Chapter 9
Exploratory Factor and Principal Component Analysis

Chapter Overview

This chapter provides an introduction to Factor Analysis (FA): A procedure to define the underlying structure among the variables in the analysis. The chapter provides general requirements, statistical assumptions and conceptual assumptions behind FA. This chapter explains the way to do FA with IBM SPSS 20.0. It shows how to determine the number of factors to retain, interpret the rotated solution, create factor scores and summarize the results. Fictitious data from two studies are analysed to illustrate these procedures. The present chapter deals only with the creation of orthogonal (uncorrelated) components.

9.1 What is Factor Analysis

According to Hair et al. (2010),[1] '*factor analysis is an interdependence technique whose primary purpose is to define the underlying structure among the variables in the analysis*'. Suppose a marketing researcher wants to identify the underlying dimensions of retail brand attractiveness. He begins by administering the retail brand attractiveness scale from the existing literature to a large sample of people ($N = 2000$) during their visit in a particular retail store. Assume that there are five different dimensions, which consist of 30 different items. What the researcher will end up with these 30 different observed variables, the mass number as such will say very little about the underlying dimension of this retail attractiveness. On average, some of the scores will be high, some will be low and some intermediate, but interpretation of these scores will be extremely difficult if not impossible. This is where the tool factor analysis (FA) comes in handy and it allows the researcher in 'data reduction' and 'data summarization' of this large pool of items to a few representative factors or dimensions, which could be used for further multivariate

[1] See Ref. Hair et al. (2010).

S. Sreejesh et al., *Business Research Methods*,
DOI: 10.1007/978-3-319-00539-3_9,

© Springer International Publishing Switzerland 2014

statistical analysis. The general purpose of FA is the orderly simplification of a large number of intercorrelated measures or condense the information contained in a number of original variables into a few representative constructs or factors with minimal loss of information. The application of FA is based on some of the following conditions: general requirement, statistical assumptions and conceptual assumptions (See Table 9.1).

FA is used in the following circumstances:

1. To identify underlying dimensions, or factors, that explains the correlations among a set of variables. For example, a set of personality trait statements may be used to measure the personality dimensions of people. These statements may then be factor analysed to identify the underlying dimensions of personality trait or factors.
2. To identify a new, smaller, set of uncorrelated variables to replace the original set of correlated variables in subsequent multivariate analysis (regression or discriminant analysis). For example, the psychographic factors identified may be used as independent variables in explaining the differences between loyal and non-loyal consumers.
3. To identify a smaller set of salient variables from a larger set for use in subsequent multivariate analysis. For example, a few of the original lifestyle

Table 9.1 Conditions for doing factor analysis

General Requirements
1. Type of scale: Observed variables should be measured in either interval or ration scales, or at least at the ordinary level
2. Number of Items: If the researcher has prior knowledge about the underlying factor structure and want to test the dimensionality, then at least five or more variables should be included to represent each factor structure
3. Sample size: The rule of thumb for sample size is to have at least five times as many cases as variables entered into factor analysis +10
Statistical Assumptions
1. Random sampling: Each participant will contribute one response for each observed variable. These sets of scores should represent a random sample drawn from the population of interest
2. Linearity: The relationship between all observed variables should be linear
3. Bivariate Normal Distribution: Each pair of observed variables should display a bivariate normal distribution (e.g. they should form an elliptical scattergram when plotted)
Conceptual Assumptions
1. Variable Selection: Factor analysis is based on the basic assumption that there exists an underlying structure for the selected set of variables. The presence of high correlation and subsequent interpretation of do not guarantee relevance, even if it meets statistical assumptions. Therefore, it is the responsibility of the researcher to select the set of variables or items that are conceptually valid and appropriate to represent the underlying dimension
2. Sample Homogeneity: Another important conceptual assumption with regard to the factor analysis is that the selected sample should be homogeneous with respect to the underlying factor structure. It is inappropriate to do factor analysis for a set of items once the researcher knows a priori that the sample of male and female is different because of gender. The ignorance of this heterogeneity, and subsequent mixing of two groups (males and females would results in getting a correlation matrix and factor structure, that will be a poor representation of the unique structure of each group

statements that correlate highly with the identified factors may be used as independent variables to explain the differences between the loyal and non-loyal users.

9.2 Factor Analysis Versus Principal Component Analysis

Most of us use both PCA and FA interchangeably and often confused with the usage. This is quite obvious because there are some important similarities between the two methods. Much of the literature in this field do not differentiated these two tools. The similarity is that both methods used generally to determine to identify groups of observed variables that tend to hang together empirically. However, the two methods are different in their goals and in their underlying models. Roughly speaking, you should use PCA when you simply need to summarize or approximate your data using fewer dimensions (to visualize it), and you should use FA when you need an explanatory model for the correlations among your data. Perhaps the most important thing that deals with the differentiating aspect is its assumption of an underlying causal structure. FA is based on the assumption that covariance in the observed variable is due to the presence of one or more latent factors. In short, any change in the observed variable is due to the influence of its latent factor or latent variable is the cause of observed variables. In FA, the researcher believes that certain latent factors exist that exert causal influence on the observed variables they are studying. Exploratory FA helps the researcher to identify the number and nature of such latent factor. An example of such causal structure is shown in Fig. 9.1.

In PCA, the researcher will not make any assumption about an underlying causal structure. PCA is simply a variable reduction procedure that (typically)

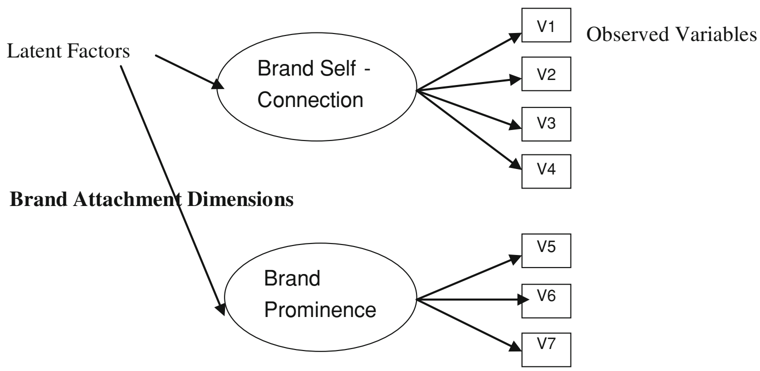


Fig. 9.1 An example of causal structure that is assumed in factor analysis

results in a relatively small number of components that account for most of the variance in a set of observed variables (i.e. groupings of observed variables vs. latent constructs).

