Principal Component Analysis

Objective

Perform **PCA** on the given data
Suppose that the researcher wanted to determine the underlying benefits that consumers seek from the purchase of a toothpaste. A sample of 30 participants was interviewed using street interviewing. The participants were asked to indicate their degree of agreement with the following statements using a seven-point scale (1 = strongly disagree, 7 = strongly agree) :

About Data Variables

V1 : It is important to buy a toothpaste that prevents cavities.
V2 : I like a toothpaste that gives shiny teeth.
V3 : A toothpaste should strengthen your gums.
V4 : I prefer a toothpaste that freshens breath.
V5 : Prevention of tooth decay should be an important benefit offered by a toothpaste.
V6 : The most important consideration in buying a toothpaste is attractive teeth.

Commands:

1. Select **ANALYZE** from the SPSS menu bar.
2. Click **Dimension Reduction**
3. Click on **Factor** enter all eight variables into the Variables box Next,
4. Click on **Descriptives** and check the boxes next to Coefficients, Significance levels, Anti-image, and KMO and Bartlett's test of sphericity.
5. Select **Reproduced**, which requests the reproduced correlation matrix that we will use for assessing the goodness-of-fit of the factor solution.
6. Select **Initial solution** and **Univariate descriptives** in the upper part of the dialog box to display useful summary statistics of your data.
7. Click **Continue** Extraction and choose Principal components in the Method drop-down menu. Under Extract, you can determine the rule for factor extraction.
8. Select **Unrotated factor** solution and **Scree plot**.
9. Click on **Continue** Under Rotation, you can choose between several orthogonal and oblique rotation methods. Select the Varimax procedure and click on Continue.
10. Click on **Continue**,
11. Click OK

Descriptive Analysis of Variables

Descriptive Statistics			
	Mean	**Std. Deviation**	**Analysis N**
V1	3.93	1.982	30
V2	3.90	1.373	30
V3	4.10	2.057	30
V4	4.10	1.373	30
V5	3.50	1.907	30
V6	4.17	1.392	30

- There are **30** observations in each variable without any missing observation.
- Mean of variable **V1** is **3.93** and Standard Deviation is **1.982**
- Mean of variable **V2** is **3.90** and Standard Deviation is **1.373**
- Mean of variable **V3** is **4.10** and Standard Deviation is **2.057**
- Mean of variable **V4** is **4.10** and Standard Deviation is **1.373**
- Mean of variable **V5** is **3.50** and Standard Deviation is **1.907**
- Mean of variable **V6** is **4.17** and Standard Deviation is **1.392**

Correlation Matrix							
		V1	**V2**	**V3**	**V4**	**V5**	**V6**
Correlation	**V1**	1.000	-.053	.873	-.086	-.858	.004
	V2	-.053	1.000	-.155	.572	.020	.640
	V3	.873	-.155	1.000	-.248	-.778	-.018
	V4	-.086	.572	-.248	1.000	-.007	.640
	V5	-.858	.020	-.778	-.007	1.000	-.136
	V6	.004	.640	-.018	.640	-.136	1.000

- The correlation between **V1** and **V2** is **-0.053**, **V1** and **V3** is **0.873**, **V1** and **V4** is **-0.086, V1** and **V5** is **-0.858** & **V1** and **V6** is **-0.004**
- The correlation between **V2** and **V1** is **-0.053**, **V2** and **V3** is **-0.155**, **V2** and **V4** is **-0.572, V2** and **V5** is **0.20** & **V2** and **V6** is **0.640**
- The correlation between **V3** and **V1** is **0.873**, **V3** and **V2** is **-0.155**, **V3** and **V4** is **-0.284, V3** and **V5** is **-0.778** & **V3** and **V6** is **-0.018**
- The correlation between **V4** and **V1** is **-0.086**, **V4** and **V2** is **0.572**, **V4** and **V3** is **-0.248, V4** and **V5** is **-0.007** & **V4** and **V6** is **0.640**
- The correlation between **V5** and **V1** is **-0.858**, **V5** and **V2** is **0.20**, **V5** and **V3** is **-0.778, V5** and **V4** is **-0.007** & **V5** and **V6** is **-0.136**
- The correlation between **V6** and **V1** is **0.004**, **V6** and **V2** is **0.640**, **V6** and **V3** is **-0.018, V6** and **V4** is **0.640** & **V6** and **V5** is **-0.136**

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

The null hypothesis, that the population correlation matrix is an identity matrix, is rejected by Bartlett's test of sphericity.

The approximate **chi-square statistic** is **111.314** with **15** degrees of freedom, which is significant at the **0.05** level.

