Chapter 401

Correlation Matrix

Introduction

This program calculates matrices of **Pearson product-moment correlations** and **Spearman-rank correlations**. It allows missing values to be deleted in a pair-wise or row-wise fashion.

When someone speaks of a correlation matrix, they usually mean a matrix of Pearson-type correlations. Unfortunately, these correlations are unduly influenced by outliers, unequal variances, nonnormality, and nonlinearities. One of the chief competitors of the Pearson correlation coefficient is the Spearman-rank correlation coefficient. The Spearman correlation is calculated by applying the Pearson correlation formula to the ranks of the data. In so doing, many of the distortions that infect the Pearson correlation are reduced considerably.

A matrix of differences can be displayed to compare the two types of correlation matrices. This allows you to determine which pairs of variables require further investigation.

Partial Correlation

This program lets you specify an optional set of *partial* variables. The linear influence of these variables is removed from the correlation matrix. This provides a statistical adjustment to the correlations among the remaining variables using multiple regression. Note that in the case of Spearman correlations, this adjustment occurs after the complete correlation matrix has been formed.

Heat Maps

Using heat maps to display the features of a correlation matrix was the topic of Friendly (2002) and Friendly and Kwan (2003). This program generates a heat map for various correlation matrices.

Plots of Eigenvectors

Friendly (2002) and Friendly and Kwan (2003) discuss the strengths of plotting the eigenvectors of a correlation matrix. They imply that such a plot is more informative that a heat map. This program generates a plot of the eigenvectors for various correlation matrices.

Another plot that is similar to the eigenvector plot is the map which is provided by a *metric multidimensional scaling* analysis (see the *Multidimensional Scaling* procedure for details).

Discussion

When there is more than one independent variable, the collection of all pair-wise correlations are succinctly represented in a matrix form. In regression analysis, the purpose of examining these correlations is two-fold: to find outliers and to identify collinearity. In the case of outliers, there should be major differences between the parametric measure, the Pearson correlation coefficient, and the nonparametric measure, the Spearman rank correlation coefficient. In the case of collinearity, high pair-wise correlations could be indicators of collinearity problems.

The Pearson correlation coefficient is unduly influenced by outliers, unequal variances, nonnormality, and nonlinearities. As a result of these problems, the Spearman correlation coefficient, which is based on the ranks of the data rather than the actual data, may be a better choice for examining the relationships between variables.

© NCSS, LLC. All Rights Reserved.

Finally, the patterns of missing values in multiple regression and correlation analysis can be very complex. As a result, missing values can be deleted in a pair-wise or a row-wise fashion. If there are only a few observations with missing values, it might be preferable to use the row-wise deletion, especially for large data sets. The row-wise deletion procedure omits the entire observation from the analysis.

On the other hand, if the pattern of missing values is randomly dispersed throughout the data and the use of the row-wise deletion would omit at least 25% of the observations, the pair-wise deletion procedure for missing values would be a safer way to capture the essence of the relationships among the variables. While this method appears to make full use of all your data, the resulting correlation matrix may have mathematical and interpretation difficulties. Mathematically, this correlation matrix may not have a positive determinant. Since each correlation may be based on a different set of rows, practical interpretations could be difficult, if not illogical.

The Spearman correlation coefficient measures the monotonic association between two variables in terms of ranks. It measures whether one variable increases or decreases with another even when the relationship between the two variables is not linear or bivariate normal. Computationally, each of the two variables is ranked separately, and the ordinary Pearson correlation coefficient is computed on the ranks. This nonparametric correlation coefficient is a good measure of the association between two variables when outliers, nonnormality, nonconstant variance, and nonlinearity may exist between the two variables being investigated.

Data Structure

The data are entered as two or more variables. An example of data appropriate for this procedure is shown in the table below. It is assumed that each row gives measurements on the same individual.

Test Scores

Test 1	Test 2	Test 3
45	54	78
87	92	58
55	77	88
44	46	53
73	45	
75	66	66
93	46	85
57	78	91
66	58	77
68	53	73
	45	68
54	65	65
	65	
59	66	72
	54	83
75	53	82

© NCSS, LLC. All Rights Reserved.

Procedure Options

This section describes the options available in this procedure.

Variables Tab

Specify the variables on which to run the analysis.

Data Variables

Correlation Variables

Specify the variables whose correlations are to be formed. Only numeric data values are analyzed.

Partial Variables

An optional set of variables that are to be "partialled out" of the correlation matrix. The influence of these variables is removed from the remaining variables using linear regression. The correlations that are formed are the *partial* correlations.