http://www.youtube.com/watch?v=znrfcV4ZvlQ

http://www.youtube.com/watch?NR=1&feature=endscreen&v=pmMPXUtkQCc

9.3 A Hypothetical Example of Factor Analysis

Consider a construction company (ABC Builders) that is planning to build an apartment complex near a university. Suppose the company is interested in identifying the perception of the university faculties regarding the value consciousness,[2] price consciousness[3] and purchase decision involvement (PDI),[4] and sale proneness[5] that would help the company to construct and design apartments according to the target people. The company hired a marketing research firm to determine how these four major response dimensions the faculties concerned about before buying an apartment. The research firm identified the scale to measure these three constructs using some existing literature in price consciousness (3-item 5-point Likert scale), value consciousness (3-item 5-point Likert scale), sale proneness (2-item 5-point Likert scale) and PDI (4-item measure on 5-point bipolar phrases). The firm surveyed 98 randomly selected respondents (faculties), using different scales to determine the perception of the faculties with regard to the PDI, price, value consciousness and sale proneness. Table 9.2 shows the selected items for measuring the above-mentioned constructs.

Variables were labelled as V_1 through V_{12}. The data were analysed using PCA and the varimax rotation procedure. Let us begin with how this can be achieved by performing FA. Given below are the steps of SPSS operations that are adopted for performing FA.

[2] Value consciousness is defined as a concern for price paid relative to quality received (Lichtenstein et al. 1993).

[3] Price consciousness is the degree to which the consumer focuses exclusively on paying low prices (Lichtenstein et al. 1993).

[4] Purchase decision involvement (PDI) is defined as the extent of interest and concern that a consumer brings to bear upon a purchase-decision task (Mittal 1989). This scale considers purchase decision task as its goal object, and is considered mindset-not response behaviour.

[5] Sale proneness is defined as an increased propensity to respond to a purchase offer because the sale form in which the price is presented positively affects purchase evaluations (Lichtenstein et al. 1990).

Table 9.2 Items for measuring PDI, price and value consciousness and sale proneness

Sl. number	Variables
(V_1)	Selecting from many types and brands of this product (apartments) available in the market, would you say that I would not care at all as to which one I buy 1 2 3 4 5 I would care a great deal as to which one I buy
(V_2)	Do you think that the various types and brands of this product available in the market are all very alike or are all very different They are alike 1 2 3 4 5 They are all different
(V_3)	How important would it be to you to make a right choice of this product from ABC? Not at all important 1 2 3 4 5 Extremely important
(V_4)	In making your selection of this product from ABC, how concerned would you be the outcome of your choice? Not at all concerned 1 2 3 4 5 Very much concerned
(V_5)	I am very concerned about low prices, but I am equally concerned about product quality
(V_6)	When purchasing a product, I always try to maximize the quality I get for the money I spend
(V_7)	When I buy products, I like to be sure that I am getting my money's worth
(V_8)	I am not willing to go to extra effort to find lower prices
(V_9)	The money saved by finding lower prices is usually not worth the time and effort
(V_{10})	The time it takes to find low prices is usually not worth the effort
(V_{11})	When I buy Brand that's on sales, I feel that I am getting a good deal
(V_{12})	Compared with most people, I am more likely buy brand that are on special

9.4 SPSS Procedures for Performing Factor Analysis on PDI, Price and Value Consciousness and Sale Proneness Data in Windows

Step 1 Create a data file with these 12 items either in Excel (APARTMENT.xlsx) or in SPSS directly (APARTMENT_PROF.SREEJESH.sav). For the sake of convenience, let us code each of these items in Table 9.3 as V_1, V_2, … V_{12}. Once you have entered the data consist of all these 98 faculties, your SPSS **Data view** screen would like the one shown in Fig. 9.2.

Step 2 **Analyse** => **Dimension Reduction** => **Factor** to get Fig. 9.3.

Step 3 Once you click the option **Factor** in the second step, you would see a window of **FA** with all the variables that are listed in the SPSS variable view in the left panel (and there would a blank space in the right side **Fig.** 9.4).

Step 4 In Step 4, select all the 12 variables by clicking on them and move these variables to the right-side panel under **Variables** window (Fig. 9.5).

Step 5 Click on **Descriptives** to produce Fig. 9.4. Then click on the following: **(1) Initial solution** (under **Statistics), (2) Coefficients, (3) Determinant, (4) KMO and Bartlett's test of sphericity** (under **Correlation Matrix).** The detailed descriptions about these components are discussed below (Fig. 9.6).

Table 9.3 Example: PDI, price and value consciousness and sale proneness data for a sample of 98 university faculties

Res. number	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}	V_{11}	V_{12}
1	4	3	3	3	4	4	3	3	4	2	4	4
2	3	4	3	3	4	4	4	4	4	3	3	4
3	2	2	2	3	4	4	5	2	2	2	2	1
4	4	3	4	4	4	3	4	4	4	4	3	5
5	4	5	5	4	5	4	4	3	3	4	5	4
6	3	2	3	2	4	3	4	4	3	3	4	4
7	4	4	3	4	3	3	4	4	5	4	4	4
8	5	5	5	5	5	5	4	5	5	4	5	5
9	5	4	4	4	4	4	5	4	5	4	5	5
10	5	4	4	5	4	5	5	4	4	3	5	4
11	4	4	3	4	3	2	3	2	2	3	5	5
12	4	4	5	4	4	5	5	5	4	5	4	4
13	3	2	3	2	3	4	3	2	2	2	5	5
14	4	3	3	3	3	3	4	3	4	3	4	3
15	3	3	3	3	4	4	4	4	3	2	4	4
16	5	4	4	4	3	3	3	5	5	4	4	3
17	3	4	4	4	4	4	5	4	5	3	4	4
18	3	3	2	4	3	4	4	3	5	2	3	4
19	3	4	4	4	4	4	4	4	4	3	4	4
20	3	4	4	4	5	4	4	5	5	4	5	4
21	3	4	3	3	4	3	4	4	4	3	4	5
22	2	3	3	3	3	3	3	4	3	3	4	4
23	3	3	3	2	3	3	3	5	5	4	3	3
24	3	4	4	5	4	4	3	3	3	3	5	4
25	5	5	5	3	5	5	5	2	3	2	4	4
26	5	5	5	5	4	4	4	5	4	4	4	5
27	5	5	4	5	2	4	3	5	3	3	5	5
28	2	4	4	4	3	3	2	3	2	3	3	3
29	3	3	3	3	3	3	3	3	2	4	4	4
30	3	5	5	4	4	3	3	5	4	5	5	5
31	3	2	2	2	3	3	2	3	3	3	4	4
32	2	3	2	3	4	4	3	4	4	4	4	4
33	4	3	4	4	3	3	2	3	3	4	5	4
34	2	2	2	2	3	4	3	3	4	4	4	3
35	4	4	3	4	4	4	4	5	3	3	5	3
36	4	4	4	4	4	3	3	5	4	4	4	4
37	4	3	2	3	2	2	3	3	4	3	2	2
38	4	5	3	4	4	3	3	3	4	3	4	3
39	5	4	3	2	4	4	3	5	5	4	4	3
40	3	3	3	3	3	2	2	3	2	3	4	3
41	4	3	3	4	3	2	2	2	2	3	3	4
42	4	3	2	3	3	3	2	3	3	2	3	4
43	5	4	5	4	4	3	3	4	4	5	5	4
44	5	4	4	5	5	4	3	4	4	4	4	5

