

Factor Analysis

- Let's Recap:
- **Correlation:**
- In market research we are often interested in summarizing the strength of association between two metric (Ratio/Interval) variables.
- Examples:
- How strongly are sales related to Ad. Expenditure.
- Market share and size of sales force.
- Consumer perception of quality and price.
- In the above cases product moment correlation is the most widely used statistic, summarizing the strength of association between two metric variables say X and Y.
- It indicates the degree to which the variation in one variable, X is related to variation in another variable, Y.

- It was originally proposed by Karl Pearson,
- Synonyms: Pearson correlation coefficient, bivariate correlation, correlation coefficient.
- $r = \text{Cov}_{xy} / S_x S_y$
- S_x , S_y represents standard deviations and Cov_{xy} measures covariance between X and Y.
- r^2 measures the proportion of variance in one variable that is explained by other.
- **Partial correlation:**
- It is a measure of association between two variables after controlling or adjusting for the effects of one or more additional variable.
- Example:
- How strongly are sales related to Ad. Expenditure, when the effect of price is controlled.

- **Factor analysis** is a general name denoting a class of procedures, primarily used for data reduction and summarization, with no loss of information.
- In Market research, there may be large number of variables, most of which are correlated and must be reduced to a manageable level. Relationship among set of inter-related variables are examined and represented in terms of few underlying factors.
- It is an Multivariate **interdependence technique**: no distinction between dependent(criterion) and independent(Predictor) variables. It does not attempt identification of causal relationship.
- Why Factor analysis is used:
 - To identify underlying dimensions, or **factors**, that 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, which can be used in subsequent MVA like regression. All the uses are exploratory in nature and therefor it is also called as Exploratory factor analysis (EFA).
- In Research it can be used in segmentation, Product research (Brand attributes influencing consumer choice), service quality etc.

- Factor Analysis Model:
- The common factors themselves can be expressed as linear combinations of the observed variables.
- $F_i = W_{i1}X_1 + W_{i2}X_2 + W_{i3}X_3 + \dots + W_{ik}X_k$

Where:

F_i = estimate of i th factor

W_i = weight or factor score coefficient

k = number of variables

Statistics Associated with Factor Analysis

- **Bartlett's test of sphericity.** Bartlett's test of sphericity is used to test the hypothesis that the variables are uncorrelated in the population (i.e., the population correlation matrix is an identity matrix)
- **Correlation matrix:** It is a set of correlation coefficients between a number of variables.
- **Correlation:** It is a numerical measure of the degree of agreement between two set of scores. It run from +1 to -1.

