2	1.14286	1.38095				
Q3	0.914286	0.809524	1.12381			
Q4	0.828571	0.690476	0.461905	0.638095		
Q5	-0.242857	-0.595238	-0.395238	-0.0047619	1.92381	
Q6	1.18571	0.880952	0.866667	0.733333	0.304762	1.6381

The covariance between variable x and variable y is calculated from

$$\text{cov}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1} \quad (6)$$