(continued)

Table 9.3 (continued)

Res. number	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}	V_{11}	V_{12}
45	5	4	3	4	4	4	3	4	4	4	3	4
46	5	5	4	4	4	4	3	4	3	5	4	4
47	4	5	5	5	5	5	4	4	4	4	4	4
48	3	4	3	4	3	4	3	4	3	2	3	3
49	3	4	3	3	4	3	3	3	2	3	3	5
50	3	3	4	5	4	4	4	5	5	4	3	4
51	4	5	4	5	4	4	3	5	2	3	4	4
52	3	4	3	4	4	3	3	3	4	4	5	5
53	5	4	4	5	4	5	3	4	5	3	5	4
54	5	5	5	5	4	5	4	4	4	3	5	5
55	3	5	5	5	3	2	2	3	4	2	5	3
56	5	4	2	4	4	3	3	4	5	4	4	4
57	3	4	3	4	3	3	3	3	4	4	3	4
58	3	4	5	5	4	3	2	4	4	2	4	3
59	4	5	5	4	3	4	3	4	5	3	3	3
60	4	3	3	3	4	3	3	4	3	4	4	4
61	5	5	5	5	4	3	3	5	5	5	4	5
62	4	4	3	3	3	3	2	3	2	2	2	3
63	4	3	3	3	4	4	3	3	4	2	4	4
64	3	4	3	3	4	4	4	4	4	3	3	4
65	2	2	2	3	4	4	5	2	2	2	2	1
66	4	3	4	4	4	3	4	4	4	4	3	5
67	4	5	5	4	5	4	4	3	3	4	5	4
68	3	2	3	2	4	3	4	4	3	3	4	4
69	4	4	3	4	3	3	4	4	5	4	4	4
70	5	5	5	5	5	5	4	5	5	4	5	5
71	5	4	4	4	4	4	5	4	5	4	5	5
72	5	4	4	5	4	5	5	4	4	3	5	4
73	4	4	3	4	3	2	3	2	2	3	5	5
74	4	4	5	4	4	5	5	5	4	5	4	4
75	3	2	3	2	3	4	3	2	2	2	5	5
76	4	3	3	3	3	3	4	3	4	3	4	3
77	3	3	3	3	4	4	4	4	3	2	4	4
78	5	4	4	4	3	3	3	5	5	4	4	3
79	3	4	4	4	4	4	5	4	5	3	4	4
80	3	3	2	4	3	4	4	3	5	2	3	4
81	3	4	4	4	4	4	4	4	4	3	4	4
82	3	4	4	4	5	4	4	5	5	4	5	4
83	3	4	3	3	4	3	4	4	4	3	4	5
84	2	3	3	3	3	3	3	4	3	3	4	4
85	3	3	3	2	3	3	3	5	5	4	3	3
86	3	4	4	5	4	4	3	3	3	3	5	4
87	5	5	5	3	5	5	5	2	3	2	4	4
88	5	5	5	5	4	4	4	5	4	4	4	5
89	5	5	4	5	2	4	3	5	3	3	5	5

(continued)

Table 9.3 (continued)

Res. number	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}	V_{11}	V_{12}
90	2	4	4	4	3	3	2	3	2	3	3	3
91	3	3	3	3	3	3	3	3	2	4	4	4
92	3	5	5	4	4	3	3	5	4	5	5	5
93	3	2	2	2	3	3	2	3	3	3	4	4
94	2	3	2	3	4	4	3	4	4	4	4	4
95	4	3	4	4	3	3	2	3	3	4	5	4
96	2	2	2	2	3	4	3	3	4	4	4	3
97	4	4	3	4	4	4	4	5	3	3	5	3
98	4	4	4	4	4	3	3	5	4	4	4	4

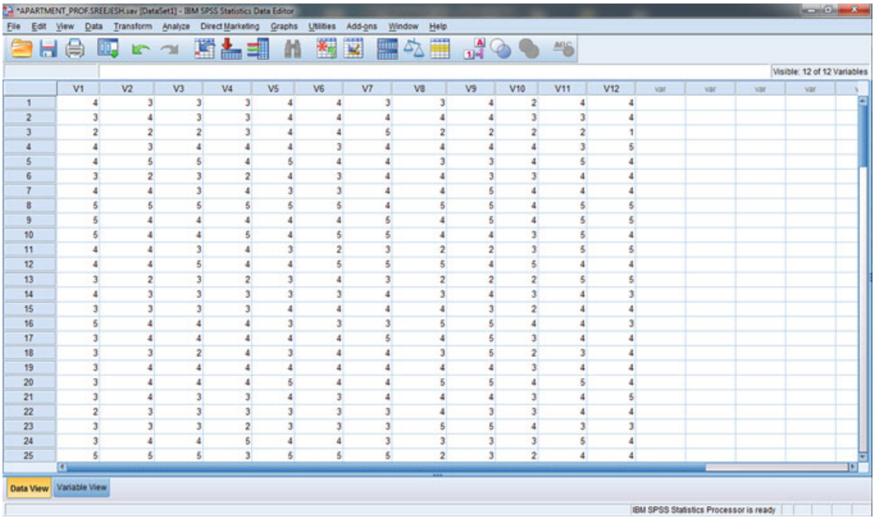


Fig. 9.2 SPSS data editor with PDI, price and value consciousness and sales proneness

