Block 17: Factor analysis

(Activity solutions can be found at the end of the document.)

In regression, a dependent variable is clearly identified. In **factor analysis**, variables are not classified as independent nor dependent. All *interdependent relationships* among variables are examined. The **factor model** is introduced followed by the steps taken in factor analysis.

Learning Objectives

- Describe the concept of factor analysis and explain how it is different from analysis of variance, multiple regression and discriminant analysis
- Discuss the procedure for conducting factor analysis, including problem formulation, construction of the correlation matrix, selection of an appropriate method, determination of the number of factors, rotation and interpretation of factors
- Explain the selection of surrogate variables and their application, with emphasis on their use in subsequent analysis
- Describe the procedure for determining the fit of a factor analysis model using the observed and the reproduced correlations.

Reading List

Malhotra, N.K., D. Nunan and D.F. Birks. Marketing Research: An Applied Approach. (Pearson, 2017) 5th edition [ISBN 9781292103129] Chapter 24.

17.1 Factor analysis

For each section of *Factor analysis*, use the LSE ELearning resources to test your knowledge with the Key terms and concepts flip cards.

Structural equation modelling (SEM)

Structural equation modelling (SEM) estimates the unknown coefficients in a *set of linear structural equations*. Variables in the equation system are usually *directly observed variables plus unmeasured latent variables* which are not observed, but relate to the observed variables.

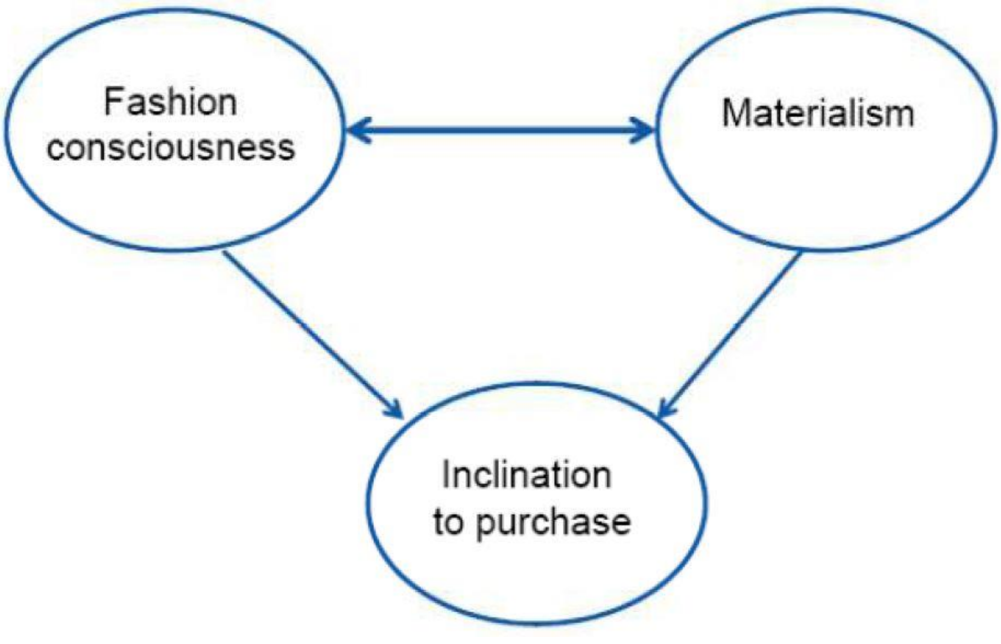
SEM assumes there is a *causal structure* among a set of latent variables, and that the observed variables are *indicators* of the latent variables. The latent variables may appear as linear combinations of observed variables, or they may be intervening variables in a causal chain.

SEM is a collection of statistical techniques including **factor analysis** and **multiple regression**. It allows the researcher to examine relationships between several continuous or discrete independent variables and several continuous or discrete dependent variables. The independent and dependent variables can be latent or measured variables.

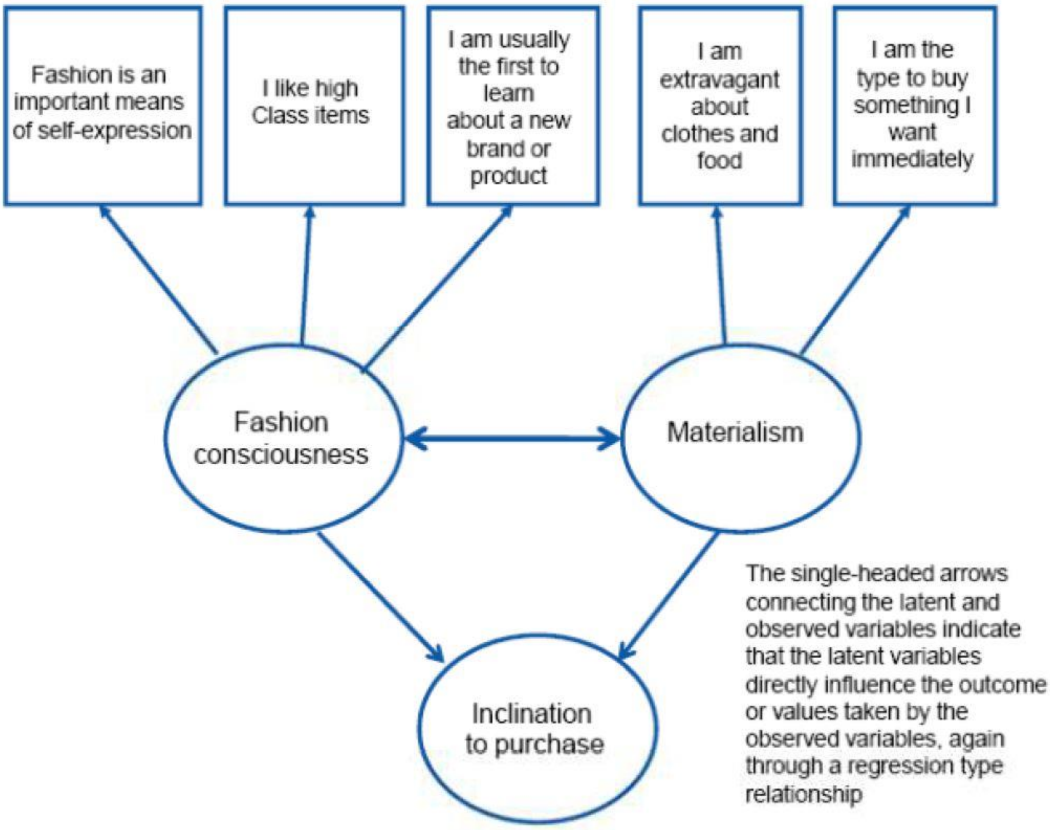
In pictorial form, latent variables can be represented as ellipses:

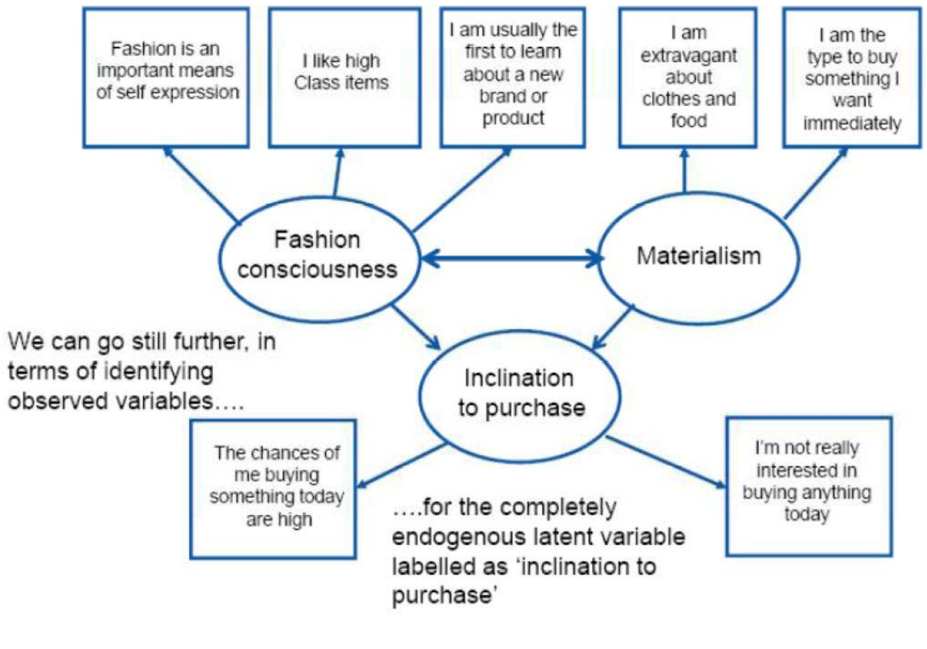
Latent variables can be correlated:

Latent variables can also influence other latent variables directly, via a regression-type relationship:

