

## Lecture 11: Factor Analysis using SPSS

### Task 1: Factor Analysis

You may be interested to investigate the reasons why customers buy a product such as a particular brand of soft drink (e.g. coca cola). Several variables were identified which influence customer to buy coca cola. Some of the variables identified as being influential include *cost of product*, *quality of product*, *availability of product*, *quantity of product*, *respectability of product*, *prestige attached to product*, *experience with product*, and *popularity of product*. From this, you designed a questionnaire to solicit customers' view on a seven point scale, where 1 = not important and 7 = very important. The results from your questionnaire are show on the table below. Only the first twelve respondents (cases) are used in this example. (Enter the *Dataset 1- 'Cunsumer Survey'* into SPSS **Data Editor** window.)

**Factor Analysis:** Factor analysis is used to find factors among observed variables. In other words, if your data contains many variables, you can use factor analysis to reduce the number of variables. Factor analysis groups variables with similar characteristics together. With factor analysis you can produce a small number of factors from a large number of variables which is capable of explaining the observed variance in the larger number of variables. The reduced factors can also be used for further analysis.

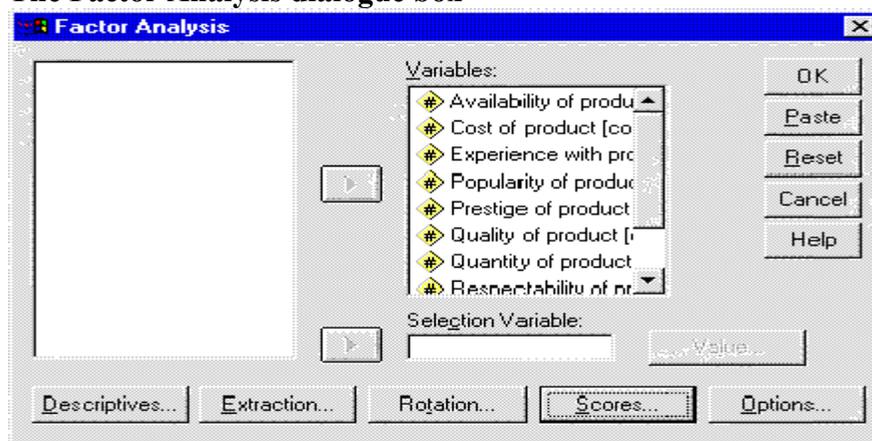
There are three stages in factor analysis:

1. First, a correlation matrix is generated for all the variables. A correlation matrix is a rectangular array of the correlation coefficients of the variables with each other.
2. Second, factors are extracted from the correlation matrix based on the correlation coefficients of the variables.
3. Third, the factors are rotated in order to maximize the relationship between the variables and some of the factors.

### Tips for Task 1: Running the Factor Analysis Procedure

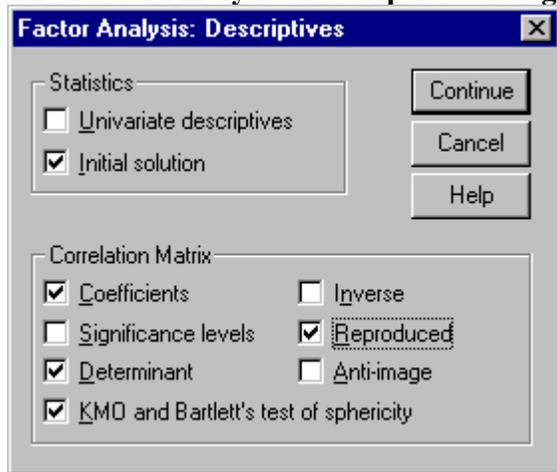
From the menu bar select **Analyze** and choose **Data Reduction** and then click on **Factor**. Highlight related variables and send them to variables lists. Then select some options and run the procedure.