1. **Initial solution** (under **Statistics)**
   The selection of this option in SPSS will produce the unrotated FA outputs such as communalities, Eigen values and percentage of variance explained. This output could be used as benchmark and compared with rotated factor solution results.
2. **Coefficients** (under **Correlation Matrix)**
   This selection will produce the output of correlation matrix (12 × 12 correlation matrix) for the 12 items, which are selected for FA. This correlation matrix summarizes the interrelationship among a set of selected variables or, as in our case, a set of items in a scale. The inadequate correlation among the selected items indicates irrelevancy of FA. The understanding of how these correlations

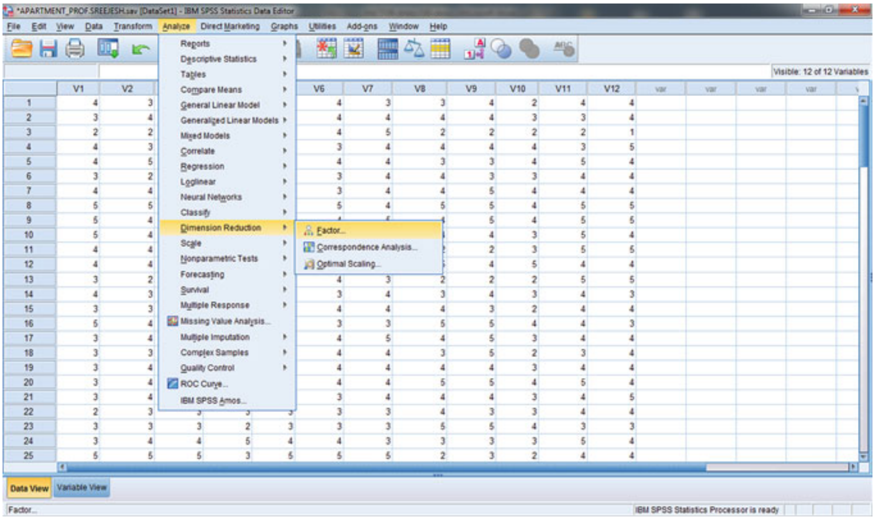


Fig. 9.3 Direction to get factor analysis in IBM SPSS 20.0

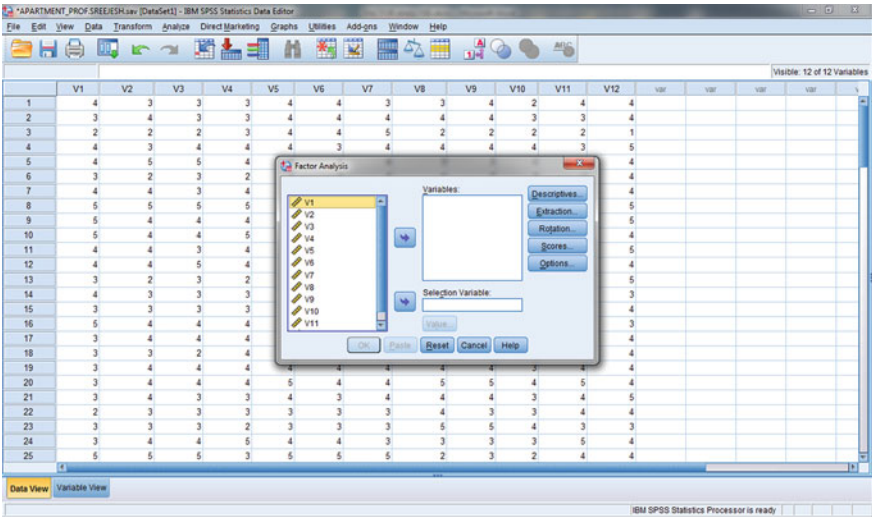


Fig. 9.4 SPSS factor analysis window

are obtained is beyond the scope of this book, and therefore, the procedure is not discussed in this chapter.

3. **Determinant** (under **Correlation Matrix**)
   This selection will produce the output of determinant of the correlation matrix. In general, the values for the determinants of the matrices can range between