For the Pearson-type correlations, the resulting matrix is the same that would be formed if the regular variables were regressed on the partial variables, the residuals were stored, and the correlation matrix of these residuals was formed.

Missing Values

Missing Value Removal

This option indicates how you want the program to handle missing values.

- **Pair-wise**

  Pair-wise removal of missing values. Each correlation is based on all pairs of data values in which no missing values occur. Missing values occurring in other variables do not influence this calculation. Note that although this method appears to make full use of all your data, the resulting correlation matrix is difficult to analyze. Mathematically, it may not have a positive determinant. Practically, each correlation may be based on a different set of rows, making it difficult to compare correlations.

- **Row-wise**

  Row-wise removal of missing values. In each row, if a missing value occurs in any of the variables specified, that row of data is ignored in the calculation of all correlations.

Reports Tab

These options specify the reports.

Select Reports

Show Individual Tables

Check this option to display a separate matrix for each statistic checked. After activating this option, you must specify which tables you would like to display.

Pearson Correlations…Count

Check all items you would like it to be displayed as separate tables.

© NCSS, LLC. All Rights Reserved.

Show Combined Tables

Check this option to display a single table containing the selected statistics. After activating this option, you must specify which items you would like to display in the table.

Pearson Correlations…Count

Check all items you would like it to be displayed in the combined table.

Add Cronbach's Alpha beneath the Pearson Correlation matrix

Check this option to display Cronbach's Alpha right after the Pearson correlation table.

Report Options Tab

Variable Names

This option lets you select whether to display variable names, variable labels, or both.

Precision

Specify the precision of numbers in the report. Single precision will display seven-place accuracy, while double precision will display thirteen-place accuracy.

Max Distance Items

This option specifies the maximum size of a distance matrix that will be displayed in the Distance Section report. Distance matrices with more items than this will not be displayed.

This option is here because for large datasets, the distance matrix may be very large.

Max Linkage Clusters

The Linkage Report can be long if the results for all links are printed. This parameter allows you to limit the number of links displayed so that only meaningful values are printed.

Decimal Places

Decimal Places (Correlation, …, Percentages)

These options specify the number of decimal places directly or using an Auto function.

If one of the Auto options is used, the ending zero digits are not shown.

Your choice here will not affect calculations; it will only affect the format of the output.

Auto

If one of the 'Auto' options is selected, the ending zero digits are not shown. For example, if 'Auto (Up to 7)' is chosen,

0.0500 is displayed as 0.05

1.314583689 is displayed as 1.314584

© NCSS, LLC. All Rights Reserved.

Table Formatting

Column Justification

Specify whether data columns in the tables will be left or right justified.

Column Widths

Specify how the widths of columns in the tables will be determined.

The options are

- **Autosize to Minimum Widths**

  Each data column is individually resized to the smallest width required to display the data in the column. This usually results in columns with different widths. This option produces the most compact table possible, displaying the most data per page.

- **Autosize to Equal Minimum Width**

  The smallest width of each data column is calculated and then all columns are resized to the width of the widest column. This results in the most compact table possible where all data columns have the same with. This is the default setting.

- **Custom (User-Specified)**

  Specify the widths (in inches) of the columns directly instead of having the software calculate them for you.

Column Widths (Single Value or List)

Enter one or more values for the widths (in inches) of columns in the tables.

- **Single Value**

  If you enter a single value, that value will be used as the width for all data columns in the table.

- **List of Values**

  Enter a list of values separated by spaces corresponding to the widths of each column. The first value is used for the width of the first data column, the second for the width of the second data column, and so forth. Extra values will be ignored. If you enter fewer values than the number of columns, the last value in your list will be used for the remaining columns.

  Type the word "Autosize" for any column to cause the program to calculate it's width for you. For example, enter "1 Autosize 0.7" to make column 1 be 1 inch wide, column 2 be sized by the program, and column 3 be 0.7 inches wide.

Wrap Column Headings onto Two Lines

Check this option to make column headings wrap onto two lines. Use this option to condense your table when your data are spaced too far apart because of long column headings.

Heat Maps Tab

Specify the six heat maps that can be displayed.

Clustering Options

Max Clusters

Specify the maximum number of clusters allowed in the heat map and reports.

© NCSS, LLC. All Rights Reserved.

Clustering Method

This option specifies which of the nine possible clustering techniques is used. These methods were described earlier. Select 'None' if you want to omit clustering and display the matrix in its original order.