### The Factor Analysis dialogue box



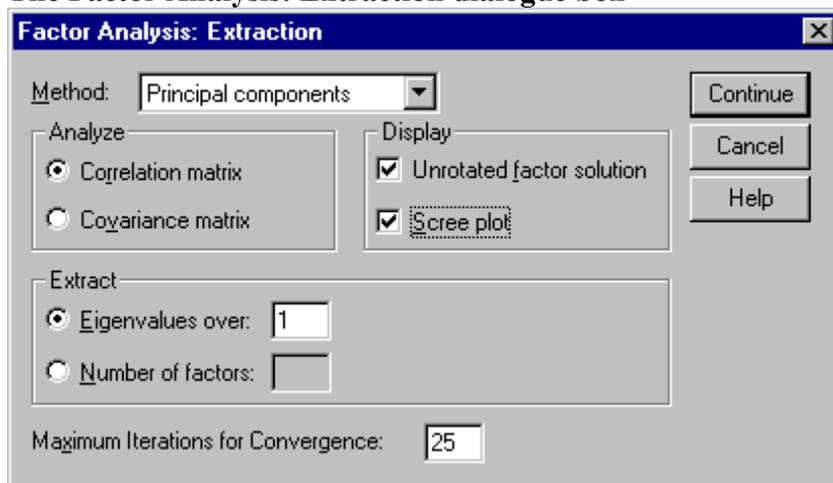
Click on the **DESCRIPTIVES** button and its dialogue box will be loaded on the screen. Within this dialogue box select the following check boxes **Univariate Descriptives**, **Coefficients**, **Determinant**, **KMO and Bartlett's test of sphericity**, and **Reproduced**. Click on **Continue** to return to the **Factor Analysis** dialogue box. The **Factor Analysis: Descriptives** dialogue box should be completed as shown below.

### The Factor Analysis: Descriptives dialogue box



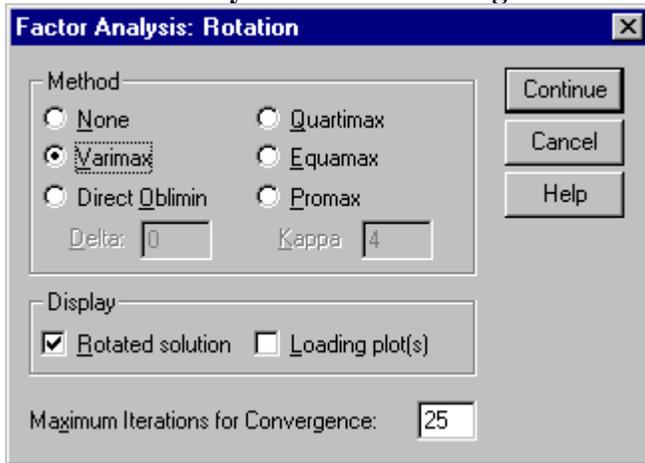
From the **Factor Analysis** dialogue box click on the **EXTRACTION** button and its dialogue box will be loaded on the screen. Select the check box for **Scree Plot**. Click on **Continue** to return to the **Factor Analysis** dialogue box. The **Factor Analysis: Extraction** dialogue box should be completed as shown below.

### The Factor Analysis: Extraction dialogue box



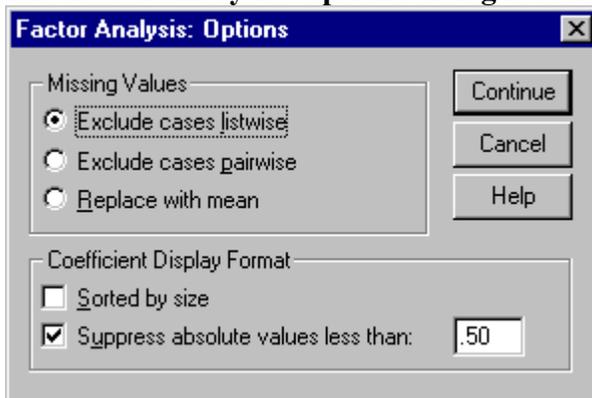
From the **Factor Analysis** dialogue box click on the **ROTATION** button and its dialogue box will be loaded on the screen. Click on the radio button next to **Varimax** to select it. Click on **Continue** to return to the **Factor Analysis** dialogue box. The **Factor Analysis: Rotation** dialogue box should be completed as shown below.

### The Factor Analysis: Rotation dialogue box



From the **Factor Analysis** dialogue box click on the **OPTIONS** button and its dialogue box will be loaded on the screen. Click on the check box of **Suppress absolute values less than** to select it. Type **0.50** in the text box. Click on **Continue** to return to the **Factor Analysis** dialogue box. Click on **OK** to run the procedure. The **Factor Analysis: Options** dialogue box should be completed as shown below.