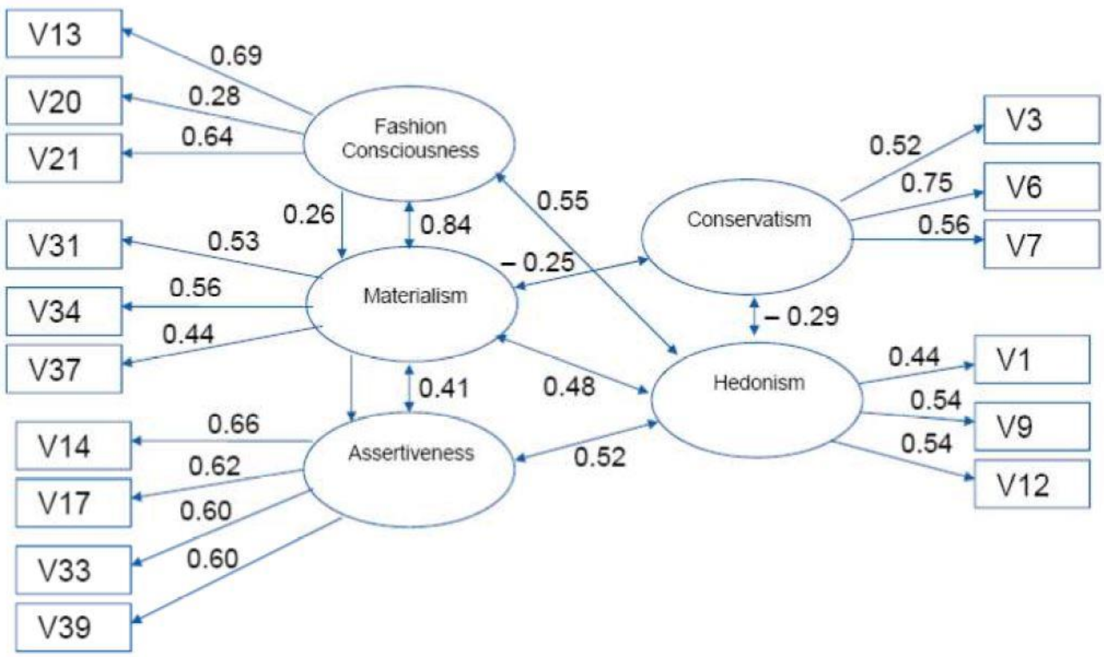


In pictorial form, observed or indicator variables can be represented as rectangles:





In the following **path diagram**, latent constructs (unmeasured variables) are shown in ellipses and the questionnaire items used to measure these latent constructs (i.e. measured variables) are shown in rectangles. Arrows pointing from the circles to the rectangles are equivalent to **factor loadings** in factor analysis.



Activity 17.1

Explain what structural equation modelling attempts to achieve. How can SEM be applied in market research?

Overview

Factor analysis is a general name denoting a class of procedures primarily used for *data reduction and summarisation*. Factor analysis is an **interdependence technique** in that an entire set of interdependent relationships is examined without making the distinction between dependent and independent variables.

Factor analysis is used in the following circumstances.

- To *identify underlying dimensions*, or factors, which explain the correlations among a set of variables.
- To *identify a new, smaller, set of uncorrelated variables* to replace the original set of correlated variables in subsequent multivariate analysis (such as regression or discriminant analysis).
- To *identify a smaller set of salient variables* from a larger set for use in subsequent multivariate analysis.

Figure 24.1 of the textbook shows factors underlying selected psychographics and lifestyles.

Activity 17.2

How is factor analysis different from multiple regression and discriminant analysis?

Activity 17.3

What are the major uses of factor analysis?

The factor analysis model

Mathematically, *each variable is expressed as a linear combination of underlying factors*. The covariation among the variables is described in terms of a small number of *common factors* plus a *unique factor* for each variable.

If the variables are *standardised*, the **factor model** may be represented as:

$$X_i = A_{i1}F_1 + A_{i2}F_2 + A_{i3}F_3 + \cdots + A_{im}F_m + V_iU_i$$

where:

- X_i = the iith standardised variable
- A_{ij} = the standardised multiple regression coefficient of variable ii on common factor jj
- F = a common factor
- V_i = the standardised regression coefficient of variable ii on unique factor ii
- U_i = the unique factor for variable ii
- M = the number of common factors.

The unique factors are correlated with each other and with the common factors. The common factors themselves can be expressed as *linear combinations* of the observed variables:

$$F_i = W_{i1}X_1 + W_{i2}X_2 + W_{i3}X_3 + \cdots + W_{ik}k$$

where:

- F_i= the estimate of the iith factor
- W_i= the weight or factor score coefficient
- k= the number of variables.

It is possible to select weights or **factor score coefficients** so that the first factor explains the largest proportion of the total variance. Next, a second set of weights can be selected so that the second factor accounts for most of the residual variance, subject to being uncorrelated with the first factor. This same principle could be applied to selecting additional weights for the additional factors.

Conducting factor analysis

The process to conduct factor analysis is as follows:

Formulate the problem

⇓⇓

Construct the correlation matrix

⇓⇓

Determine the method of factor analysis

⇓⇓

Determine the number of factors

⇓⇓

Rotate factors

⇓⇓

Interpret factors

⇓⇓⇓⇓

Calculate factor scores **or** Select surrogate variables

⇓⇓

Determine the model fit

Conducting factor analysis

Formulate the problem:

The objectives of factor analysis should be identified. The variables to be included in the factor analysis should be specified *based on past research, theory and/or judgement of the researcher*.

It is important that the variables be appropriately measured on an *interval or ratio scale*. An appropriate sample size should be used. As a rough guideline, there should be at least four or five times as many observations (i.e. sample size) as there are variables.

Construct the correlation matrix:

The analytical process is based on a *matrix of correlations* between the variables. Bartlett's test of sphericity can be used to test the null hypothesis that the variables are uncorrelated in the population. In other words, the *population correlation matrix is an identity matrix*. If this hypothesis cannot be rejected, then the *appropriateness of factor analysis* should be questioned. Small values of the KMO statistic indicate that the correlations between pairs of variables cannot be explained by other variables and that factor analysis may not be appropriate.

Determine the method of factor analysis:

In **principal components analysis**, the total variance in the data is considered. The diagonal of the correlation matrix consists of unities, and full variance is brought into the factor matrix. Principal components analysis is recommended when the primary concern is to determine the *minimum number of factors which will account for maximum variance* in the data for use in subsequent multivariate analysis. The factors are called **principal components**.

In **common factor analysis**, the factors are estimated based only on the common variance. *Communalities* are inserted in the diagonal of the correlation matrix. This method is appropriate when the primary concern is to *identify the underlying dimensions* and the common variance is of interest. This method is also known as **principal axis factoring**.

Determine the number of factors:

***A priori* determination -** Sometimes, because of prior knowledge, the researcher knows how many factors to expect and, therefore, can specify the number of factors to be extracted beforehand.

Determination based on eigenvalues - Here, only factors with eigenvalues greater than 1.0 are retained. An eigenvalue represents the amount of variance associated with the factor. Hence only factors with a variance greater than 1.0 are included. Factors with a variance less than 1.0 are no better than a single variable since, due to standardisation, each variable has a variance of 1.0. If the number of variables is less than 20, this approach will result in a conservative number of factors.

Determination based on the scree plot - A scree plot is a plot of the eigenvalues against the number of factors in the order of extraction. Experimental evidence indicates that the point just before the scree begins denotes the true number of factors.

Determination based on the percentage of variance - In this approach, the number of factors extracted is determined so that the cumulative percentage of variance extracted by the factors reaches a satisfactory level. It is recommended that the factors extracted should account for at least 60% of the variance.

Determination based on split-half reliability - The sample is split in half and factor analysis is performed on each half. Only factors with a high correspondence of factor loadings across the two subsamples are retained.

Determination based on significance tests - It is possible to determine the statistical significance of the separate eigenvalues and retain only those factors which are statistically significant. A drawback is that with large samples (sample sizes greater than 200), many factors are likely to be statistically significant, although from a practical viewpoint many of these account for only a small proportion of the total variance.

Rotate factors:

Although the initial or unrotated factor matrix indicates the relationship between the factors and the individual variables, it *seldom results in factors which can be interpreted*, because the factors are correlated with many variables. Therefore, through *rotation* (we ignore the technical details), the factor matrix is transformed into a simpler one which is easier to interpret.

In rotating the factors, we would like each factor to have non-zero, or significant, loadings or coefficients for only some of the variables. Likewise, we would like each variable to have non-zero or significant loadings with only a few factors, if possible with only one. The rotation is called **orthogonal rotation** if the axes are maintained at right angles.

The most commonly-used method for rotation is the **varimax procedure**. This is an orthogonal method of rotation which minimises the number of variables with high loadings on a factor, thereby enhancing the interpretability of the factors.Orthogonal rotation results in factors which are uncorrelated. The rotation is called **oblique rotation** when the axes are not maintained at right angles, and the factors are correlated. Sometimes, allowing for correlations among factors can simplify the factor pattern matrix. Oblique rotation should be used when factors in the population are likely to be strongly correlated.

Figure 24.5 of the textbook shows a factor matrix before and after rotation.

Interpret factors:

A factor can then be interpreted in terms of the variables which load highly on it. Another useful aid in interpretation is to plot the variables, *using the factor loadings as coordinates*. Variables at the end of an axis are those which have high loadings on only that factor and hence describe the factor.

Calculate factor scores:

The factor scores for the iith factor may be *estimated* as follows:

$$F_i = W_{i1}X_1 + W_{i2}X_2 + W_{i3}X_3 + \cdots + W_{ik}k$$

Factor scores can be used as independent or predictor variables in subsequent multivariate analysis, such as multiple regression or discriminant analysis.