Alpha

Only displayed when the *Flexible Strategy* method is selected. Specifies the values of α_i and α_j. You may enter a number or the letters "NI/NK." The "NI/NK" will cause this constant to be calculated and used as it is in the Centroid and Group Average methods.

Beta

Only displayed when the *Flexible Strategy* method is selected. Specifies the values of β. You may enter a number between -1 and 1 or the letters "NIJ/NK." The "NIJ/NK" will cause this constant to be calculated and used as it is in the Centroid method.

Gamma

Only displayed when the *Flexible Strategy* method is selected. Specifies the values of γ. You may enter any number.

Heat Maps

Diagonal Elements

Specify whether the diagonal elements of heat map are colored or missing.

The choices are

- **Set as missing**

  The diagonal elements will be set to missing values. These elements are displayed using the missing value color of the heat map.

- **Set to 1**

  The diagonal elements will be shown with the color associated with a value of 1.

Select Heat Maps (Pearson Correlations, …, Squared Spearman Correlations)

Check the boxes corresponding to the heat maps to be displayed. For both the Pearson and the Spearman correlation matrices, you can choose to display heat maps of the regular correlation matrix, the absolute values of the correlation matrices, and the squared values of the correlations.

Heat Map Format Buttons

Click the format button to change the heat map settings of the two correlation matrices shown directly above.

Edit During Run

Checking this option will cause the clustered heat map format window to appear when the procedure is run. This allows you to modify the format of the graph with the actual data.

Cluster Analysis Reports

These options let you select which cluster analysis reports you want displayed for each heat map selected.

© NCSS, LLC. All Rights Reserved.

Eigenvectors Tab

These options control the eigenvector plots and reports.

Number of Eigenvectors

Specify the number of eigenvectors (principal components) displayed in the Eigenvector Report. This is also the number of eigenvectors (principal components) plotted. Each unique pair of eigenvectors will be plotted on a separate scatter plot.

Usually, 2 or 3 is all you will need.

Eigenvector Label

Specify the letters, word, or phrase to be used as the labels of the eigenvectors in the plots and reports.

For example, if you enter 'PC', the labels would be PC1, PC2, PC3, etc.

Eigenvector Plots

Select which eigenvector plots you want displayed. A separate plot is displayed for each pair of eigenvectors.

Eigenvector Plots

Select which eigenvector plots you want displayed. A separate plot is displayed for each pair of eigenvectors.

Show the eigenvalue percentage in the eigenvector labels

If checked, the eigenvalue percentage will be added to the eigenvector labels on the eigenvector plots.

For example, if checked, the labels would appear as PC1(61%), PC2(20%), PC3(11%).

Scatter Plot Format Button

Click the format button to change the scatter plot settings of the eigenvector plots.

Edit During Run

Checking this option will cause the corresponding format window to appear when the procedure is run. This allows you to modify the format of the graph while viewing the actual data.

Eigenvalue and Eigenvector Reports

Select which eigenvalue and eigenvector reports you want displayed. This report includes Bartlett's Sphericity Test when the Missing Value Removal option has been set to 'Row wise'.

Storage Tab

Specify if and where the correlation matrices are to be stored.

Data Storage Variables

Pearson Correlations

Specifies columns to receive the Pearson correlation (or partial correlation) matrix.

If you leave this option blank, the matrix is not saved.

If columns are specified, the correlation matrix will be automatically saved into these columns. You must specify as many columns here as there are in your analysis.

Warning
Existing data will be replaced.

© NCSS, LLC. All Rights Reserved.

Spearman Correlations

Specifies columns to receive the Spearman correlation (or partial correlation) matrix.

If you leave this option blank, the matrix is not saved.

If columns are specified, the correlation matrix will be automatically saved into these columns. You must specify as many columns here as there are in your analysis.

Warning

Existing data will be replaced.

Example 1 – Creating a Correlation Matrix

This section presents an example of how to run an analysis of the data contained in the IQ dataset.

You may follow along here by making the appropriate entries or load the completed template **Example 1** by clicking on Open Example Template from the File menu of the procedure window.

1 Open the IQ dataset.

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Click on the file **IQ.NCSS**.
- Click **Open**.

2 Open the Correlation Matrix window.

- Using the Analysis menu or the Procedure Navigator, find and select the **Correlation Matrix** procedure.
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

3 Specify the variables.

- On the Correlation Matrix window, select the **Variables tab**.
- Double-click in the Correlation Variables text box. This will bring up the variable selection window.
- Select **Test1, Test2, Test3, Test4, Test5, IQ** from the list of variables and then click **Ok**. These variables will appear in the Correlation Variables box.
- Enter **Row Wise** in the Missing Value Removal box.