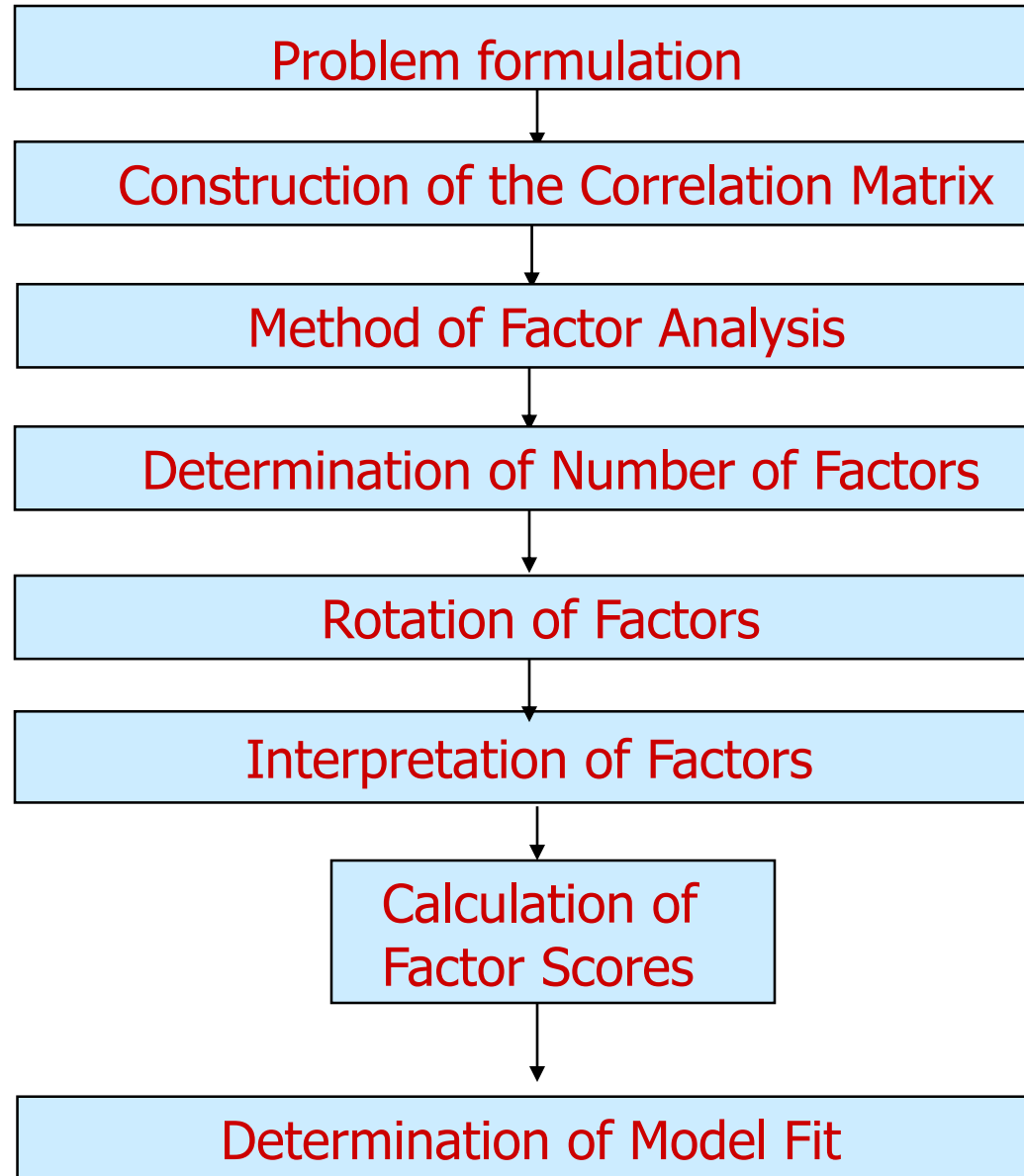
Statistics Associated with Factor Analysis

- **Communality.** Amount of variance a variable shares with all the other variables. This is the proportion of variance explained by the common factors.
- **Eigenvalue.** Represents the total variance explained by each factor.
- **Factor loadings.** Are simple Correlations between the variables and the factors.
- **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.

Statistics Associated with Factor Analysis

- **Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy.** Used to examine the appropriateness of factor analysis. High values (between 0.5 and 1.0) indicate appropriateness. Values below 0.5 imply not.
- **Percentage of variance.** The percentage of the total variance attributed to each factor.
- **Scree plot.** A scree plot is a plot of the Eigenvalues against the number of factors in order of extraction.

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. The variables should be 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 (sample size) as there are variables.

- Illustration: Suppose the researcher wants to determine the underlying benefits consumer seeks from the purchase of a tooth paste. A sample of 30 respondents was taken. The respondents were asked to indicate their degree of agreement on a 7 point scale (1= strongly disagree,7= strongly agree)
- To determine benefits from toothpaste
- Responses were obtained on 6 variables:
 - V1: It is imp to buy toothpaste to prevent cavities
 - V2: I like a toothpaste that gives shiny teeth
 - V3: A toothpaste should strengthen your gums
 - V4: I prefer a toothpaste that freshens breathe
 - V5: Prevention of tooth decay is not imp
 - V6: The most imp consideration is attractive teeth
- Data obtained are coded in table 19.1