Select surrogate variables:

By examining the factor matrix, one could select for each factor the variable with the highest loading on that factor which could then be used as a *surrogate variable* for the associated factor. However, the choice is not as easy if two or more variables have similarly high loadings. In such a case, the choice between these variables should be based on theoretical and/or measurement considerations.

Determine model fit:

The correlations between the variables can be deduced, or reproduced, from the estimated correlations between the variables and the factors. The differences between the observed correlations (as given in the input correlation matrix) and the reproduced correlations (as estimated from the factor matrix) can be examined to determine the model fit. These differences are called **residuals**.

Activity 17.4

Describe the factor analysis model.

Statistics associated with factor analysis

Bartlett's test of sphericity - A test statistic used to examine the hypothesis that the variables are uncorrelated in the population, i.e. the population correlation matrix is an identity matrix; each variable correlates perfectly with itself (ρ=1), but has no correlation with the other variables (ρ=0).

Correlation matrix - A correlation matrix is a lower triangular matrix showing the simple bivariate correlations, rr, between all possible pairs of variables included in the analysis. The diagonal elements, which are all 1, are usually omitted.

Communality - Communality is the amount of variance a variable shares with all the other variables being considered. This is also the proportion of variance explained by the common factors.

Eigenvalue - The eigenvalue represents the total variance explained by each factor.

Factor loadings - Factor loadings are simple correlations between the variables and the factors.

Factor loading plot - A factor loading plot is a plot of the original variables using the factor loadings as coordinates.

Factor matrix - A factor matrix contains the factor loadings of all the variables on all the factors extracted.

Factor scores - Factor scores are composite scores estimated for each participant on the derived factors.

Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy - An index used to examine the appropriateness of factor analysis. High values (between 0.5 and 1.0) indicate that factor analysis is appropriate. Values below 0.5 imply that factor analysis may not be appropriate.

Percentage of variance - The percentage of the total variance attributed to each factor.

Residuals - The differences between the observed correlations, as given in the input correlation matrix, and the reproduced correlations, as estimated from the factor matrix.

Scree plot - A scree plot is a plot of the eigenvalues against the number of factors in the order of extraction.

Activity 17.5

What hypothesis is examined by Bartlett's test of sphericity? For what purpose is this test used?

Activity 17.6
What is meant by the term 'communality of a variable'?

Activity 17.7
Briefly define the following: eigenvalue, factor loadings, factor matrix and factor scores.

Activity 17.8
For what purpose is the Kaiser–Meyer–Olkin measure of sampling adequacy used?

Activity 17.9
What is the major difference between principal components analysis and common factor analysis?

Activity 17.10
Explain how eigenvalues are used to determine the number of factors.

Activity 17.11
What is a scree plot? For what purpose is it used?

Activity 17.12
Why is it useful to rotate the factors? Which is the most common method of rotation?

Activity 17.13
What guidelines are available for interpreting the factors?

Activity 17.14
When is it useful to calculate factor scores?

Activity 17.15
What are surrogate variables? How are they determined?

Activity 17.16
How is the fit of the factor analysis model examined?

Factor analysis example
Suppose you want to determine the underlying benefits consumers seek from the purchase of toothpaste. A sample of 30 participants was interviewed using street interviewing.

Participants were asked to indicate their degree of agreement with the following statements using a 7-point scale (1 = strongly disagree, 7 = strongly agree).

- V_1 = It is important to buy toothpaste which prevents cavities.
- V_2 = I like toothpaste which gives shiny teeth.
- V_3 = Toothpaste should strengthen your gums.
- V_4 = I prefer toothpaste which freshens breath.
- V_5 = Prevention of tooth decay should be an important benefit offered by toothpaste.

- V_6 = The most important consideration in buying toothpaste is attractive teeth.

Table 24.1 of the textbook provides the toothpaste attribute ratings. The data can be downloaded from the file *Toothpaste.sav* or viewed in the table below.

Number	V1	V2	V3	V4	V5	V6
1	7	3	6	4	2	4
2	1	3	2	4	5	4
3	6	2	7	4	1	3
4	4	5	4	6	2	5
5	1	2	2	3	6	2
6	6	3	6	4	2	4
7	5	3	6	3	4	3
8	6	4	7	4	1	4
9	3	4	2	3	6	3
10	2	6	2	6	7	6
11	6	4	7	3	2	3
12	2	3	1	4	5	4
13	7	2	6	4	1	3
14	4	6	4	5	3	6
15	1	3	2	2	6	4
16	6	4	6	3	3	4
17	5	3	6	3	3	4
18	7	3	7	4	1	4
19	2	4	3	3	6	3
20	3	5	3	6	4	6
21	1	3	2	3	5	3
22	5	4	5	4	2	4
23	2	2	1	5	4	4
24	4	6	4	6	4	7
25	6	5	4	2	1	4
26	3	5	4	6	4	7
27	4	4	7	2	2	5
28	3	7	2	6	4	3
29	4	6	3	7	2	7
30	2	3	2	4	7	2

Table 24.2 of the textbook reports the correlation matrix for the data.
Table 24.3 of the textbook provides the results of principal components analysis for the data.
Figure 24.4 of the textbook shows the scree plot with the eigenvalues plotted in the order of extraction.
Figure 24.6 of the textbook shows the factor loading plot.

Discussion forum and activities

To access the solutions to these questions and case study, click here to access the printable Word document or click here to go to LSE's Elearning resources.

Activities on the block’s topics

1. Replicate the factor analysis results from the lecture example using the data file *Toothpaste.sav*. (An Excel version of the dataset is *Toothpaste.xlsx*.)

Video walkthrough of activity 1.

2. In a study of the relationship between household behaviour and shopping behaviour, data on the following lifestyle statements were obtained on a seven-point scale (1 = disagree, 7 = agree).

    - V1=V1= I would rather spend a quiet evening at home than go out to a party.
    - V2=V2= I always check prices, even on small items.
    - V3=V3= Magazines are more interesting than movies.
    - V4=V4= I would not buy products advertised on billboards.
    - V5=V5= I am a homebody.
    - V6=V6= I save and cash coupons.
    - V7=V7= Companies waste a lot of money advertising.

The data obtained from a pretest sample of 25 participants can be found in the file *Behaviour.sav*. (An Excel version of the dataset is *Behaviour.xlsx*.)

h. Analyse these data using principal components analysis, using the varimax rotation procedure.

i. Interpret the factors extracted.

j. Calculate factor scores for each participant.

k. If surrogate variables were to be selected, which ones would you select?

l. Examine the model fit.

m. Analyse the data using common factor analysis.

Video walkthrough of activity 2.

3. In a survey pre-test, data were obtained from 45 participants on Benetton clothes. These data are given in the file *Benetton.sav*, which gives the usage, gender, awareness, attitude, preference, intention and loyalty toward Benetton of a sample of Benetton users. Usage was coded as 1, 2 or 3, representing light, medium or heavy users, respectively. Gender was coded as 1 for females and 2 for males. Awareness, attitude, preference, intention and loyalty were measured on a 7-point Likert-type scale (1 = Very unfavourable, 7 = Very favourable). Note that five participants have missing values which are denoted by 9. (An Excel version of the dataset is *Benetton.xlsx*.)

Analyse the Benetton data using factor analysis. Consider only the following variables: awareness, attitude, preference, intention and loyalty toward Benetton.

[Video walkthrough of activity 3.]

4. Complete the following portion of an output from principal components analysis:

Variable	**Communality**	**Factor**	**Eigenvalue**	**% of variance**
V1V1	1.0	1	3.25	
V2V2	1.0	2	1.78	
V3V3	1.0	3	1.23	
V4V4	1.0	4	0.78	
V5V5	1.0	5	0.35	
V6V6	1.0	6	0.30	
V7V7	1.0	7	0.19	
V8V8	1.0	8	0.12	

a. Draw a scree plot based on this data.
b. How many factors should be extracted? Explain your reasoning.

Learning outcomes checklist

Use this to assess your own understanding of the chapter. You can always go back and amend the checklist when it comes to revision!

- Describe the concept of factor analysis and explain how it is different from analysis of variance, multiple regression and discriminant analysis
- Discuss the procedure for conducting factor analysis, including problem formulation, construction of the correlation matrix, selection of an appropriate method, determination of the number of factors, rotation and interpretation of factors
- Explain the selection of surrogate variables and their application, with emphasis on their use in subsequent analysis
- Describe the procedure for determining the fit of a factor analysis model using the observed and the reproduced correlations.

Block 17: Factor analysis

Solution to Activity 17.1

Structural equation modelling (SEM) enables the researcher to examine relationships between several continuous or discrete independent variables and several continuous or discrete dependent variables. The independent and dependent variables can be latent or measured variables. SEM estimates the unknown coefficients in a set of linear structural equations. Variables in the equation system are usually directly observed variables plus unmeasured latent variables which are not observed but relate to the observed variables. SEM assumes there is a causal structure among a set of latent variables, and that the observed variables are indicators of the latent variables. The latent variables may appear as linear combinations of observed variables, or they may be intervening variables in a causal chain. The application in market research occurs in areas where researchers cannot directly measure and, therefore, represent phenomena such as 'fashion consciousness'. Such phenomena are termed latent variables and represent a whole array of abstract concepts or theoretical constructs which cannot be measured directly but are vitally important to market researchers, for example 'stylish', 'prestigious', 'sexy' and 'edgy'. Such variables are often referred to as 'factors' or 'common factors'. That is, they are presumed to underlie what can be observed, in the sense that the latent variables directly influence the outcome or values taken by the observed variables.