4 Specify the reports.

- On the Correlation Matrix window, select the **Reports tab**.
- Check the **Show Individual Tables** box.
    - Check the **Pearson Correlations** box.
    - Check the **Spearman Correlations** box.
    - Check the **Difference** box.
- Check the **Show Combined Table** box.
    - Check the **Pearson Correlations** box.
    - Check the **Spearman Correlations** box.
    - Check the **Pearson P-Value** box.
    - Check the **Count** box.
- Check the **Add Cronbach's Alpha…** box.

5 Run the procedure.

- From the Run menu, select **Run Procedure**. Alternatively, just click the green Run button.

© NCSS, LLC. All Rights Reserved.

Individual Reports

Pearson Correlation Report
Row-Wise Missing Value Deletion

Variables	Test1	Test2	Test3	Test4	Test5	IQ
Test1	1.0000	0.1000	-0.2608	0.7539	0.0140	0.2256
Test2	0.1000	1.0000	0.0572	0.7196	-0.2814	0.2407
Test3	-0.2608	0.0572	1.0000	-0.1409	0.3473	0.0741
Test4	0.7539	0.7196	-0.1409	1.0000	-0.1729	0.3714
Test5	0.0140	-0.2814	0.3473	-0.1729	1.0000	-0.0581
IQ	0.2256	0.2407	0.0741	0.3714	-0.0581	1.0000

Cronbach's Alpha = 0.4519 Standardized Cronbach's Alpha = 0.4785

Spearman Correlation Report
Row-Wise Missing Value Deletion

Variables	Test1	Test2	Test3	Test4	Test5	IQ
Test1	1.0000	0.0098	-0.3539	0.6517	0.0000	0.2202
Test2	0.0098	1.0000	0.0430	0.6971	-0.3118	0.2303
Test3	-0.3539	0.0430	1.0000	-0.2143	0.3982	0.1238
Test4	0.6517	0.6971	-0.2143	1.0000	-0.1577	0.3772
Test5	0.0000	-0.3118	0.3982	-0.1577	1.0000	-0.0125
IQ	0.2202	0.2303	0.1238	0.3772	-0.0125	1.0000

Difference (Pearson - Spearman) Report
Row-Wise Missing Value Deletion

Variables	Test1	Test2	Test3	Test4	Test5	IQ
Test1	0.0000	0.0902	0.0931	0.1022	0.0140	0.0054
Test2	0.0902	0.0000	0.0142	0.0225	0.0304	0.0104
Test3	0.0931	0.0142	0.0000	0.0734	-0.0509	-0.0497
Test4	0.1022	0.0225	0.0734	0.0000	-0.0152	-0.0058
Test5	0.0140	0.0304	-0.0509	-0.0152	0.0000	-0.0455
IQ	0.0054	0.0104	-0.0497	-0.0058	-0.0455	0.0000

The above tables display the Pearson Correlation Report, Spearman Correlation Report, and the Difference Report. Cronbach's Alpha is displayed at the bottom of the first report.

The Difference report displays the difference between the Pearson and the Spearman correlation coefficients. The report lets you find those variable pairs for which these two correlation coefficients are very different. A large difference indicates the presence of outliers, nonlinearity, nonnormality, and the like. You should investigate scatter plots of pairs of variables with large differences.

Reliability

Because of the central role of measurement in science, scientists of all disciplines are concerned with the accuracy of their measurements. Item analysis is a methodology for assessing the accuracy of measurements that are obtained in the social sciences where precise measurements are often hard to secure. The accuracy of a measurement may be broken down into two main categories: validity and reliability. The validity of an instrument refers to whether it accurately measures the attribute of interest. The reliability of an instrument concerns whether it produces identical results in repeated applications. An instrument may be reliable but not valid. However, it cannot be valid without being reliable.

The methods described here assess the reliability of an instrument. They do not assess its validity. This should be kept in mind when using the techniques of item analysis since they address reliability, not validity.

An instrument may be valid for one attribute but not for another. For example, a driver's license exam may accurately measure an individual's ability to drive. However, it does not accurately measure that individual's ability to do well in college. Hence the exam is reliable and valid for measuring driving ability. It is reliable and invalid for measuring success in college.

© NCSS, LLC. All Rights Reserved.