Another Example of Factor Analysis

Table 19.1

RESPONDENT NUMBER	V1	V2	V3	V4	V5	V6
1	7.00	3.00	6.00	4.00	2.00	4.00
2	1.00	3.00	2.00	4.00	5.00	4.00
3	6.00	2.00	7.00	4.00	1.00	3.00
4	4.00	5.00	4.00	6.00	2.00	5.00
5	1.00	2.00	2.00	3.00	6.00	2.00
6	6.00	3.00	6.00	4.00	2.00	4.00
7	5.00	3.00	6.00	3.00	4.00	3.00
8	6.00	4.00	7.00	4.00	1.00	4.00
9	3.00	4.00	2.00	3.00	6.00	3.00
10	2.00	6.00	2.00	6.00	7.00	6.00
11	6.00	4.00	7.00	3.00	2.00	3.00
12	2.00	3.00	1.00	4.00	5.00	4.00
13	7.00	2.00	6.00	4.00	1.00	3.00
14	4.00	6.00	4.00	5.00	3.00	6.00
15	1.00	3.00	2.00	2.00	6.00	4.00
16	6.00	4.00	6.00	3.00	3.00	4.00
17	5.00	3.00	6.00	3.00	3.00	4.00
18	7.00	3.00	7.00	4.00	1.00	4.00
19	2.00	4.00	3.00	3.00	6.00	3.00
20	3.00	5.00	3.00	6.00	4.00	6.00
21	1.00	3.00	2.00	3.00	5.00	3.00
22	5.00	4.00	5.00	4.00	2.00	4.00
23	2.00	2.00	1.00	5.00	4.00	4.00
24	4.00	6.00	4.00	6.00	4.00	7.00
25	6.00	5.00	4.00	2.00	1.00	4.00
26	3.00	5.00	4.00	6.00	4.00	7.00
27	4.00	4.00	7.00	2.00	2.00	5.00
28	3.00	7.00	2.00	6.00	4.00	3.00
29	4.00	6.00	3.00	7.00	2.00	7.00
30	2.00	3.00	2.00	4.00	7.00	2.00

Construct the Correlation Matrix

- The analytical process is based on a matrix of correlations between the variables.
- If the Bartlett's test of sphericity is not rejected, then factor analysis is not appropriate.
- If the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy is small, then the correlations between pairs of variables cannot be explained by other variables and factor analysis may not be appropriate.

Correlation Matrix

Table 19.2

Variables	V1	V2	V3	V4	V5	V6
V1	1.000					
V2	-0.530	1.000				
V3	0.873	-0.155	1.000			
V4	-0.086	0.572	-0.248	1.000		
V5	-0.858	0.020	-0.778	-0.007	1.000	
V6	0.004	0.640	-0.018	0.640	-0.136	1.000

Determine the Method of Factor Analysis

- In **Principal components analysis**, the total variance in the data is considered.
 - Used to determine the min number of factors that will account for max variance in the data.
- In **Common factor analysis**,(excluding it for the time being) the factors are estimated based only on the common variance.
 - Communalities are inserted in the diagonal of the correlation matrix.
 - Used to identify the underlying dimensions and when the common variance is of interest.

Results of Principal Components Analysis

Table 19.3

Bartlett's Test Apprx. chi-square=111.3, df=15, significance=0.00 Kaiser-Meyer-Olkin msa=0.660				
<u>Communalities</u>				
Variables	Initial	Extraction		
V1	1.000	0.926		
V2	1.000	0.723		
V3	1.000	0.894		
V4	1.000	0.739		
V5	1.000	0.878		
V6	1.000	0.790		
<u>Initial Eigen values</u>				
	Factor	Eigen value	% of variance	Cumulat. %
	1	2.731	45.520	45.520
	2	2.218	36.969	82.488
	3	0.442	7.360	89.848
	4	0.341	5.688	95.536
	5	0.183	3.044	98.580
	6	0.085	1.420	100.000

Results of Principal Components Analysis

Table 19.3, cont.

<u>Extraction Sums of Squared Loadings</u>			
Factor	Eigen value	% of variance	Cumulat. %
1	2.731	45.520	45.520
2	2.218	36.969	82.488

<u>Factor Matrix</u>		
Variables	Factor 1	Factor 2
V1	0.928	0.253
V2	-0.301	0.795
V3	0.936	0.131
V4	-0.342	0.789
V5	-0.869	-0.351
V6	-0.177	0.871

<u>Rotation Sums of Squared Loadings</u>			
Factor	Eigenvalue	% of variance	Cumulat. %
1	2.688	44.802	44.802
2	2.261	37.687	82.488