Solution to Activity 17.2

In multiple regression and discriminant analysis, one variable is considered as the dependent or criterion variable and the other variables as independent or predictor variables. However, in factor analysis there is no such distinction. Here, the whole set of interdependent relationships among sets of many interrelated variables is examined and represented in terms of a few underlying factors.

Solution to Activity 17.3

The major uses of this technique are the following.

- Identifying underlying dimensions or factors, which explain the correlation among a set of interrelated variables.
- Identifying a new smaller set of uncorrelated variables to replace the original set of interrelated variables in subsequent multivariate analysis, such as regression or discriminant analysis.
- Identifying a smaller set of salient variables from a larger set for use in subsequent multivariate analysis.

Solution to Activity 17.4

In the factor analysis model, each variable is expressed as a linear combination of underlying factors. The covariation among the variables is described in terms of a few common factors and a unique factor for each variable. If the variables are standardised the factor model is represented as:

$$X_i + A_{i1}F_1 + A_{i2}F_2 + A_{i3}F_3 + \cdots + A_{im}F_m + V_iU_i$$

where:

- Xi = the *i*th standardised variable

- A_{ij}= the standardised multiple regression coefficient of variable ii on common factor jj
- F= a common factor
- V_i= the standardised regression coefficient of variable ii on unique factor ii
- U_i= the unique factor
- M = the number of common factors.

The factors in turn can be expressed as linear combinations of the observed variables:

$$F_i = W_{i1}X_1 + W_{i2}X_2 + W_{i3}X_3 + \cdots + W_{ik}X_k$$

where:

- F_i= the estimate of the iith factor
- W_i= the weight or factor score coefficient
- k = the number of variables.

Solution to Activity 17.5

It examines the hypothesis that the population correlation matrix is an identity matrix and is used to test the appropriateness of the factor model.

Solution to Activity 17.6

In factor analysis, the amount of variance a variable shares with all other variables included in the analysis is referred to as the communality. This is also the proportion of variance explained by the common factors.

Solution to Activity 17.7

The following are all statistics associated with factor analysis.

- *Eigenvalue*. The total variance explained by each factor is represented by the corresponding eigenvalue.
- *Factor loadings*. This statistic is basically a simple correlation between a variable and the factors.
- *Factor matrix*. This is a matrix containing the factor loadings of all the variables on all the factors extracted.
- *Factor scores*. This represents the composite scores estimated for each participant on the derived factors.

Solution to Activity 17.8

The index is used to test the appropriateness of factor analysis. Values of the Kaiser–Meyer–Olkin (KMO) statistic between 0.5 and 1.0 indicate appropriateness, whereas values less than 0.5 imply inappropriateness of the factor analysis model.

Solution to Activity 17.9

The major difference lies in the approach used to derive the weights or factor score coefficients. In principal components analysis, the diagonal of the correlation matrix consists of ones and, therefore, the full variance is considered in the factor matrix. It is used to determine the minimum number of factors accounting for the maximum variance in the data, which can then be used in subsequent multivariate analysis.

In common factor analysis, communalities are inserted in the diagonal of the correlation matrix and, therefore, the factors are estimated based only on the common variance. This method is appropriate when the primary concern is to identify the underlying dimensions and the common variance is of interest.

Solution to Activity 17.10

To summarise the information contained in the original variables, a few factors are usually extracted. One of the methods of determining these factors is on the basis of eigenvalues. In this approach, only the factors with eigenvalues greater than 1.0 are retained. As the eigenvalue basically represents the amount of variance associated with the factor, the factors which exhibit a larger amount of variance are more useful in summarising information than the set of single variables (which have variance 1.0 due to standardisation). This approach is generally used if the number of variables is less than 20.

Solution to Activity 17.11

A scree plot is a plot of the eigenvalues against the number of factors in the order of extraction. This plot displays a distinct break between the steep slope of factors with large eigenvalues and the gradual trailing off, called scree, associated with the rest of the factors. The shape of the plot is used to determine the number of factors which summarise the information contained in the original variables. Experimental results show that the point just before the scree begins denotes the optimal number of factors.

Solution to Activity 17.12

Rotation of factors enhances the interpretability of the factors by achieving the simplest structure. The coefficients of the factor matrix obtained from factor analysis are used to interpret the factors. However, if several factors have near-zero or insignificant coefficients for the same variable or if each variable has significant loadings with a large number of factors, then it becomes very difficult to interpret the factors.

The varimax procedure is the most common method of rotation used. This method enhances the interpretability of the factors by minimising the number of variables which have high loadings on a factor.

Solution to Activity 17.13

Interpretation of the factors is done on the basis of the rotated factor matrix. The variables which have large loadings on the same factor are identified and that factor is then interpreted in terms of those variables. A plot of the factor loadings can also help with interpretation. Variables at the end of an axis have high loadings on that factor, whereas variables near the origin have small loadings on all the factors.

Usually a factor which cannot be clearly defined in terms of the original variables is labelled as undefined or a general factor.

Solution to Activity 17.14

It is useful when one is using factor analysis to reduce the original set of variables to a smaller set of composite variables (factors) for use in subsequent multivariate analysis.

Solution to Activity 17.15

Surrogate variables act as representatives of the factors in subsequent analysis. By examining the factor matrix and selecting a variable with the highest, or near highest, loading for each factor, a surrogate variable may be selected. However, in complex cases, the choice may be based on theoretical and/or precise measurement considerations. Often, prior knowledge may also influence the judgement.

Solution to Activity 17.16

Model fit is assessed via an examination of the residuals – the differences between the observed correlations obtained from the input correlation matrix and the reproduced correlations estimated from the factor matrix. If many large residuals exist, then one can infer that the factor model does not provide a good fit to the data. This analysis is based on the implicit assumption that the observed correlation between the variables is due to the common factors. Therefore, the correlations between the variables can be reproduced from the estimated correlations between the variables and the factors.

Solutions to Activitys on the block’s topics

1. SPSS output from running factor analysis on the dataset is as follows.

Descriptive Statistics

	Mean	**Std. Deviation**	**Analysis N**
It is important to buy toothpaste that prevents cavities	3.93	1.982	30
I like toothpaste that gives shiny teeth	3.90	1.373	30
Toothpaste should strength your gums	4.10	2.057	30
I prefer toothpaste that freshens breath	4.10	1.373	30
Prevention of tooth decay should be an important benefit offered by toothpaste	3.50	1.907	30
The most important consideration in buying toothpaste is attractive teeth	4.17	1.392	30

The above descriptive statistics show average agreement ratings between 3.50 and 4.17 for the six statements, with responses for each variable exhibiting some variation. There are no missing values as there are 30 responses for each statement.

Correlation Matrix

		It is important to buy toothpaste that prevents cavities	I like toothpaste that gives shiny teeth	Toothpaste should strength your gums	I prefer toothpaste that freshens breath	Prevention of tooth decay should be an important benefit offered by toothpaste	The most important consideration in buying toothpaste is attractive teeth
Correlation	**It is important to buy toothpaste that prevents cavities**	1.000	-.053	.873	-.086	-.858	.004
	I like toothpaste that gives shiny teeth	-.053	1.000	-.155	.572	.020	.640
	Toothpaste should strength your gums	.873	-.155	1.000	-.248	-.778	-.018
	I prefer toothpaste that freshens breath	-.086	.572	-.248	1.000	-.007	.640
	Prevention of tooth decay should be an important benefit offered by toothpaste	-.858	.020	-.778	-.007	1.000	-.136
	The most important consideration in buying toothpaste is attractive teeth	.004	.640	-.018	.640	-.136	1.000
Sig. (1-tailed)	**It is important to buy toothpaste that prevents cavities**		-.390	.000	.325	.000	.491

	I like toothpaste that gives shiny teeth	.390		.207	.000	.459	.000
	Toothpaste should strength your gums	.000	.207		.093	.000	.462
	I prefer toothpaste that freshens breath	.325	.000	.093		.486	.000
	Prevention of tooth decay should be an important benefit offered by toothpaste	.000	.459	.000	.486		.236
	The most important consideration in buying toothpaste is attractive teeth	.491	.000	.462	.000	.236	

Looking at the correlation matrix, we observe some sample correlation coefficients which are significantly different from zero. Recall the test of $H_0: \rho = 0$ vs. $H_1: \rho \neq 0$uses the test statistic:

$$\frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}$$

The sample correlation coefficients are reported in the top half of the correlation matrix, with the pp-values for all the pairwise correlation tests in the bottom half. There are significant correlations between the following pairs of variables:

- V_1 and V_3
- V_1 and V_5
- V_2 and V_4
- V_2 and V_6
- V_3 and V_5
- V_4 and V_6 .