Several methods have been proposed for assessing the reliability of an instrument. These include the retest method, alternative-form method, split-halves method, and the internal consistency method. We will focus on internal consistency here.

Cronbach's Alpha

Cronbach's alpha (or *coefficient alpha*) is the most popular of the internal consistency coefficients. It is calculated as follows.

$$\alpha = \frac{K}{K-1}\left[1 - \frac{\sum_{i=1}^{K}\sigma_{ii}}{\sum_{i=1}^{K}\sum_{j=1}^{K}\sigma_{ij}}\right]$$

where K is the number of items (questions) and σ_{ij} is the estimated covariance between items i and j. Note the σ_{ii} is the variance (not standard deviation) of item i.

If the data are standardized by subtracting the item means and dividing by the item standard deviations before the above formula is used, we obtain the standardized version of Cronbach's alpha. A little algebra will show that this is equivalent to the following calculations based directly on the correlation matrix of the items.

$$\alpha = \frac{K\overline{\rho}}{1+\overline{\rho}(K-1)}$$

where K is the number of items (variables) and $\overline{\rho}$ is the average of all the correlations among the K items.

Cronbach's alpha has several interpretations. It is equal to the average value of alpha coefficients obtained for all possible combinations of dividing *2K* items into two groups of K items each and calculating the two-half tests. Also, alpha estimates the expected correlation of one instrument with an alternative form containing the same number of items. Furthermore, alpha estimates the expected correlation between an actual test and a hypothetical test which may never be written.

Since Cronbach's alpha is supposed to be a correlation, it should range between -1 and 1. However, it is possible for alpha to be less than -1 when several of the covariances are relatively large, negative numbers. In most cases, alpha is positive, although negative values arise occasionally.

What value of alpha should be achieved? Carmines (1990) stipulates that as a rule, a value of at least 0.8 should be achieved for widely used instruments. An instrument's alpha value may be improved by either adding more items or by increasing the average correlation among the items.

Combined Report

Combined Correlation Report
Row-Wise Missing Value Deletion

Variables		Test1	Test2	Test3	Test4	Test5	IQ
Test1	Pearson Correlation	1.0000	0.1000	-0.2608	0.7539	0.0140	0.2256
	Spearman Correlation	1.0000	0.0098	-0.3539	0.6517	0.0000	0.2202
	Pearson P-Value		0.7228	0.3478	0.0012	0.9606	0.4187
	Count	15	15	15	15	15	15
Test2	Pearson Correlation	0.1000	1.0000	0.0572	0.7196	-0.2814	0.2407
	Spearman Correlation	0.0098	1.0000	0.0430	0.6971	-0.3118	0.2303
	Pearson P-Value	0.7228		0.8395	0.0025	0.3095	0.3876
	Count	15	15	15	15	15	15

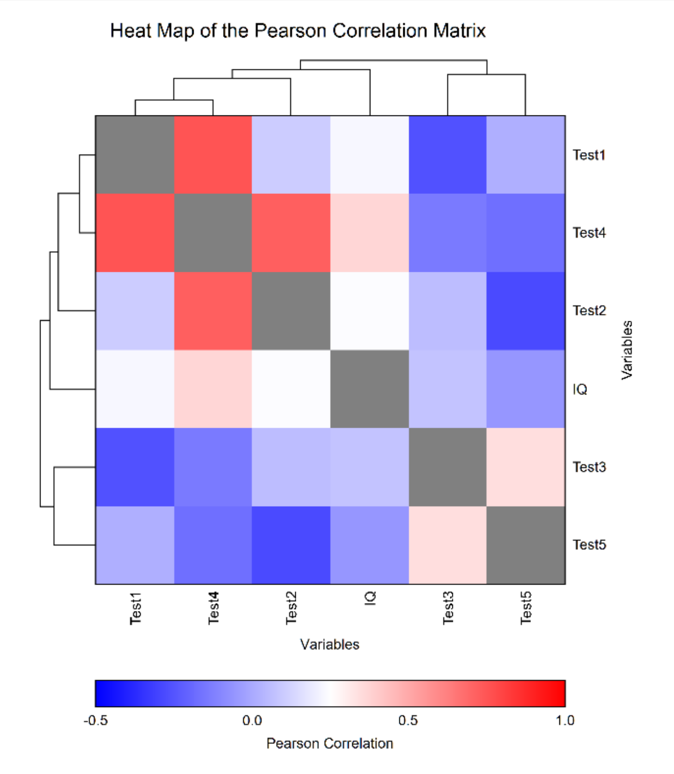
© NCSS, LLC. All Rights Reserved.