The value of the **KMO statistic (0.660)** is also **large (>0.5).**

Thus, factor analysis may be considered an appropriate technique for analyzing the correlation matrix of Table

Anti-image Matrices							
		V1	V2	V3	V4	V5	V6
Anti-image Covariance	V1	.141	-.048	-.106	-.051	.104	.073
	V2	-.048	.520	.044	-.093	-.049	-.219
	V3	-.106	.044	.186	.114	.025	-.073
	V4	-.051	-.093	.114	.457	.025	-.206
	V5	.104	-.049	.025	.025	.237	.067
	V6	.073	-.219	-.073	-.206	.067	.413
Anti-image Correlation	V1	.621[a]	-.179	-.657	-.203	.570	.303
	V2	-.179	.697[a]	.140	-.192	-.139	-.474
	V3	-.657	.140	.679[a]	.390	.118	-.265
	V4	-.203	-.192	.390	.637[a]	.074	-.474
	V5	.570	-.139	.118	.074	.769[a]	.213
	V6	.303	-.474	-.265	-.474	.213	.561[a]

a. Measures of Sampling Adequacy(MSA)

Communalities		
	Initial	Extraction
V1	1.000	.926
V2	1.000	.723
V3	1.000	.894
V4	1.000	.739
V5	1.000	.878
V6	1.000	.790
Extraction Method : Principal Component Analysis.		

- The Communalities of **V1** after extraction is **0.926** which is initially **1.00**
- The Communalities of **V2** after extraction is **0.723** which is initially **1.00**
- The Communalities of **V3** after extraction is **0.894** which is initially **1.00**
- The Communalities of **V4** after extraction is **0.739** which is initially **1.00**
- The Communalities of **V5** after extraction is **0.878** which is initially **1.00**
- The Communalities of **V6** after extraction is **0.990** which is initially **1.00**

The commonalities for the variances under 'Extraction' are different from those under 'Initial' because all of the variances associated with the variables are not explained unless all the factors are retained.

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	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						
Extraction Method: Principal Component Analysis.									

The section labelled 'Initial eigenvalues' gives the eigenvalues.

The eigenvalues for the factors are, as expected, in decreasing order of magnitude as we go from factor 1 to factor 6. The eigenvalue for a factor indicates the total variance attributed to that factor. The total variance accounted for by all the six factors is 6.00, which is equal to the number of variables.

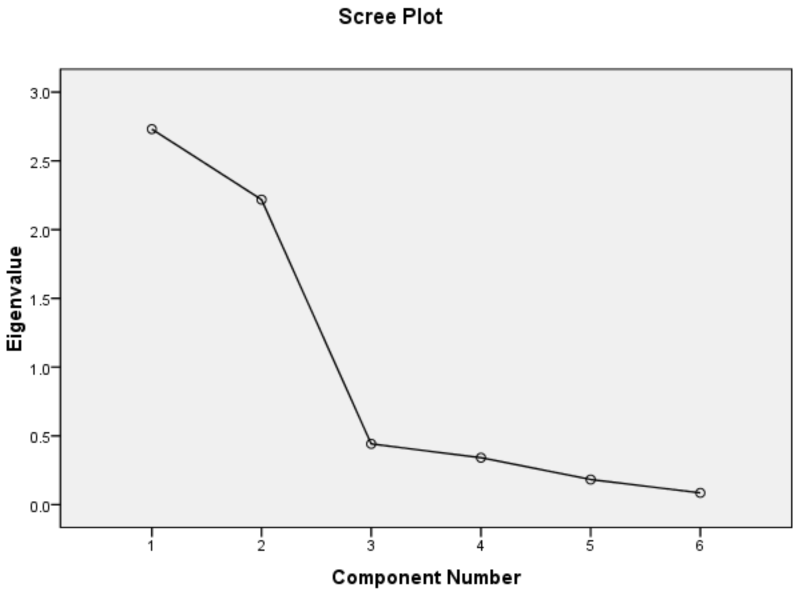
- Factor 1 accounts for a variance of **2.731**, which is **(2.731/6)** or **45.52%** of the total variance
- Factor 2 accounts for a variance of **2.218**, which is **(2.218/6)** or **36.97%** of the total variance and cumulative **82.488**
- Factor 2 accounts for a variance of **0.442**, which is **(0.442/6)** or **7.360%** of the total variance and cumulative **89.848**

Several considerations are involved in determining the number of factors that should be used in the analysis.

In Table, we see that the eigenvalue greater than **1.0** (default option) results in two factors being extracted. Our a priori knowledge tells us that toothpaste is bought for two major reasons.