Therefore, it seems some of the variables are correlated suggesting factor analysis may be appropriate. However, this should be confirmed using the Kaiser-Meyer-Olkin statistic and/or Bartlett's test of sphericity.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.660
Bartlett's Test of Sphericity	Approx. Chi-Square	111.314
	df	15
	Sig.	.000

We see that the Kaiser-Meyer-Olkin statistic is 0.660 which is between 0.5 and 1.0, indicating that factor analysis is appropriate. This is confirmed when we look at Bartlett's test of sphericity which has a test statistic value of 111.314 which is from an approximate x_{15}^2 distribution. The *p*-value is 0.000, i.e. it is a highly significant test statistic value hence we reject the null hypothesis that in the population the correlation matrix is an identity matrix, that is:

$$H_0\colon \begin{pmatrix} \rho_{11} & \rho_{12} & \rho_{12} & \rho_{14} & \rho_{15} & \rho_{16} \\ \rho_{21} & \rho_{22} & \rho_{23} & \rho_{24} & \rho_{25} & \rho_{26} \\ \rho_{31} & \rho_{32} & \rho_{33} & \rho_{34} & \rho_{35} & \rho_{36} \\ \rho_{41} & \rho_{42} & \rho_{43} & \rho_{44} & \rho_{45} & \rho_{46} \\ \rho_{51} & \rho_{52} & \rho_{53} & \rho_{54} & \rho_{55} & \rho_{56} \\ \rho_{61} & \rho_{62} & \rho_{63} & \rho_{64} & \rho_{65} & \rho_{66} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Therefore, we proceed with factor analysis as we have strong evidence that there are some correlated variables among $V_1,\ V_2, \ldots, V_6$

Communalities

	Initial	Extract
It is important to buy toothpaste that prevents cavities	1.000	.926
I like toothpaste that gives shiny teeth	1.000	.723
Toothpaste should strength your gums	1.000	.894
I prefer toothpaste that freshens breath	1.000	.739
Prevention of tooth decay should be an important benefit offered by toothpaste	1.000	.878
The most important consideration in buying toothpaste is attractive teeth	1.000	.790

Extraction Method: Principal Component Analysis.

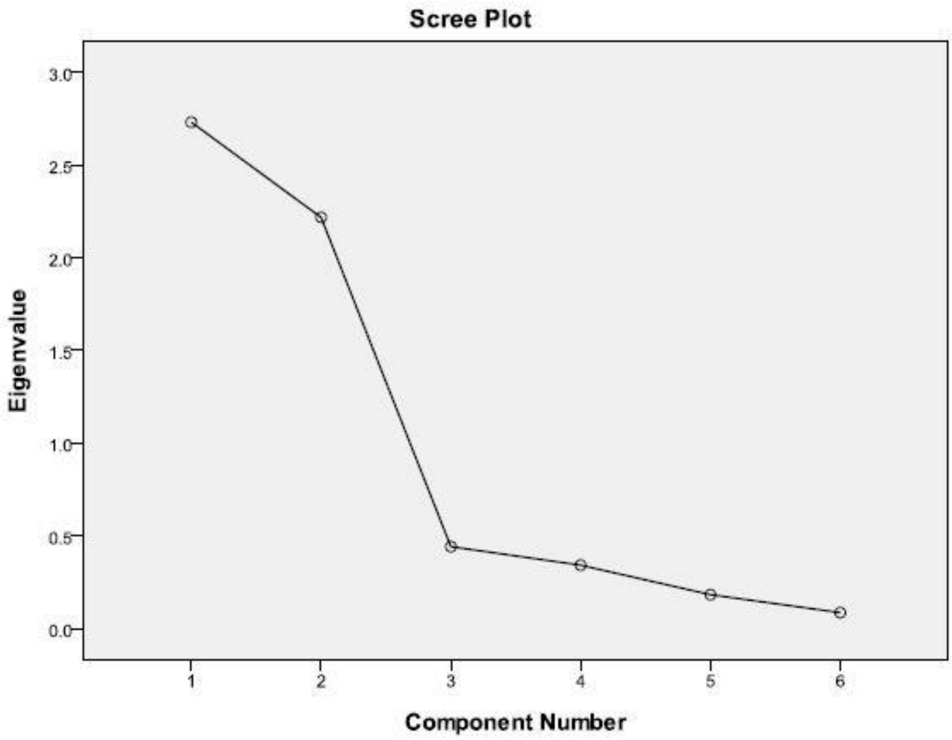
After standardising the variables, each variable then has a variance of 1.0. The communalities represent the amount of variance a variable shares with all the other variables being considered, as well as the proportion of variance explained by the (to be determined) common factors. We see that all the variables have 'high' communalities.

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	2.731	45.520	45.520	2.731	45.520	45.520	2.688	44.802	44.802
2	2.218	36.969	82.488	2.218	36.969	82.488	2.261	37.687	82.488
3	.442	7.360	89.848						
4	.341	5.688	95.536						
5	.183	3.044	98.580						
6	.085	1.420	100.000						

Extraction Method Principal Component Analysis.

With k variables, we can identify a maximum of k common factors (here $k = 6$). However, not all extracted factors are necessarily useful. To determine the number of factors we can look at a variety of methods (hopefully we decide on the same number of factors using each approach). First, we consider the eigenvalues for each factor - these are, in descending order, 2.731, 2.218, 0.442, 0.341, 0.183 and 0.085. Given the standardised variables have a variance of 1.0, only factors with eigenvalues greater than 1 should be used. Therefore, here we can extract two common factors. Second, we seek common factors which cumulatively account for a large percentage of the total variance (typically at least 60%). We see the first factor accounts for 45.520%, while the second accounts for 36.969%, so the first two factors combined account for a cumulative percentage of 45.520 + 36.969 = 82.488% of the total variance. Subsequent factors add little to this cumulative percentage so, again, two factors seem appropriate.



A scree plot is a plot of the eigenvalues against the number of factors in the order of extraction. The point just before the scree indicates the 'right' number of factors so, again, two factors seem reasonable.

Component Matrix[a]

	Component	
	1	**2**
It is important to buy toothpaste that prevents cavities	.928	.253
I like toothpaste that gives shiny teeth	-.301	.795
Toothpaste should strength your gums	.936	.131
I prefer toothpaste that freshens breath	-.342	.789
Prevention of tooth decay should be an important benefit offered by toothpaste	-.869	-.351
The most important consideration in buying toothpaste is attractive teeth	-.177	.871

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

SPSS has itself concluded that two factors are appropriate (remember in the 'Extraction……' option box we chose 'Based on Eigenvalue' for values greater than 1). The component (factor) matrix reports the factor loadings (which are simple correlations) between each variable and both factors (components). Although it *seems* that V_1, V_3, …, V_5 load highly on factor 1, and V_2, V_4, …, V_6 load highly on factor 2, we can rotate this matrix to ease the interpretability.

Rotated Component Matrix[a]

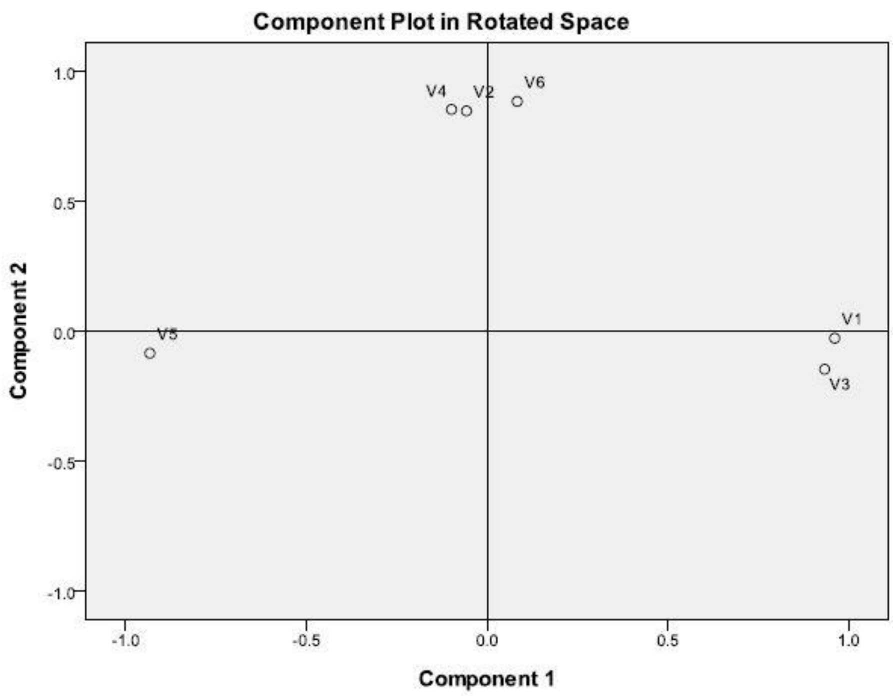
	Component	
	1	**2**
It is important to buy toothpaste that prevents cavities	.962	-.027
I like toothpaste that gives shiny teeth	-.057	.848
Toothpaste should strength your gums	.934	-.146
I prefer toothpaste that freshens breath	-.098	.854
Prevention of tooth decay should be an important benefit offered by toothpaste	-.933	-.084
The most important consideration in buying toothpaste is attractive teeth	.083	.885

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

Using the varimax rotation procedure to obtain the simplest structure of the factor matrix, we obtain factor loadings which more clearly load highly on one, and only one, common factor. Our previous suspicions that V_1, V_3, …, V_5

load highly on factor 1, and V_2, V_4, …, V_6 load highly on factor 2 are confirmed by this rotated factor matrix. Note there is much less ambiguity about which variables load highly on a factor using the rotated matrix.