Test3	Pearson Correlation	-0.2608	0.0572	1.0000	-0.1409	0.3473	0.0741
	Spearman Correlation	-0.3539	0.0430	1.0000	-0.2143	0.3982	0.1238
	Pearson P-Value	0.3478	0.8395		0.6164	0.2046	0.7931
	Count	15	15	15	15	15	15
Test4	Pearson Correlation	0.7539	0.7196	-0.1409	1.0000	-0.1729	0.3714
	Spearman Correlation	0.6517	0.6971	-0.2143	1.0000	-0.1577	0.3772
	Pearson P-Value	0.0012	0.0025	0.6164		0.5378	0.1729
	Count	15	15	15	15	15	15
Test5	Pearson Correlation	0.0140	-0.2814	0.3473	-0.1729	1.0000	-0.0581
	Spearman Correlation	0.0000	-0.3118	0.3982	-0.1577	1.0000	-0.0125
	Pearson P-Value	0.9606	0.3095	0.2046	0.5378		0.8371
	Count	15	15	15	15	15	15
IQ	Pearson Correlation	0.2256	0.2407	0.0741	0.3714	-0.0581	1.0000
	Spearman Correlation	0.2202	0.2303	0.1238	0.3772	-0.0125	1.0000
	Pearson P-Value	0.4187	0.3876	0.7931	0.1729	0.8371	
	Count	15	15	15	15	15	15

Cronbach's Alpha = 0.4519 Standardized Cronbach's Alpha = 0.4785

The above report displays the Pearson and Spearman correlations, the significance level of a test of the Pearson correlation (Pearson P-Value) and count for each pair of variables.

Heat Map of the Pearson Correlation Matrix



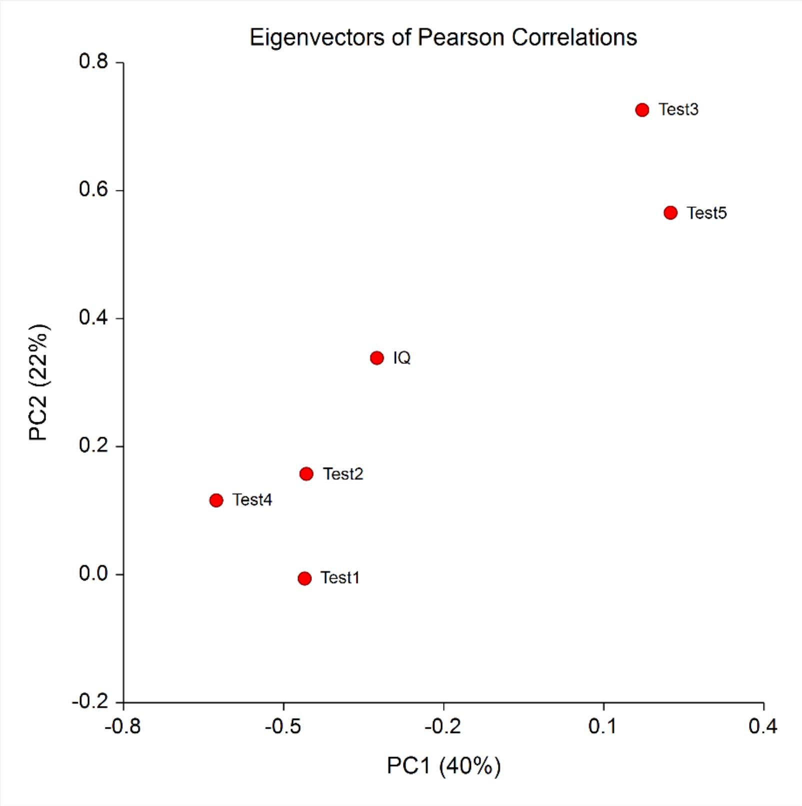
This report displays a heat map of the correlation matrix. Note that the rows and columns are sorted in the order suggested by the hierarchical clustering.

This plot allows you to discover various subsets of the variables that seem to be highly correlated within the subset. You can see that Test1, Test4, and Test2 seem to be highly related. Similarly, Test3 and Test5 seem to be related.

This plot was suggested by Friendly (2002) and Friendly and Kwan (2003).

© NCSS, LLC. All Rights Reserved.