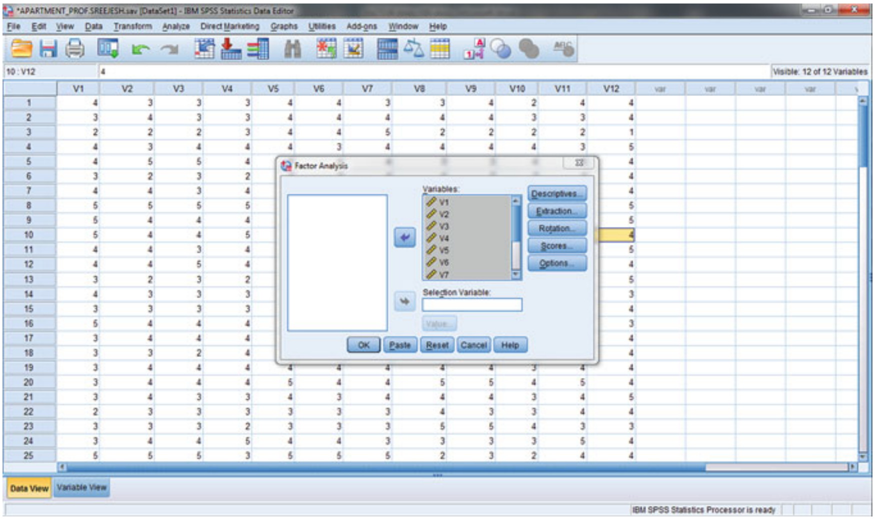


Fig. 9.5 After entering Variables into the factor analysis window

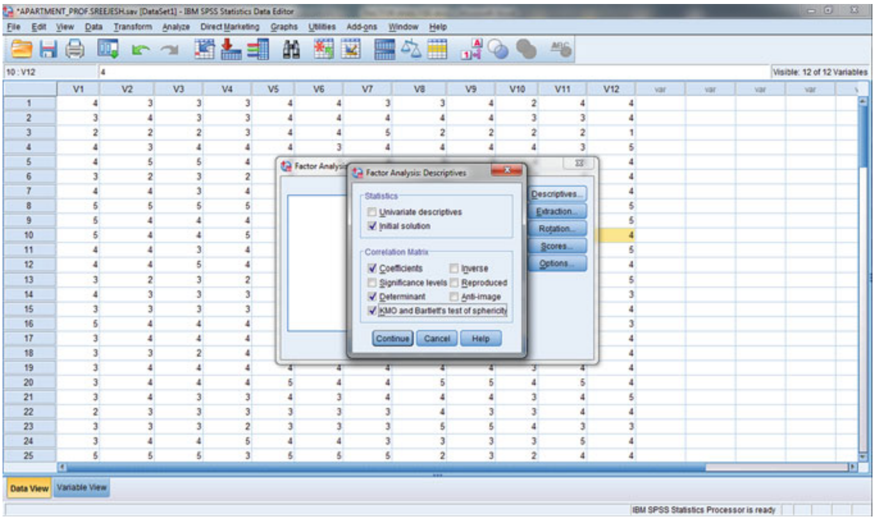


Fig. 9.6 Factor analysis descriptive window

$-\infty$ and $+\infty$. However, values for the determinant of a correlation matrix range only between 0 and 1.00. When all of the off-diagonal elements in the correlation matrix are equal to 0, the determinant of that matrix will be equal to 1. That would mean that matrix is an identity matrix and there is no correlation between the items. Therefore, FA will result into as many factors as there are items. A value of 0 for a determinant indicates that there is at least one linear dependency in the matrix. That means that one or more columns (or rows) in the matrix can be obtained by linear transformations of other columns (or rows) or combinations of columns (or rows). Linear dependency could occur, for example, when one item is highly correlated with the other (e.g. $r > 0.80$). It could also occur when one person's answers are the exact replica or are a linear combination of another person's answers. When this occurs, SPSS for Windows issues the following warning: ***Determinant*** = 0.000 and ***this Matrix is Not Positive Definite***. Therefore, the ideal range of determinant in FA would be in between 0 and 1.00 (neither exact 0 nor exact 1).

9.5 KMO and Bartlett's Test of Sphericity (Under Correlation Matrix)

KMO stands for Kaiser–Meyer–Olkin and named after statisticians, and it is considered to be the measure of sampling adequacy. As a general guideline, it is considered that a value greater than 0.60 shows acceptable sampling adequacy, greater than 0.70 shows good sampling adequacy, greater than 0.80 shows very good sampling adequacy and greater than 0.90 shows excellent sampling adequacy. It means that a larger values indicates greater likelihood that the correlation matrix is not an identity matrix and null hypothesis will be rejected (null hypothesis = the correlation matrix is an identity matrix).

Once you complete the selection of these four components, click on **Continue** to go the main window of **FA.**

Step 6 In step 6, click on **Extraction** at the bottom of Fig. 9.7. In this window, select **PCA** from the **Methods** pull-down. *Select (2)* **Unrotated factor solution** (under **Display), Correlation matrix** and also select the **Scree plot** box, Check **Based on Eigenvalues eigen values greater than one** under **Extract**. This setting instructs the computer to extract based on eigen values greater than one criteria. Click on **Continue.**

9.6 Principle Component Analysis

The objective behind the usage of principle component analysis other than other methods is that PCA summarizes the interrelationships among a set of original variables in terms of a smaller set of orthogonal (i.e. uncorrelated) principal components that are linear combinations of original variables.

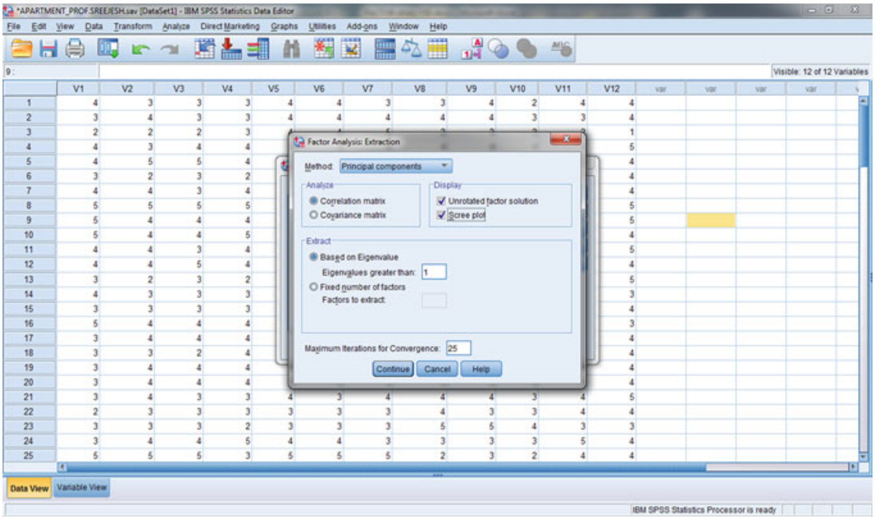


Fig. 9.7 Factor analysis extraction window

9.7 Unrotated Factor Solution

In the **Extraction** window, we selected **Unrotated factor solution** in the **Display** side (right side). This option usually selected to compare this unrotated solution with the rotated solution (the importance of rotation would be discussed in the later part of this chapter).

9.8 Scree Plot

We generally use scree plot to select the number of extracted factors. The selection of scree plot produces a graphical display in which eigen values on the Y-axis and number of factors on the *X*-axis. The word 'scree' typically represents a kink or distinct binding or a trailing point. For identifying the number of extracted factors, we can have look into the scree plot, in which we would consider only those factors that are present before the scree or kink begins.