The factor loading plot is a plot of the original variables using the factor loadings (based on the rotated factor matrix) as coordinates. Remember the factor loadings are just correlation coefficients, so the axes of the factor loading plot go from −1 to +1. This plot allows us to easily visualise which variables load highly onto which factors. Again, V_1, V_3 and V_5load highly on factor 1 (at the end of the horizontal axis), and V_2, V_4 and V_6load highly on factor 2 (at the end of the vertical axis). We can now interpret the two common factors. Recall the original statements to which survey participants had to give their level of agreement.

- V_1 = It is important to buy toothpaste which prevents cavities.
- V_2 = I like toothpaste which gives shiny teeth.
- V_3 = Toothpaste should strengthen your gums.
- V_4 = I prefer toothpaste which freshens breath.
- V_5 = Prevention of tooth decay should be an important benefit offered by toothpaste.
- V_6 = The most important consideration in buying toothpaste is attractive teeth.

It is clear that V_1, V_3 and V_5 are all related to dental health, so we might label factor 1 as 'health benefits', while V_2, V_4 and V_6 are all related to appearance, so we might label factor 2 as 'social benefits'.

Component Score Coefficient Matrix

	Component	
	1	**2**
It is important to buy toothpaste that prevents cavities	.358	.011
I like toothpaste that gives shiny teeth	-.001	.375
Toothpaste should strength your gums	.345	-.043
I prefer toothpaste that freshens breath	-.017	.377
Prevention of tooth decay should be an important benefit offered by toothpaste	-.350	-.059
The most important consideration in buying toothpaste is attractive teeth	.052	.395

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
Component Scores.

We are now in a position to calculate factor scores for each participant, i.e. the 'health benefits' score and 'social benefits' score. These are returned in SPSS as new columns in the original data matrix (check this!) and are calculated using:

$$F_i = W_{i1}X_1 + W_{i2}X_2 + W_{i3}X_3 + \cdots + W_{ik}X_k$$

where the weights are from the component score coefficient matrix. Note the X_is are standarised variables (not the original values in our data matrix). These two factor scores can now be used instead of the original six variables in subsequent multivariate analyses (such as multiple regression or discriminant analysis). By construction, since we used the varimax rotation procedure, these factor scores are uncorrelated and hence there is absolutely no chance of multicollinearity if these served as independent variables! Note we can check that these are uncorrelated by producing a correlation matrix of the factor scores.

Alternatively, if we would rather not use the factor scores (as they are a synthetic creation and not directly observable) we could instead use surrogate variables. A surrogate variable is one of the original variables which is used instead of a common factor. We choose the original variable which loads most highly with an extracted factor (remember a factor loading is just a correlation, so we seek the original variable which is most highly correlated with a factor). Looking at the rotated factor matrix, we see that V1V1 would be the surrogate variable for common factor 1 (a factor loading of 0.962), and V6V6 would be the surrogate variable for common factor 2 (a factor loading of 0.885).

Reproduced Correlations

		It is important to buy toothpaste that prevents cavities	**I like toothpaste that gives shiny teeth**	**Toothpaste should strength your gums**	**I prefer toothpaste that freshens breath**	**Prevention of tooth decay should be an important benefit offered by toothpaste**	**The most important consideration in buying toothpaste is attractive teeth**
Reproduced Correlation	**It is important to buy toothpaste that prevents cavities**	.926[a]	-.078	.902	-.177	-.895	.057
	I like toothpaste that gives shiny teeth	-.078	.723[a]	-.177	.730	-.018	.746
	Toothpaste should strength your gums	.902	-.177	.894[a]	-.217	-.859	-.051
	I prefer toothpaste that freshens breath	-.117	.730	-.217	.739[a]	.020	.748
	Prevention of tooth decay should be an important benefit offered by toothpaste	-.895	-.018	-.859	.020	.878[a]	-.152
	The most important consideration in buying toothpaste is	.057	.746	-.051	.748	-.152	.790[a]

	attractive teeth						
Residual[b]	**It is important to buy toothpaste that prevents cavities**		.024	-.029	.031	.038	-.052
	I like toothpaste that gives shiny teeth	.024		.022	-.158	.038	-.105
	Toothpaste should strength your gums	-.029	.022		-.031	.081	.033
	I prefer toothpaste that freshens breath	.031	-.158	-.031		-.027	-.107
	Prevention of tooth decay should be an important benefit offered by toothpaste	.038	.038	.081	-.027		.016
	The most important consideration in buying toothpaste is attractive teeth	-.052	-.105	.033	-.107	.016	

Extraction Method: Principal Component Analysis

a. Reproduced communalities
b. Residuals are computed between observed and reproduced correlations. There are 5 (33.0%) nonredundant residuals with absolute values greater than 0.05.

Finally, we assess the model fit by examining the residuals - the differences between the observed correlations (the input correlation matrix) and the reproduced correlations (estimated from the factor matrix). As a guide we would treat residuals with absolute values less than 0.05 as ‘small’, and anything greater than 0.05 as signalling a less-than-ideal fit. SPSS helpfully reports the percentage of residuals with absolute values *greater* than 0.05, which here is 33.0%, which is 3 out of 15. The output above displays the reproduced correlations in the top part and the residuals in the bottom part (note we already examined the input correlations earlier).

2. Selected SPSS output using principal components analysis follows.

Correlation Matrix

	I would rather spend a quiet evening at home than	**I always check prices, even on small items**	**Magazines are more interesting than movies**	**I would not buy products advertised on billboards**	**I am a homebody**	**I save and cash coupons**	**Companies waste a lot of money advertising**

		go out to a party						
Correlation	I would rather spend a quiet evening at home than go out to a party	1.000	-.004	.628	.082	.675	-.100	-.338
	I always check prices, even on small items	-.004	1.000	.151	-.248	.048	.582	-.251
	Magazines are more interesting than movies	.628	.151	1.000	-.182	.480	.090	-.588
	I would not buy products advertised on billboards	.082	-.248	-.182	1.000	.272	.017	.469
	I am a homebody	.675	.048	.480	.272	1.000	-.110	-.082
	I save and cash coupons	-.100	.582	.090	.017	-.110	1.000	.014
	Companies waste a lot of money advertising	-.338	-.251	-.588	.469	-.082	.014	1.000
Sig. (1-tailed)	I would rather spend a quiet evening at home than go out to a party		.493	.000	.348	.000	.316	.049
	I always check prices, even on small items	.493		.236	.116	.409	.001	.113
	Magazines are more interesting than movies	.000	.236		.192	.008	.334	.001
	I would not buy products	.348	.116	.192		.094	.469	.009

	advertised on billboards							
	I am a homebody	.000	.409	.008	.094		.301	.348
	I save and cash coupons	.316	.001	.334	.469	.301		.473
	Companies waste a lot of money advertising	.049	.113	.001	.009	.348	.473	

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.550
Bartlett's Test of Sphericity	**Approx. Chi-Square**	57.994
	df	21
	Sig.	.000

Communalities

	Initial	**Extraction**
I would rather spend a quiet evening at home than go out to a party	1.000	.818
I always check prices, even on small items	1.000	.796
Magazines are more interesting than movies	1.000	.790
I would not buy products advertised on billboards	1.000	.800
I am a homebody	1.000	.805
I save and cash coupons	1.000	.841
Companies waste a lot of money advertising	1.000	.796

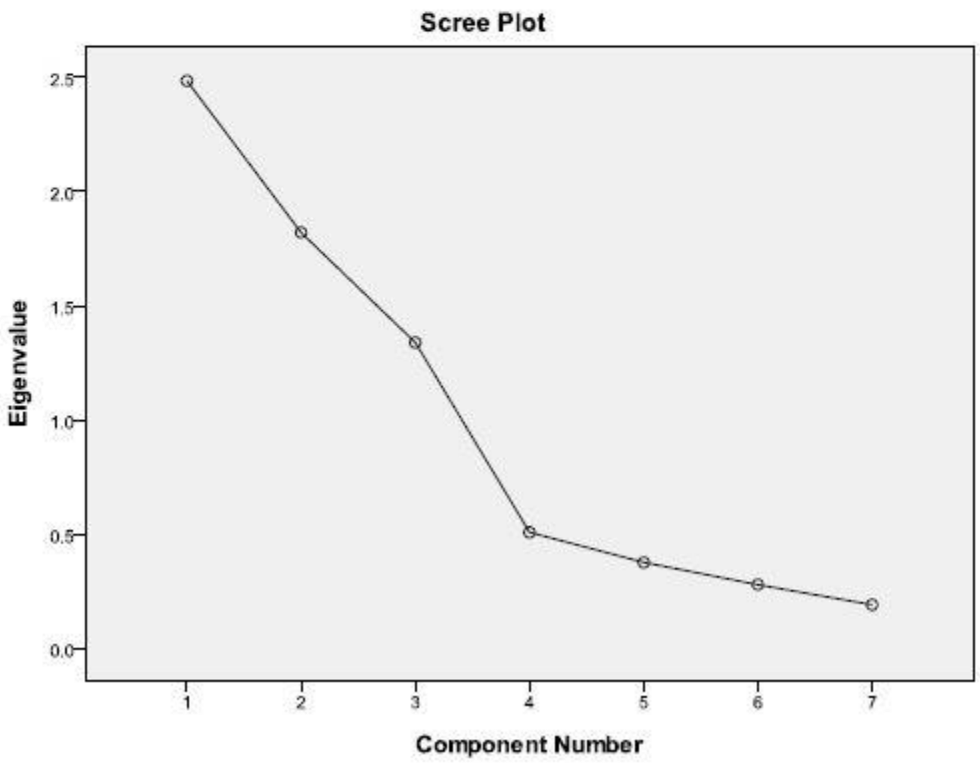
Extraction Method: Principal Component Analysis

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	2.485	35.505	35.505	2.485	35.505	35.505	2.315	33.076	33.076
2	1.821	26.013	61.518	1.821	26.013	61.518	1.731	24.729	57.805
3	1.339	19.131	80.649	1.339	19.131	80.649	1.599	22.844	80.649
4	.508	7.258	87.907						
5	.376	5.373	93.280						
6	.279	3.990	97.270						
7	.191	2.730	100.000						

Extraction Method Principal Component Analysis.