Pearson Eigenvectors Plot(s)



This plot displays a scatter plot of PC1 (the first eigenvector) on the horizontal axis and PC2 (the second eigenvector) on the vertical axis. The number within the parentheses are the percentage of the sum of the eigenvalues that that is accounted for by the corresponding eigenvector. For example, in this plot, 40% of the variability in the correlation matrix is accounted for by the first eigenvector and 22% of the variability is accounted for by the second eigenvector. Thus, the two eigenvectors in this plot account for 62% of the variation among the correlations.

Note that this plot lets you see which variables to be clustered. In this case, Test3 and Test5 are related as are Test1, Test2, and Test4. The IQ variable seems to be by itself, although it is somewhat similar to the second three variables.

This is the same interpretation that we obtained from the heat map, but perhaps it is easier to see subtleties in this plot.

This plot was suggested by Friendly (2002) and Friendly and Kwan (2003).

Storing the Correlations on the Database

When you specify variables in either the Pearson Correlations or the Spearman Correlations boxes, the correlation matrix will be stored in those variables during the execution of the program.

© NCSS, LLC. All Rights Reserved.

Example 2 – Bartlett's Sphericity Test

This section presents an example of how to run Bartlett's Sphericity test of the data contained in the IQ dataset. Note that Bartlett's test is only available when Missing Value Removal is set to *Row Wise*.

You may follow along here by making the appropriate entries or load the completed template **Example 2** by clicking on Open Example Template from the File menu of the procedure window.

1 Open the IQ dataset.

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Click on the file **IQ.NCSS**.
- Click **Open**.

2 Open the Correlation Matrix window.

- Using the Analysis menu or the Procedure Navigator, find and select the **Correlation Matrix** procedure.
- On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

3 Specify the variables.

- Select the **Variables tab**.
- Double-click in the Correlation Variables text box. This will bring up the variable selection window.
- Select **Test1, Test2, Test3, Test4, Test5, IQ** from the list of variables and then click **Ok**. These variables will appear in the Correlation Variables box.
- Select **Row Wise** for the Missing Value Removal method.

4 Specify the reports.

- Select the **Reports tab**.
- Check the **Show Individual Tables** box.
    - Check the **Pearson Correlations** box.

5 Specify the Eigenvectors.

- Select the **Eigenvectors tab**.
- Check the **Show Individual Tables** box.
    - Check the **Pearson Correlations** box.
- Check the **Pearson Eigenvector Plot(s)** box.
- Check the **Show the eigenvalue percentages …** box.
- Check the **Eigenvalues and Eigenvectors of the Pearson Correlations …** box.

6 Run the procedure.

- From the Run menu, select **Run Procedure**. Alternatively, just click the green Run button.

Individual Reports

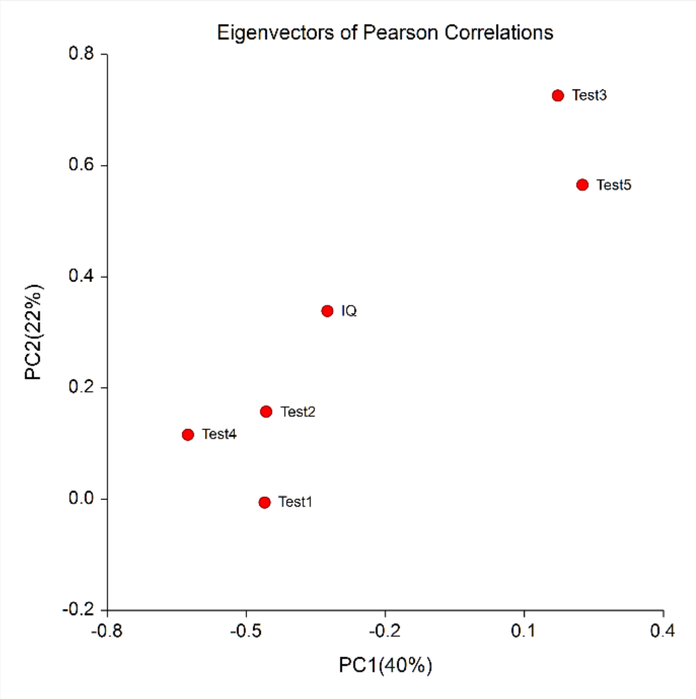
Pearson Correlation Report
Row-Wise Missing Value Deletion

Variables	Test1	Test2	Test3	Test4	Test5	IQ
Test1	1.0000	0.1000	-0.2608	0.7539	0.0140	0.2256
Test2	0.1000	1.0000	0.0572	0.7196	-0.2814	0.2407
Test3	-0.2608	0.0572	1.0000	-0.1409	0.3473	0.0741
Test4	0.7539	0.7196	-0.1409	1.0000	-0.1729	0.3714
Test5	0.0140	-0.2814	0.3473	-0.1729	1.0000	-0.0581
IQ	0.2256	0.2407	0.0741	0.3714	-0.0581	1.0000

The above table displays the Pearson Correlation Report.