9.9 Eigen Values and Eigen Values Greater than One

Eigen value represents the amount of variance in all of the items that can be explained by a given principal component or factor. In PCA, the total amount of variance available is equal to the number of items; therefore, dividing the eigen

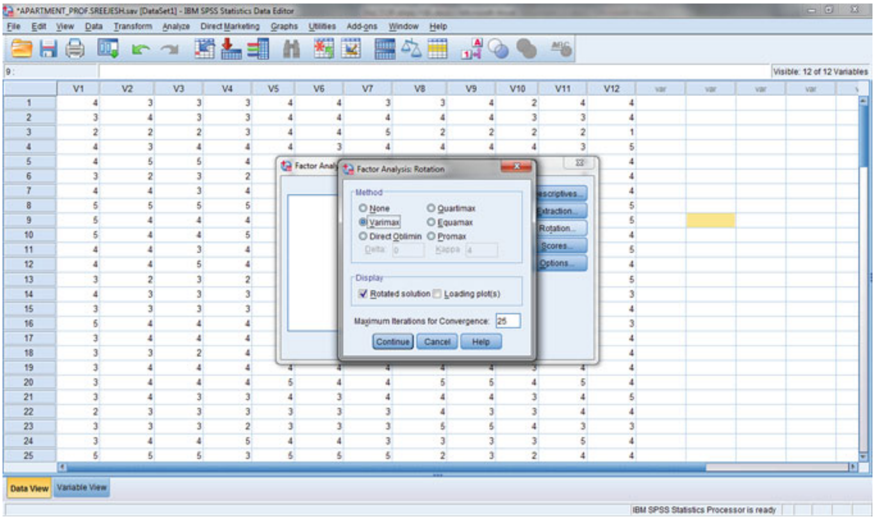


Fig. 9.8 Factor analysis rotation window

value by the number of items gives the proportion of total item variance accounted for by the given principal component or factor. The rationale for the eigen values greater than one criterion is that any individual factor should account for the variance of at least a single variable if it is to be retained for interpretation. This criterion is considered to be more reliable when the number of variables under study is between 20 and 50.

Step 7 In step 7, click on **Rotation,** which will give you Fig. 9.8. Click on **Varimax** and then make sure **Rotated solution** is also checked. Click on **Continue.**

9.10 Rotated Solution

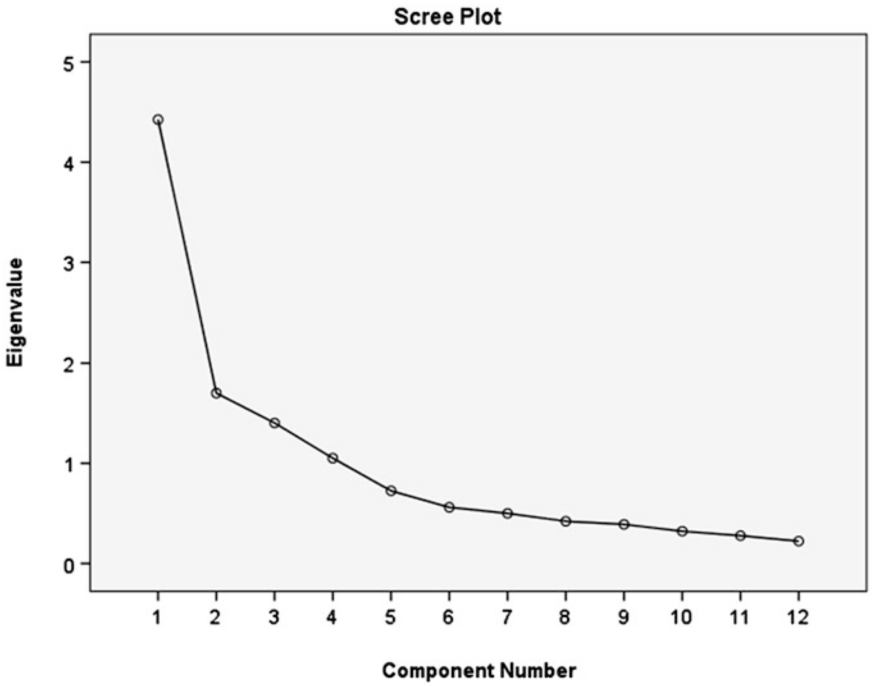
Unrotated factor solutions achieve the objective of data reduction, but it would not provide information that offers the most adequate interpretation of the variables under examination. Therefore, for achieving more theoretically meaningful factor solution, we employ a rotational method. In most of the cases, rotation of the factors improves the interpretation by reducing some of the ambiguities that often accompany initial unrotated factor solutions.

Step 8 Click on **Options,** which will give you Fig. 9.8.
Click on **Suppress absolute values less than** and type 4 (point 4) in the box (see Fig. 9.8).
Suppressing small factor loadings makes the output easier to read.
Click on **Continue** then **OK.** Compare Output 1 to your output and syntax.

9.11 SPSS Syntax Method

```
FACTOR
  /VARIABLES V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12
  /MISSING LISTWISE
  /ANALYSIS V1 V2 V3 V4 V5 V6 V7 V8 V9 V10 V11 V12
  /PRINT INITIAL CORRELATION DET KMO EXTRACTION ROTATION
  /FORMAT BLANK (.40)
  /PLOT EIGEN
  /CRITERIA MINEIGEN (1) ITERATE (25)
  /EXTRACTION PC
  /CRITERIA ITERATE (25)
  /ROTATION VARIMAX
  /METHOD = CORRELATION.
```

9.12 Output 1: IBM SPSS 20.0 Output for Factor Analysis



9.13 Results and Interpretation

The aforementioned steps (Figs. 9.2, 9.3, 9.4, 9.5, 9.6, 9.7, 9.8, 9.9 and 9.10) give a number of tables depending on the option selected by the researcher for doing FA in IBM SPSS 20.0. The first table in FA output is **correlation matrix.** Table 9.4: Presents 12 × 12 correlation matrix for the 20 items specified in the study. This correlation matrix summarizes the interrelationship among set of variables or, as in this example, a set of items in a scale. The correlation ranges between −1.00 and +1.00, with higher absolute values indicating a stronger relationship between two variables. A positive value indicates direct relationship between two items. In Table 9.4, for example, the correlation between V2 (*Do you think that the various types and brands of this product available in the market are all very alike or are all very different*) and V3 (*How important would it be to you to make a right choice of this product from ABC?*) was 0.734. This means that respondents who scored high on V2 also scored high on V3. A negative value indicates an inverse relationship between two items: high scores on one item are associated with low scores on the second item. Given the magnitude of correlation between variables, it is clear that the hypothesized factor model appears to be appropriate. Looking at the correlation table for larger number of variable is a tiresome job, and therefore, we have some other measures to check the adequacy of correlation or interrelationship between the factored items, and these measures are as follows:

1 The determinant

This is the determinant of the matrix (12 × 12), and the value is located under the correlation matrix (Table 9.4). In our example, we got a value of 0.007, its neither exact zero or exact one, which is greater than the cut-off value of 0.00001. Therefore, we can conclude that the correlation matrix is neither an identity matrix nor a singular matrix. This value confirms the assumption that there are sufficient interrelationships among our study items.