Component Matrix[a]

	Component		
	1	**2**	**3**
I would rather spend a quiet evening at home than go out to a party	.817	.378	.087
I always check prices, even on small items	.279	-.714	.457

Magazines are more interesting than movies	.887	-.027	-.043
I would not buy products advertised on billboards	-.204	.634	.597
I am a homebody	.664	.505	.329
I save and cash coupons	.050	-.604	.689
Companies waste a lot of money advertising	-.684	.383	.426

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

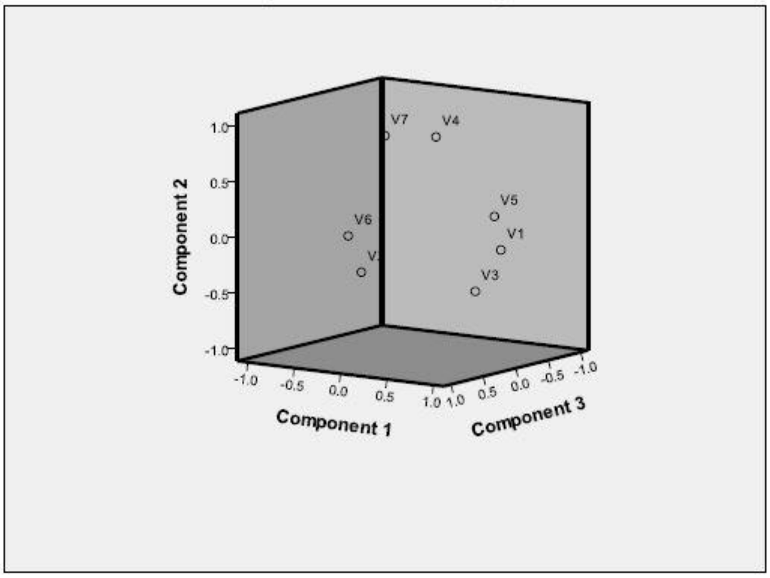
Rotated Component Matrix[a]

	Component		
	1	**2**	**3**
I would rather spend a quiet evening at home than go out to a party	.897	-.082	-.076
I always check prices, even on small items	.049	-.232	.860
Magazines are more interesting than movies	.762	-.440	.125
I would not buy products advertised on billboards	.214	.867	-.052
I am a homebody	.868	.224	-.017
I save and cash coupons	-.057	.091	.911
Companies waste a lot of money advertising	-.351	.817	-.073

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 4 iterations.

Component Plot in Rotated Space



Component Score Coefficient Matrix

	Component		
	1	**2**	**3**
I would rather spend a quiet evening at home than go out to a party	.391	.020	-.048
I always check prices, even on small items	.010	-.048	.530
Magazines are more interesting than movies	.302	-.189	.043
I would not buy products advertised on billboards	.170	.542	.059
I am a homebody	.405	.211	.021
I save and cash coupons	-.008	.145	.595
Companies waste a lot of money advertising	-.086	.461	.035

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
Component Scores.

Reproduced Correlations

	I would rather spend a quiet	I always check prices, even on	Magazines are more interesting than movies	I would not buy products advertised	I am a homebody	I save and cash coupons	Companies waste a lot of money advertising

		evening at home than go out to a party	small items		on billboards			
Reproduced Correlation	I would rather spend a quiet evening at home than go out to a party	.818[a]	-.002	.711	.125	.762	-.127	-.377
	I always check prices, even on small items	-.002	.796[a]	.247	-.236	-.025	.760	-.269
	Magazines are more interesting than movies	.711	.247	.790[a]	-.224	.561	.031	-.636
	I would not buy products advertised on billboards	.125	-.236	-.224	.800[a]	.381	.019	.637
	I am a homebody	.762	-.025	.561	.381	.805[a]	-.045	-.121
	I save and cash coupons	-.127	.760	.031	.019	-.045	.841[a]	.028
	Companies waste a lot of money advertising	-.377	-.269	-.636	.637	-.121	.028	.796[a]
Residual[b]	I would rather spend a quiet evening at home than go out to a party		-.001	-.083	-.043	-.087	.027	.040
	I always check prices, even on small items	-.001		-.096	-.012	.073	-.177	.018
	Magazines are more interesting than movies	-.083	-.096		.042	-.081	.060	.048
	I would not buy products advertised on billboards	-.043	-.012	.042		-.110	-.002	-.167
	I am a homebody	-.087	.073	-.081	-.110		-.065	.038
	I save and cash coupons	.027	-.177	.060	-.002	-.065		-.013
	Companies waste a lot of money advertising	.040	.018	.048	-.167	.038	-.013	

Extraction Method: Principal Component Analysis.

a. Reproduced communalities.
b. Residuals are computed between observed and reproduced correlations.There are 10 (47.0%) nonredundant residuals with absolute values greater than 0.05.

a. The value of the KMO statistic is 0.550 and hence is greater than 0.5, and Bartlett's test for sphericity is highly significant (pp-value is 0.000) indicating that factor analysis is appropriate. Three factors are extracted, accounting for 80.649% of the variance. These results are found from the 'KMO and Bartlett's Test' and 'Total Variance Explained' boxes.
b. The rotated factors should be interpreted based on factor loadings (found in the 'Rotated Component Matrix' box), or the factor loading plot. Note that since three factors are extracted, the factor loading plot is in three dimensions with coordinates for each of the three factors.

Component	**Variable**	**Loading**
Factor 1:	V_1 Spend evening at home	0.897
	V_3 Magazines over movies	0.762
	V_5 Homebody	0.868
Factor 2:	V_4 Won't buy products on billboards	0.867
	V_7 Companies waste money on advertising	0.817
Factor 3:	V_2 Always check prices	0.860
	V_6 Save and cash coupons	0.911

Factor 1 - Quiet homebody; Factor 2 - Anti-advertising; Factor 3 - Price conscious.

c. Factor scores can be requested using a standard option in the software (see the instructions) and appear as new columns in the data matrix. These can then be used in other multivariate analysis procedures, for example as explanatory variables in regression analysis.
d. Surrogate variables (the variable with the highest loading for each factor) are as follows.
   - Factor 1 - V_1.
   - Factor 2 - V_4.
   - Factor 3 - V_6.
e. Looking at the ‘Reproduced Correlations’ box, there are 10 residuals (47%) greater than 0.05. Therefore, the model fit is poor. However, note the small sample size.
f. Repeating the procedure with common factor analysis (select ‘Principal axis factoring’ in the ‘Extraction’ option window) we obtain the following (SPSS output omitted here).

Factor	**Variable**	**Loading**
Factor 1:	V_1 Spend evening at home	0.853
	V_3 Magazines over movies	0.703
	V_5 Homebody	0.807
Factor 2:	V_4 Won't buy products on billboards	0.680

	V_7 Companies waste money on advertising	0.791
Factor 3:	V_2 Always check prices	0.697
	V_6 Save and cash coupons	0.858

Factor 1 - Quiet homebody; Factor 2 - Anti-advertising; Factor 3 - Price conscious.

Surrogate variables are as follows.

- Factor 1 - V_1.
- Factor 2 - V_7.
- Factor 3 - V_6.

There are four residuals (19%) greater than 0.05. Therefore, the model fit is adequate, but not excellent.