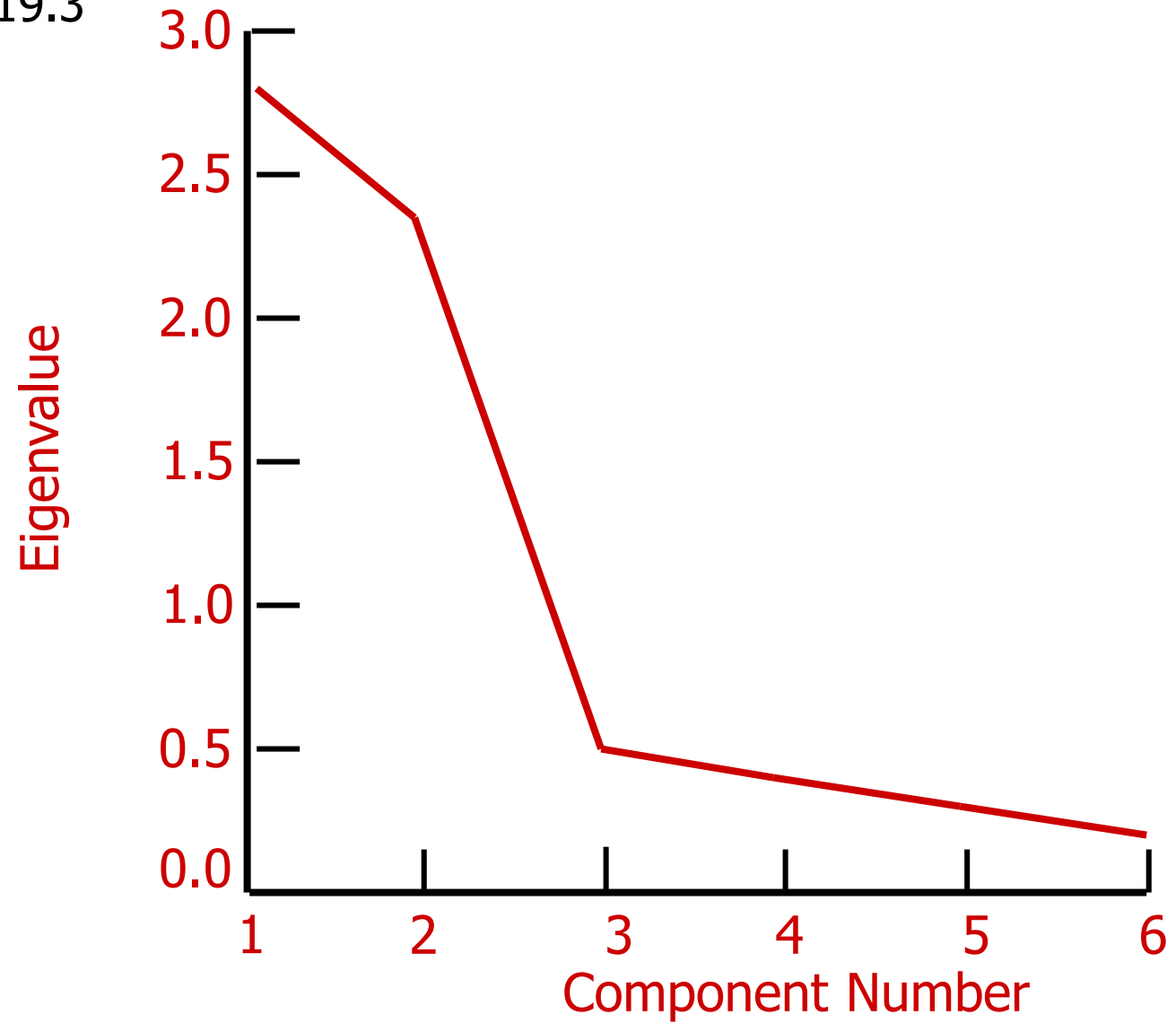
Results of Principal Components Analysis

Table 19.3, cont.

<u>Rotated Factor Matrix</u>		
Variables	Factor 1	Factor 2
V1	0.962	-0.027
V2	-0.057	0.848
V3	0.934	-0.146
V4	-0.098	0.845
V5	-0.933	-0.084
V6	0.083	0.885

Scree Plot

Fig. 19.3



Determine the Number of Factors

- **Determination Based on Eigenvalues.** Only factors with Eigenvalues greater than 1.0 are retained.
- **Determination Based on Scree Plot.** A scree plot is a plot of the Eigenvalues against the number of factors in order of extraction. The point at which the scree begins denotes the true number of factors.
- **Determination Based on Percentage of Variance.**

Interpret Factors

- A factor can be interpreted in terms of the variables that load high on it.

- In order to understand the technique let us take a research problem where Factor analysis is used.
- **Research Objective:**
- To explore the underlying factors related to consumer decision making for selecting health insurance policy.

customers who have claimed their policies and have gone through the tough process of claiming. Hence, a research attempt was made to ascertain if there is any logical basis to choose health insurance policies.

Research Methodology

Problem

- Do most of the customers have the knowledge of how to select a health insurance policy?
- What features do they look for in health insurance policy?
- How much importance do they give to certain features?