2 Bartlett's test of Spherecity and the KMO

Table 9.5 gives the results of KMO and Bartlett's test (Bartlett 1950). Bartlett's test of Spherecity tests the null hypothesis that the correlation matrix is an identity matrix (there is no relationship between items) and follows Chi square distribution. Larger the value of Bartlett's test indicates greater likelihood the correlation matrix is not an identity matrix and null hypothesis will be rejected. In this example, The Bartlett's test value (452.25) is significant (i.e. a significance value of less than 0.05); this means that we may reject the null hypothesis that our correlation matrix is an identity matrix and will conclude that the variables are correlated highly enough to provide a reasonable basis for FA. The KMO test is a measure of sampling adequacy. The KMO measure should be greater than 0.70 and is inadequate if less than 0.60. All these three measures (determinant, Bartlett's test and KMO) show the evidence that there are good interrelationships

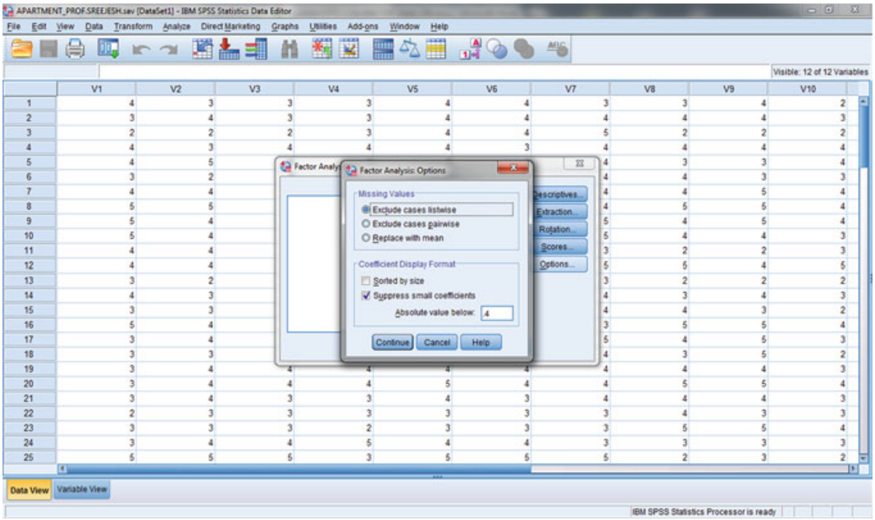


Fig. 9.9 Factor analysis options window

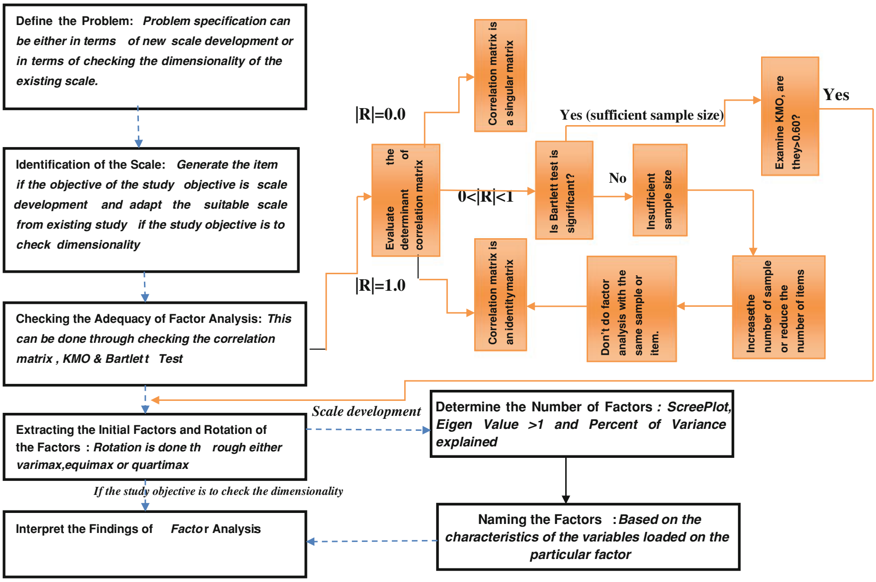


Fig. 9.10 Decision-making process behind factor analysis

between study items and measures. Therefore, we can go for extracting factors using these items.

Table 9.6 presents the communality of each item or measure to the common factor (i.e. the proportion of variance in each variable accounted for by the

Table 9.4 Correlation matrix[a]

		V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12
Correlation	V1	1.000	0.543	0.468	0.467	0.201	0.246	0.169	0.283	0.286	0.248	0.309	0.322
	V2	0.543	1.000	0.734	0.706	0.362	0.257	0.138	0.392	0.247	0.269	0.323	0.317
	V3	0.468	0.734	1.000	0.641	0.444	0.323	0.204	0.380	0.203	0.346	0.421	0.341
	V4	0.467	0.706	0.641	1.000	0.246	0.252	0.115	0.344	0.246	0.214	0.322	0.284
	V5	0.201	0.362	0.444	0.246	1.000	0.534	0.502	0.223	0.227	0.199	0.238	0.226
	V6	0.246	0.257	0.323	0.252	0.534	1.000	0.598	0.225	0.242	−0.033	0.203	0.114
	V7	0.169	0.138	0.204	0.115	0.502	0.598	1.000	0.169	0.314	0.015	0.055	0.076
	V8	0.283	0.392	0.380	0.344	0.223	0.225	0.169	1.000	0.565	0.535	0.175	0.202
	V9	0.286	0.247	0.203	0.246	0.227	0.242	0.314	0.565	1.000	0.380	0.077	0.106
	V10	0.248	0.269	0.346	0.214	0.199	−0.033	0.015	0.535	0.380	1.000	0.240	0.278
	V11	0.309	0.323	0.421	0.322	0.238	0.203	0.055	0.175	0.077	0.240	1.000	0.504
	V12	0.322	0.317	0.341	0.284	0.226	0.114	0.076	0.202	0.106	0.278	0.504	1.000

[a] Determinant = 0.007

Table 9.5 KMO and Bartlett's test

Kaiser–Meyer–Olkin measure of sampling adequacy		0.806
	Approx. Chi Square	452.251
Bartlett's test of sphericity	df	66
	Sig.	0.000

Table 9.6 Initial communalities

	Initial	Extraction
V1	1.000	0.494
V2	1.000	0.825
V3	1.000	0.745
V4	1.000	0.759
V5	1.000	0.640
V6	1.000	0.749
V7	1.000	0.759
V8	1.000	0.744
V9	1.000	0.711
V10	1.000	0.717
V11	1.000	0.715
V12	1.000	0.722

Extraction method: Principal component analysis

common factors). While using PCA for factor extraction, we could get as many factors as variables. When all factors are included in the solution, all of the variance of each variable is accounted for by the common factors. Thus, the proportion of variance accounted for by the common factors, or the communality of a variable is 1 for all the variables.