3. Selected SPSS output using principal components analysis follows.

Correlation Matrix

		Awareness	Attitude	Preference	Intention	Loyalty
Correlation	**Awareness**	1.000	.799	.598	-.030	.014
	Attitude	.799	1.000	.594	.042	.053
	Preference	.598	.594	1.000	.189	-.195
	Intention	-.030	.042	.189	1.000	.765
	Loyalty	.014	.053	.195	.765	1.000
Sig. (1-tailed)	**Awareness**		.000	.000	.428	.467
	Attitude	.000		.000	.398	.372
	Preference	.000	.000		.122	.114
	Intention	.428	.398	.122		.000
	Loyalty	.467	.372	.114	.000	

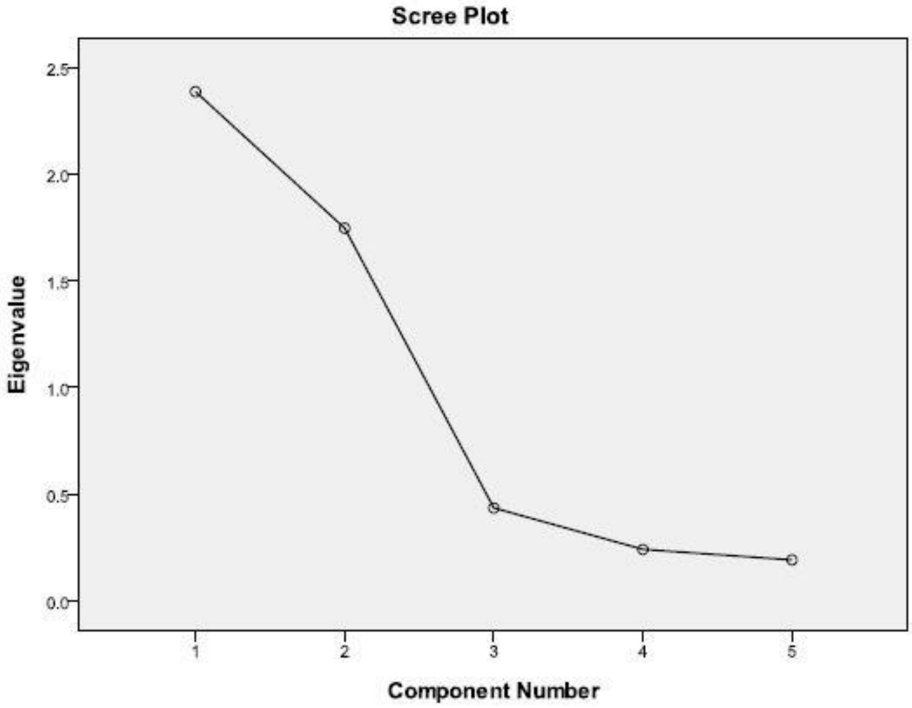
KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.626
Bartlett's Test of Sphericity	Approx. Chi-Square	90.478
	df	10
	Sig.	.000

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	2.386	47.716	47.716	2.386	47.716	47.716	2.331	46.627	46.627
2	1.746	34.923	82.639	1.746	34.923	82.639	1.801	36.012	82.639
3	.435	8.703	91.342						
4	.241	4.818	96.161						
5	.192	3.839	100.000						

Extraction Method: Principal Component Analysis.



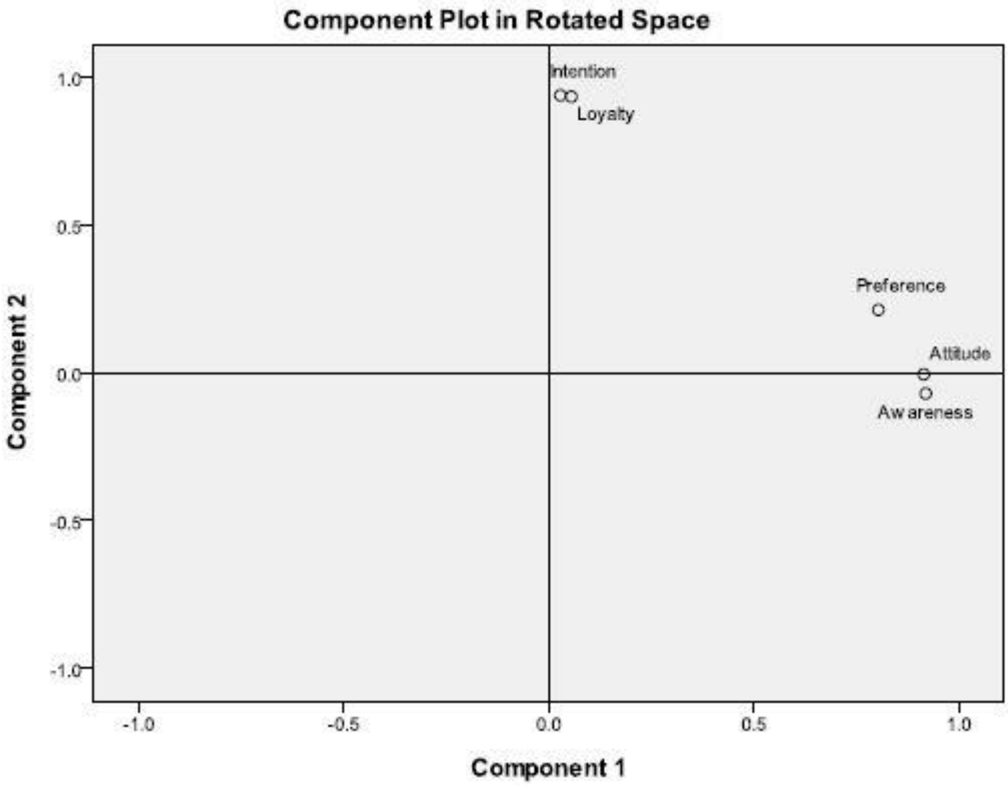
Rotated Component Matrix[a]

	Component	
	1	**2**
Awareness	.919	-.071
Attitude	.914	-.006
Preference	.804	.212

Intention	.028	.938
Loyalty	.055	.933

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.



Reproduced Correlations

		Awareness	Attitude	Preference	Intention	Loyalty
Reproduced Correlation	**Awareness**	.850[a]	.841	.724	-.040	-.016
	Attitude	.841	.836[a]	.734	.021	.045
	Preference	.724	.734	.692[a]	.222	.243
	Intention	-.040	.021	.222	.880[a]	.877
	Loyalty	-.016	.045	.243	.877	.874[a]
Residual[b]	**Awareness**		-.042	-.126	.011	.029

	Attitude	-.042		-.140	.021	.008
	Preference	-.126	-.140		-.033	-.047
	Intention	.011	.021	-.033		-.112
	Loyalty	.029	.008	-.047	-.112	

Extraction Method: Principal Component Analysis.
a. Reproduced communalities
b. Residuals are computed between observed and reproduced correlations. There are 3 (30.0%) nonredundant residuals with absolute values greater than 0.05.

The value of the KMO statistic is 0.626 which is greater than 0.5, and Bartlett's test for sphericity is highly significant indicating that factor analysis is appropriate. Two factors are extracted, accounting for 82.639% of the variance.

The rotated factors should be interpreted based on factor loadings or the factor loading plot.

Component	**Variable**	**Loading**
Factor 1:	V_1 Awareness of Benetton	0.919
	V_2 Attitude towards Benetton	0.914
	V_3 Preference for Benetton	0.804
Factor 2:	V_4 Purchase intention for Benetton	0.938
	V_5 Loyalty for Benetton	0.933

Factor 1 - Perception of Benetton; Factor 2 - Commitment to Benetton.

Surrogate variables are as follows.

1. Factor 1 - Either awareness or attitude (both have very similar factor loadings).
2. Factor 2 - Either purchase intention or loyalty for Benetton (again, both have very similar factor loadings).

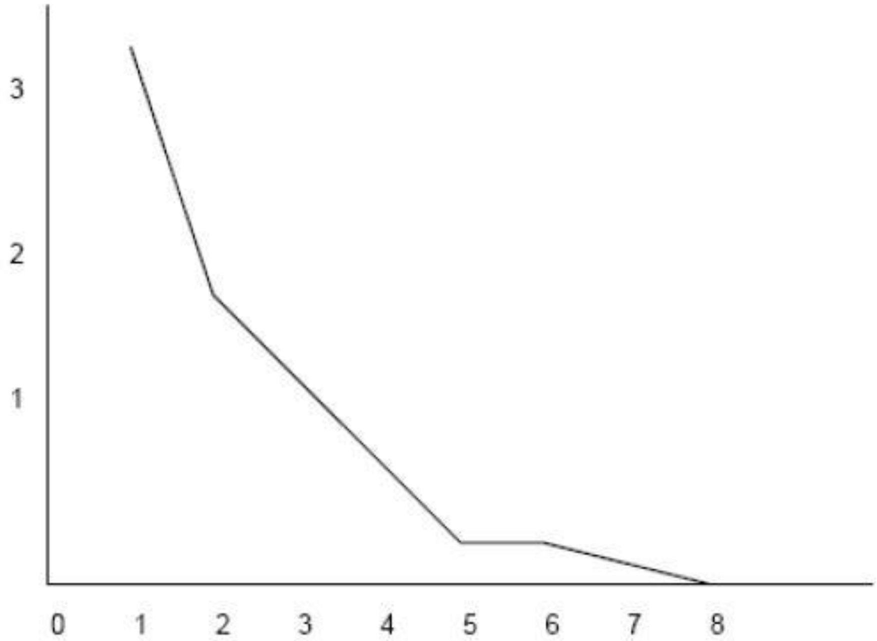
There are three residuals (30%) greater than 0.05. Therefore, the model fit is not great.

4. The percentage of variance is calculated as the eigenvalue divided by the sum of all eigenvalues.

Variable	**Communality**	**Factor**	**Eigenvalue**	**% of variance**
V_1	1.0	1	3.25	40.625
V_2	1.0	2	1.78	22.250
V_3	1.0	3	1.23	15.375
V_4	1.0	4	0.78	9.750

V_5	1.0	5	0.35	4.375
V_6	1.0	6	0.30	3.750
V_7	1.0	7	0.19	2.375
V_8	1.0	8	0.12	1.500

a. The scree plot is:



. Based on the scree plot, three or four factors should be considered because it is at V_5 that the graph flattens. Based on the eigenvalues, only three factors should be extracted since the percentage of variance accounted for by V_5 is too small to warrant consideration.