Design and Methods of Study

The study is mainly based on field survey and is exploratory in nature. The sources of data are mainly primary and were collected through questionnaires.

used to explore the underlying factor structure of a set of observed variables without imposing a preconceived structure on the outcome. ANOVA was done to observe the relationship between demographic variables and the resultant factors.

Objectives

- To explore the underlying factors related to consumer decision making for selecting health insurance policy; and
- To study the association between customers of different socioeconomic backgrounds and the importance they assigned to the resultant factor.

Hypothesis

H_{01} = *There is no significant relationship between the resultant factor; transparency, product feature, service feature, network hospital, brand image and price with respect to age, education and occupation.*

Table 1: Customer Profile

		Frequency	%	Cumulative %
Gender	Male	278	76.2	76.2
	Female	87	23.8	100.0
Age	< 30	112	30.7	30.7
	30-45	179	49.0	79.7
	45-60	74	20.3	100.0
Education	Up to HSC	5	1.4	1.4
	Graduation	152	41.6	43.0
	Post-Graduation	135	37.0	80.0
	Professional	73	20.0	100.0
Occupation	Salaried	220	60.3	60.3
	Self-Employed	35	9.6	69.9
	Professional	36	9.9	79.7
	Business	50	13.7	93.4
	Student	12	3.3	96.7
	Housewife	12	3.3	100.0
Income	Less than ₹3 lakh	103	28.2	28.2
	₹3-4 lakh	126	34.5	62.7
	₹4-5 lakh	76	20.8	83.6
	Above ₹5 lakh	60	16.4	100.00

Source: Computed with the help of SPSS Data analysis

Table 2: KMO and Bartlett's Test (Factor Analysis)

Bartlett's Test of Sphericity	Kaiser-Meyer-Olkin Measure of Sampling Adequacy	0.924
	Approx. Chi-Square	4846.923
	Df	435
	Sig.	0.000

Table 3: Eigen Values and Sum of Square Loading

Component	Initial Eigen Values			Extraction Sums of Squared Loadings		
	Total	Variance	Cumulative	Total	Variance	Cumulative
1	10.441	34.804	34.804	10.441	34.804	34.804
2	1.953	6.510	41.314	1.953	6.510	41.314
3	1.596	5.319	46.633	1.596	5.319	46.633
4	1.329	4.430	51.062	1.329	4.430	51.062
5	1.120	3.732	54.794	1.120	3.732	54.794
6	1.028	3.428	58.222	1.028	3.428	58.222
7	0.957	3.190	61.412			
8	0.924	3.079	64.491			
9	0.855	2.849	67.339			
10	0.778	2.593	69.932			
11	0.737	2.457	72.389			
12	0.695	2.318	74.707			
13	0.631	2.103	76.811			
14	0.608	2.026	78.837			
15	0.593	1.978	80.814			
16	0.551	1.835	82.650			
17	0.525	1.749	84.398			
18	0.500	1.668	86.066			
19	0.471	1.572	87.638			
20	0.465	1.551	89.188			
21	0.430	1.432	90.621			
22	0.407	1.357	91.978			
23	0.392	1.307	93.284			
24	0.369	1.229	94.513			
25	0.321	1.071	95.584			
26	0.304	1.013	96.597			
27	0.299	0.996	97.593			
28	0.272	0.907	98.500			
29	0.236	0.788	99.288			
30	0.214	0.712	100.000			

58% extraction

Note: Extraction Method: Principal Component Analysis. Eigen value above 1 is taken into consideration for determining the factors.

Factors Affecting the Choice of Health Insurance Policies in India

Table 4: Total Variance Explained

Component	Rotation Sums of Squared Loadings		
	Total	of Variance	Cumulative
1	4.028	13.428	13.428
2	3.960	13.201	26.628
3	3.179	10.598	37.226
4	2.255	7.515	44.742
5	2.141	7.136	51.878
6	1.903	6.344	58.222

Note: Extraction Method: Principal Component Analysis.