The "Extraction Sums od Squared Loadings" of first two factors is **82.488%**

The "Rotation Sums od Squared Loadings" of first two factors is **82.488%**



From the scree plot, a distinct break occurs at three factors. Finally, from the cumulative percentage of variance accounted for, we see that the first two factors account for **82.49%** of the variance and that the gain achieved in going to three factors is marginal. Furthermore, split-half reliability also indicates that two factors are appropriate. Thus, two factors appear to be reasonable in this situation.

Component Matrix[a]		
	Component	
	1	**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
Extraction Method: Principal Component Analysis.		
a. 2 components extracted.		

Factor 1 is at least somewhat correlated with five of the six variables (absolute value of factor loading greater than 0.3).

Factor 2 is at least somewhat correlated with four of the six variables.

Moreover, variables 2 and 5 load at least somewhat on both the factors.

Reproduced Correlations							
		V1	**V2**	**V3**	**V4**	**V5**	**V6**
Reproduced Correlation	**V1**	.926[a]	-.078	.902	-.117	-.895	.057
	V2	-.078	.723[a]	-.177	.730	-.018	.746
	V3	.902	-.177	.894[a]	-.217	-.859	-.051
	V4	-.117	.730	-.217	.739[a]	.020	.748
	V5	-.895	-.018	-.859	.020	.878[a]	-.152
	V6	.057	.746	-.051	.748	-.152	.790[a]
Residual[b]	**V1**		.024	-.029	.031	.038	**-.052**
	V2	.024		.022	**-.158**	.038	**-.105**
	V3	-.029	.022		-.031	**.081**	.033
	V4	.031	-.158	-.031		-.027	**-.107**
	V5	.038	.038	.081	-.027		.016
	V6	-.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.							

If there are many large residuals, the factor model does not provide a good fit to the data and the model should be reconsidered. .In the upper-right triangle of the 'Reproduced correlation matrix' of Table , we see that only five residuals are larger than **0.05**, indicating an acceptable model fit.

Rotated Component Matrix[a]		
	Component	
	1	**2**
V1	0.962	-0.027
V2	-0.057	0.848
V3	0.934	-0.146
V4	-0.098	0.854
V5	-0.933	-0.084
V6	0.083	0.885
Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.		
a. Rotation converged in 3 iterations.		

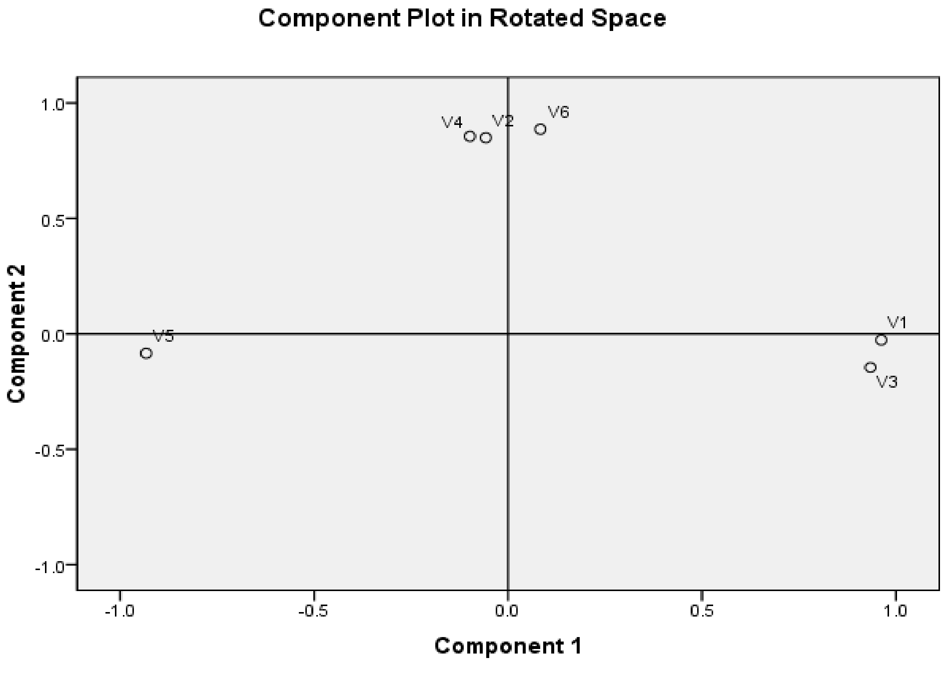
From above Table, by comparing the varimax rotated factor matrix with the unrotated matrix (entitled simply 'Factor matrix'), we can see how rotation achieves simplicity and enhances interpretability. Whereas five variables correlated with factor 1 in the unrotated matrix, only variables V1, V3 and V5 correlate highly with factor 1 after rotation. The remaining variables, V2, V4 and V6, correlate highly with factor 2. Furthermore, no variable correlates highly with both the factors. The rotated factor matrix forms the basis for interpretation of the factors.