### The Factor Analysis: Options dialogue box



### Interpretation of the Output

#### Descriptive Statistics

The first output from the analysis is a table of descriptive statistics for all the variables under investigation. Typically, the *mean*, *standard deviation* and *number of respondents* (N) who participated in the survey are given. Looking at the *mean*, one can conclude that *respectability of product* is the most important variable that influence customers to buy the product. It has the highest *mean* of 6.08.

## Descriptive Statistics

	Mean	Std. Deviation	Analysis N
Availability of product	5.08	.79	12
Cost of product	2.42	1.00	12
Experience with product	3.92	1.16	12
Popularity of product	4.25	.87	12
Prestige of product	3.42	1.16	12
Quality of product	3.58	.79	12
Quantity of product	5.75	1.29	12
Respectability of product	6.08	1.00	12

## The Correlation matrix

The next output from the analysis is the correlation coefficient. A correlation matrix is simply a rectangular array of numbers which gives the correlation coefficients between a single variable and every other variables in the investigation. The correlation coefficient between a variable and itself is always 1, hence the principal diagonal of the correlation matrix contains 1s. The correlation coefficients above and below the principal diagonal are the same. The determinant of the correlation matrix is shown at the foot of the table below.

Correlation Matrix<sup>a</sup>

	Availability of product	Cost of product	Experience with product	Popularity of product	Prestige of product	Quality of product	Quantity of product	Respectability of product
Correlation Availability of product	1.000	.527	-.386	-.298	-.041	.349	.467	.105
Cost of product	.527	1.000	-.202	-.237	.072	.585	.372	.420
Experience with product	-.386	-.202	1.000	.563	.564	-.238	-.682	-.620
Popularity of product	-.298	-.237	.563	1.000	.248	.165	-.509	-.237
Prestige of product	-.041	.072	.564	.248	1.000	-.484	-.470	-.660
Quality of product	.349	.585	-.238	.165	-.484	1.000	.245	.508
Quantity of product	.467	.372	-.682	-.509	-.470	.245	1.000	.655
Respectability of product	.105	.420	-.620	-.237	-.660	.508	.655	1.000

a. Determinant = 1.731E-03

**Kaiser-Meyer-Olkin (KMO) and Bartlett's Test** : measures strength of the relationship among variables

The **KMO** measures the *sampling adequacy which should be greater than 0.5 for a satisfactory factor analysis* to proceed. If any pair of variables has a value less than this , consider dropping one of them from the analysis. The off-diagonal elements should all be very small (close to zero) in a good model. Looking at the table below, the KMO measure is 0.417.

There is no significant answer to question “How many cases do I need to factor analysis?”, and methodologies differ. A common rule is to suggest that a researcher has at least 10-15 participant per variable. Fiedel (2005) says that in general over 300 cases for sampling analysis is probably adequate. There is universal agreement that factor analysis is inappropriate when sample size is below 50. Kaisen (1974) recommend 0.5 as minimum (barely accepted), values between 0.7-0.8 acceptable, and values above 0.9 are superb.

**Bartlett's test** is another indication of the strength of the relationship among variables. This tests the null hypothesis that the correlation matrix is an identity matrix. An identity matrix is

matrix in which all of the diagonal elements are 1 and all off diagonal elements are 0. You want to reject this null hypothesis. From the same table, we can see that the **Bartlett's test** of sphericity is significant. That is, its associated probability is less than 0.05. In fact, it is actually 0.012, i.e. the significance level is small enough to reject the null hypothesis. This means that correlation matrix is not an identity matrix

**KMO and Bartlett's Test**

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.417
Bartlett's Test of Sphericity	Approx. Chi-Square	47.694
	df	28
	Sig.	.012

**Communalities**

The next item from the output is a table of communalities which shows how much of the variance in the variables has been accounted for by the extracted factors. For instance over 90% of the variance in *quality of product* is accounted for while 73.5% of the variance in *availability of product* is accounted for.

**Communalities**

	Initial	Extraction
Availability of product	1.000	.735
Cost of product	1.000	.844
Experience with product	1.000	.800
Popularity of product	1.000	.804
Prestige of product	1.000	.865
Quality of product	1.000	.918
Quantity of product	1.000	.768
Respectability of product	1.000	.814

Extraction Method: Principal Component Analysis.