In Table 9.7, total variance is divided into 12 possible factors, because the use of PCA. In our factor extraction option in SPSS, we have selected factor extraction option as 'Based on eigen value and eigen value > 1' criteria. Which means that the factor should explains more information than a single item would have explained. Based on eigen value criteria, we have retained only four factor solution. These four factors account for 23.33, 18.07, 16.47 and 13.60 % of the total variance, respectively. That is, almost 71.49 % of the total variance is attributable to these three factors. The remaining eight factors together account for only approximately 28.51 % of the variance. Thus, a model with three factors may be adequate to represent the data. From the scree plot, it again appears that a four-factor model should be sufficient to represent the data set.

Table 9.8 shows the component matrix, it is an unrotated component analysis factor matrix. The values inside the table show correlation of each variable to the respective extracted factor. Here in our example, we have extracted four factors, the value of V1 (0.650) for the component 1 shows the correlation of item number one to the component 1. These coefficients, called *factor loadings*, indicate how closely the variables are related to each factor. However, as the factors are

Table 9.7 Total variance explained

Component	Initial eigen values			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	Variance	Cumulative (%)	Total	Variance (%)	Cumulative (%)	Total	Variance (%)	Cumulative (%)
1	4.425	36.876	36.876	4.425	36.876	36.876	2.801	23.339	23.339
2	1.701	14.171	51.046	1.701	14.171	51.046	2.168	18.070	41.409
3	1.403	11.689	62.736	1.403	11.689	62.736	1.977	16.475	57.884
4	1.051	8.757	71.493	1.051	8.757	71.493	1.633	13.609	71.493
5	0.726	6.046	77.539						
6	0.561	4.673	82.212						
7	0.500	4.166	86.378						
8	0.422	3.515	89.893						
9	0.391	3.255	93.148						
10	0.322	2.686	95.834						
11	0.277	2.311	98.145						
12	0.223	1.855	100.000						

Extraction method: Principal component analysis

Table 9.8 Component matrix[a]

	Component			
	1	2	3	4
V1	0.650			
V2	0.786			
V3	0.807			
V4	0.718			
V5	0.592	0.492		
V6	0.519	0.672		
V7	0.412	0.763		
V8	0.617		0.597	
V9	0.509		0.652	
V10	0.498		0.535	
V11	0.531			0.504
V12	0.510			0.570

Extraction method: Principal component analysis
[a] 4 components extracted

unrotated (the factors were extracted on the basis of the proportion of total variance explained), significant cross-loadings have occurred, thus it becomes very difficult to identify which variables actually throng to each component or factor. There the role of 'rotation' comes and helps to give good and meaningful interpretation. Technically, rotation means tilting the axes of each factor toward right in order to facilitate the variables to have closer association or affinity with only a single factor.

Table 9.9 gives the Rotated Factor Matrix, which contains four loadings, is key for understanding the results of the analysis. The FA using PCA has sorted the items (V1 to V12) into four overlapping groups of items, each which has a loading

Table 9.9 Rotated component matrix[a]

	Component			
	1	2	3	4
V1	0.641			
V2	0.876			
V3	0.771			
V4	0.855			
V5		0.725		
V6		0.839		
V7		0.861		
V8			0.805	
V9			0.779	
V10			0.759	
V11				0.799
V12				0.819

Extraction method: Principal component analysis
Rotation method: Varimax with Kaiser normalization
[a] Rotation converged in 5 iterations

Table 9.10 Component transformation matrix

Component	1	2	3	4
1	0.702	0.421	0.437	0.372
2	−0.291	0.889	−0.102	−0.337
3	−0.296	−0.124	0.884	−0.339
4	−0.578	0.127	0.129	0.795

Extraction method: Principal component analysis
Rotation method: Varimax with Kaiser normalization

of |0.40| or higher (|0.40| means the absolute value, or value without considering the sign, is greater than 0.40). Actually, every item has some loading from every factor, but there are blanks in the matrix where weights were less than |0.40|, which had achieved using the suppress option in SPSS.

The loading coefficients in this table generated through an orthogonal rotation (Varimax), which shows the correlation coefficient of each item to the component or factor, so they ranges from -1.0 to $+1.0$. The negative loading coefficient simply means that the relationship of the respective item to the component or factor in opposite direction. As a rule of thumb, it is considered that a factor loading lower than |0.40| is considered as bad, greater than |0.40| considered as good (Table 9.10).

In summary, it can be concluded that FA has identified four factors from the list of 12 variables. In the main, these factors are represented by the specific statements written to reflect the four different perception constructs: Product decision involvement, price consciousness, value consciousness and sales proneness.

9.14 Key Statistics

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.

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

Eigen value. The eigen value 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 respondent on the derived factors.

KMO measure of sampling adequacy. The KMO measure of sampling adequacy is an index used to examine the appropriateness of FA. High values (between 0.5 and 1.0) indicate that FA is appropriate. Values below 0.5 imply that FA may not be appropriate.

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

Residuals. Residuals are 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 eigen values against the number of factors in order of extraction.

9.15 Review Questions

1. Discuss the possible reasons for the use of FA with the data **(FACTOR)**.
2. Produce a correlation matrix for the 12 variables (scale items). Does it appear that FA would be appropriate for these data?
3. Do a principal component analysis (with rotation if necessary for interpretation) using the data. How many factors should be retained? What is the percentage of variance accounted for each factor? Interpret the factors.

Reference

Hair JF Jr, Black WC, Babin BJ, Anderson RE (2010) Multivariate data analysis: a global perspective. Pearson, London