© NCSS, LLC. All Rights Reserved.

Pearson Eigenvector Plot



This plot displays a scatter plot of PC1 (the first eigenvector) on the horizontal axis and PC2 (the second eigenvector) on the vertical axis. The number within the parentheses are the percentage of the sum of the eigenvalues that that is accounted for by the corresponding eigenvector. For example, in this plot, 40% of the variability in the correlation matrix is accounted for by the first eigenvector and 22% of the variability is accounted for by the second eigenvector. Thus, the two eigenvectors in this plot account for 62% of the variation among the correlations.

Note that this plot lets you see which variables could be clustered. In this case, Test3 and Test5 are related as are Test1, Test2, and Test4. The IQ variable seems to be by itself, although it is somewhat similar to the second three variables.

Eigenvalues Report

Eigenvalues of Pearson Correlation Matrix
Row-Wise Missing Value Deletion

Eigenvector	Eigenvalue	Individual Percent	Cumulative Percent	Scree Plot
PC1	2.374012	39.57	39.57	||||||||
PC2	1.297129	21.62	61.19	|||||
PC3	1.109029	18.48	79.67	||||
PC4	0.779485	12.99	92.66	|||
PC5	0.435845	7.26	99.92	||
PC6	0.004500	0.08	100.00	|

Log(Det|R) = -5.254947 Bartlett Sphericity Test = 58.68 DF = 15 Prob Level = 0.000000

The above report displays the Pearson and Spearman correlations, the significance level of a test of the Pearson correlation (Pearson P-Value) and count for each pair of variables.

Eigenvector

This column gives the label of the eigenvector whose eigenvalue is displayed. Note that you can modify the label.

© NCSS, LLC. All Rights Reserved.

Eigenvalue

The eigenvalues. Often, these are used to determine how many eigenvectors to retain. (In this example, we would retain the first three.)

One rule-of-thumb is to retain those eigenvectors whose eigenvalues are greater than one. The sum of the eigenvalues is equal to the number of variables. Hence, in this example, the first eigenvector retains the information contained in 2.37 of the original variables.

Individual and Cumulative Percents

The first column gives the percentage of the total variation in the variables accounted for by this eigenvector. The second column is the cumulative total of the percentage. Some authors suggest that the user pick a cumulative percentage, such as 80% or 90%, and keep enough factors to attain this percentage.

Scree Plot

This is a rough bar plot of the eigenvalues. It enables you to quickly note the relative size of each eigenvalue. Many authors recommend it as a method of determining how many eigenvectors to plot.

The word *scree*, first used by Cattell (1966), is usually defined as the rubble at the bottom of a cliff. When using the scree plot, you must determine which eigenvalues form the “cliff” and which form the “rubble.” You keep the eigenvectors that make up the cliff. Cattell and Jaspers (1967) suggest keeping those that make up the cliff plus the first eigenvector of the rubble.

Log(Det|R|)

This is the log (base e) of the determinant of the correlation matrix.

Bartlett Test, DF, Prob Level

This is Bartlett’s sphericity test (Bartlett, 1950) for testing the null hypothesis that the correlation matrix is an identity matrix (all correlations are zero). If you get a significance level (Prob Level) greater than 0.05, there is no evidence that any of the correlations are different from zero. The test is valid for large samples ($N>150$). It uses a Chi-square distribution with $p(p-1)/2$ degrees of freedom.

Note that this test is only available when the Missing Value Removal option is set to *Row Wise*.

The formula for computing this test is:

$$\chi^2 = \frac{(11 + 2p - 6N)}{6} \operatorname{Log}_e |R|$$

Eigenvectors Report

Eigenvectors of Pearson Correlation Matrix
Row-Wise Missing Value Deletion

Variables	Eigenvectors PC1	PC2
Test1	-0.4608	-0.0060
Test2	-0.4575	0.1575
Test3	0.1720	0.7261
Test4	-0.6263	0.1161
Test5	0.2251	0.5656
IQ	-0.3253	0.3386

The eigenvectors show the direction of each factor (principal component) after the correlation matrix is suitably scaled and rotated. These are the values that are plotted in the Eigenvector plots shown above.

© NCSS, LLC. All Rights Reserved.