Source: Computed with the help of SPSS Data analysis

Table 5 : Factor Loadings

> 0.4

	Component					
	1	2	3	4	5	6
q61	0.111	0.139	0.222	0.176	0.002	0.713
q62	0.116	0.135	-0.092	-0.139	0.437	0.617
q63	0.168	0.645	0.190	-0.017	0.232	0.067
q64	0.232	0.136	0.104	0.654	0.040	0.289
q65	0.187	0.190	0.185	0.719	0.161	0.031
q66	0.091	0.461	0.566	0.247	0.168	0.074
q67	0.461	0.180	0.136	0.436	0.123	-0.068
q68	0.233	0.599	0.234	0.073	0.329	-0.013
q69	0.167	0.201	0.193	-0.054	0.677	-0.075
q610	0.378	-0.070	0.447	0.222	0.295	0.180
q611	0.088	0.322	0.696	0.124	0.064	0.194
q612	0.159	0.329	0.711	0.177	0.181	-0.045
q613	0.258	0.057	0.566	0.380	-.038	0.052
q614	0.114	0.247	0.598	-0.036	0.225	0.251
q615	0.199	0.069	0.222	0.270	-0.066	0.620
q616	0.268	0.345	0.454	-0.275	0.200	0.372
q617	0.577	0.193	-0.025	0.032	0.336	0.046

Table 5 (Cont.)

	Component					
	1	2	3	4	5	6
q618	0.665	0.177	0.307	-0.017	0.138	0.156
q619	0.720	0.141	0.144	0.107	0.081	0.055
q620	0.599	0.081	0.090	0.328	0.274	0.130
q621	0.056	0.308	0.154	0.213	0.593	0.133
q622	0.174	0.312	0.151	0.288	0.555	0.131
q623	0.436	0.227	0.310	0.381	-0.092	0.230
q624	0.650	0.289	0.127	0.088	-0.158	0.199
q625	0.620	0.393	0.025	0.143	0.033	0.090
q626	0.327	0.580	0.178	0.032	0.269	0.111
q627	0.210	0.703	0.132	0.121	0.180	0.116
q628	0.235	0.693	0.322	0.264	0.002	0.146
q629	0.193	0.707	0.167	0.218	0.092	0.123
q630	0.539	0.128	0.150	0.231	0.127	0.073

Note: Extraction Method: Principal Component Analysis.

Source: Computed with the help of SPSS Data analysis

Table 6 (Cont.)

Component					
1	2	3	4	5	6
Q619	Q627	Q612			
Q620	Q628	Q613			
Q623	Q629	Q614			
Q624		Q616			
Q625					
Q630					

Source: Computed with the help of SPSS Data analysis

Table 7: Component Explanation (1)

Factor1(Transparency)	Explanation
Q 67	Hassle-free claim settlement
Q617	Ethical conduct
Q618	Cashless card and bond
Q619	Competitive pricing for services rendered
Q620	Mandatory disclosure of product features
Q623	Reputed hospitals in network lists
Q624	Minimum premium
Q625	Sum assured
Q630	Provider communicates openly with its customers

Source: Computed with the help of SPSS Data analysis

Table 8: Component Explanation (2)

Factor 2 (Product Feature)	Explanation
Q 63	Portability benefits
Q 68	Death claims
Q626	Tax benefits
Q627	Image of TPA
Q628	Turnaround time maintained during reimbursement
Q629	Pre-policy health checkups

Source: Computed with the help of SPSS Data analysis

Table 9: Component Explanation (3)

Factor 3 (Service Feature)	Explanation
Q66	Accidental protection offered
Q610	Pre-existing diseases
Q611	Assistance by company people
Q612	Assistance by TPA
Q613	Proactiveness of network hospitals
Q614	Waiting period
Q616	Capping and loading

Source: Computed with the help of SPSS Data analysis

Table 10: Component Explanation (4)

Factor 4 (Network)	Explanation
Q64	Critical illness benefit
Q65	Number of network hospitals

Source: Computed with the help of SPSS Data analysis

Table 11: Component Explanation (5)

Factor 5 (Brand Image)	Explanation
Q69	Return or growth offered on premiums
Q621	Brand equity of the provider
Q622	Computerized operations (Record keeping)

Source: Computed with the help of SPSS Data analysis

- SPSS steps for Factor analysis
- Select Analyze from Menu bar.
- Click Dimension reduction and then factor analysis.
- Move all the variables to the Variable Box.
- Click on descriptive in the pop up window. In the statistics box check Initial solution. In the correlation Matrix box check KMO and Bartlets test and also check Reproduced. Click Continue..
- Click on Extraction. In the pop up window for the Method select principal components. In the analyze box, check CORRELATION MATRIX, in the extract box select based on EIGEN VALUES and enter 1 for EIGEN VALUES greater than box. In the display box check UNROTATED FACTOR SOLUTION. Click continue.
- Click on Rotation, in the Method box check VARIMAX. In the DISPLAY box check ROTATED SOLUTION. Click continue..
- Click on SCORES. In the pop up window, check display factor score coefficient matrix. Click continue...OK....