**Total Variance Explained**

The next item shows all the factors extractable from the analysis along with their eigenvalues, the percent of variance attributable to each factor, and the cumulative variance of the factor and the previous factors. Notice that the first factor accounts for 46.367% of the variance, the second 18.471% and the third 17.013%. All the remaining factors are not significant.

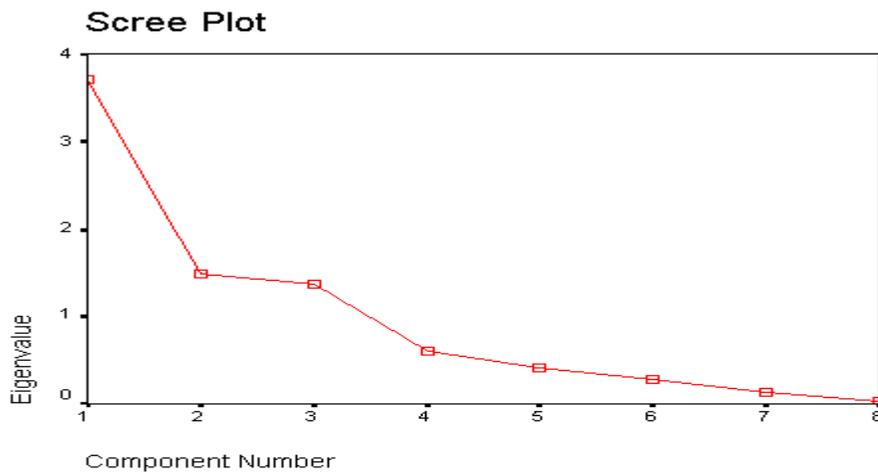
Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.709	46.367	46.367	3.709	46.367	46.367	2.501	31.260	31.260
2	1.478	18.471	64.838	1.478	18.471	64.838	2.045	25.566	56.826
3	1.361	17.013	81.850	1.361	17.013	81.850	2.002	25.024	81.850
4	.600	7.499	89.349						
5	.417	5.214	94.563						
6	.281	3.508	98.071						
7	.129	1.608	99.679						
8	2.569E-02	.321	100.000						

Extraction Method: Principal Component Analysis.

**Scree Plot**

The scree plot is a graph of the eigenvalues against all the factors. The graph is useful for determining how many factors to retain. The point of interest is where the curve starts to flatten. It can be seen that the curve begins to flatten between factors 3 and 4. Note also that factor 4 has an eigenvalue of less than 1, so only three factors have been retained.



**Eigenvalue:** The standardized variance associate with a particular factor. The sum of the eigenvalues can not exceed the number of items in the analysis, since each item contributes one to the sum of variances.

**Component (Factor) Matrix**

The table below shows the loadings of the eight variables on the three factors extracted. The higher the absolute value of the loading, the more the factor contributes to the variable. The gap on the table represent loadings that are less than 0.5, this makes reading the table easier. We suppressed all loadings less than 0.5.

**Component Matrix**

	Component		
	1	2	3
Availability of product	.546		
Cost of product	.563	.697	
Experience with product	-.817		
Popularity of product	-.534		.598
Prestige of product	-.660		-.526
Quality of product	.578	.570	.508
Quantity of product	.841		
Respectability of product	.815		

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

### Rotated Component (Factor) Matrix

The idea of rotation is to reduce the number factors on which the variables under investigation have high loadings. Rotation does not actually change anything but makes the interpretation of the analysis easier. Looking at the table below, we can see that *availability of product*, and *cost of product* are substantially loaded on Factor (Component) 3 while *experience with product*, *popularity of product*, and *quantity of product* are substantially loaded on Factor 2. All the remaining variables are substantially loaded on Factor 1. These factors can be used as variables for further analysis.

**Rotated Component Matrix**

	Component		
	1	2	3
Availability of product			.760
Cost of product			.908
Experience with product	-.560	.684	
Popularity of product		.893	
Prestige of product	-.901		
Quality of product	.637		.634
Quantity of product		-.652	
Respectability of product	.845		

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 7 iterations.

### Conclusion

You should now be able to perform a factor analysis and interpret the output. Many other items are produce in the output, for the purpose of this illustration they have been ignored. Note that the correlation matrix can used as input to factor analysis. In this case you have to use SPSS command syntax which is outside the scope of this document.