In the rotated factor matrix of Table, factor 1 has high coefficients for variables V1 (prevention of cavities) and V3 (strong gums), and a negative coefficient for V5 (prevention of tooth decay is not important). Therefore, this factor may be labelled a health benefit factor. Note that a negative coefficient for a negative variable (V5) leads to a positive interpretation that prevention of tooth decay is important. Factor 2 is highly related with variables V2 (shiny teeth), V4 (fresh breath) and V6 (attractive teeth).

Thus factor 2 may be labelled a social benefit factor. A plot of the factor loadings, given in Figure, confirms this interpretation. Variables V1, V3 and V5 (denoted 1, 3 and 5, respectively) are at the end of the horizontal axis (factor 1), with V5 at the end opposite to V1 and V3, whereas variables V2, V4 and V6 (denoted 2, 4 and 6) are at the end of the vertical axis (factor 2).

One could summarize the data by stating that consumers appear to seek two major kinds of benefits from toothpaste: health benefits and social benefits.

Component Transformation Matrix		
Component	**1**	**2**
1	0.957	-0.290
2	0.290	0.957
Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.		



Logistic Regression

Objective:

Perform the **Binary Logistic Regression** of given data.

Commands:

1. Click on **Analyze**
2. Click on **Regression**
3. Click on **Binary Logistic Regression**
4. From the pop-up window move Loyalty variable into **Dependent**.
5. Move all the independent (Brand, Production and Shopping) variables into **Covariates**.
6. Click on **Continue**.
7. Click on **OK**

Case Processing Summary			
Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	30	100.0
	Missing Cases	0	.0
	Total	30	100.0
Unselected Cases		0	.0
Total		30	100.0

There are 30 observations in our dataset without any missing value.

Dependent Variable Encoding	
Original Value	Internal Value
0	0
1	1

Here 0 denotes No while 1 denotes the Yes.

Block 0: Beginning Block

Classification Table[a,b]					
	Observed		Predicted		
			Loyalty		Percentage Correct
			0	1	
Step 0	Loyalty	0	0	15	.0
		1	0	15	100.0
	Overall Percentage				50.0

In Beginning the classification is 100% correct.

Variables in the Equation							
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	.000	.365	.000	1	1.000	1.000

Variables not in the Equation					
			Score	df	Sig.
Step 0	Variables	Brand	13.264	1	.000
		Production	.462	1	.497
		Shopping	4.727	1	.030
	Overall Statistics		14.019	3	.003

Block 1: Method = Enter

Omnibus Tests of Model Coefficients				
		Chi-square	df	Sig.
Step 1	Step	18.117	3	.000
	Block	18.117	3	.000
	Model	18.117	3	.000

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	23.471[a]	.453	.604

- The value of **Cox & Snell R Square** is **0.45** means that **45%** variability is explained by the explanatory variables.
- The value of **Negalkerke R Square** is **0.604** means that **60%** variability is explained by the explanatory variables.

Classification Table[a]					
	Observed		Predicted		
			Loyalty		Percentage Correct
			0	1	
Step 1	Loyalty	0	12	3	80.0
		1	3	12	80.0
	Overall Percentage				80.0

From the above table, we can easily see that the correct classification of 0 is 12 out of 15, which means 80% and the correct classification of 1 is 12 out of 15, which means 80%.

Variables in the Equation							
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Brand	1.274	.479	7.075	1	.008	3.575
	Production	.186	.322	.335	1	.563	1.205
	Shopping	.590	.491	1.442	1	.230	1.804
	Constant	-8.642	3.346	6.672	1	.010	.000

$$P_{(Loyalty)} = \frac{e^{-8.642+1.274\times Brand+0.186\times Production+0.590\times Shopping}}{1+\ e^{-8.642+1.274\times Brand+0.186\times Production+0.590\times Shopping}}$$

Here intercept (**-8.642**) can be interpreted as the value of log odds when the value of all explanatory variables are set to be 0 and slopes shows the average change in log odds associated with one unit increase in x.

Independent variables production and shopping have p-values **0.56** and **0.23** respectively which is far more than level of significance 0.05, so there is no clear evidence of a real association between loyalty and these two independent variables

The diagonal elements of the confusion matrix indicate correct predictions, while the off-diagonals represent incorrect predictions.

Only the Brand variable is significant because the p-value is 0.008.

The Wald Statistics of Brand is 7.075 means it is significant.
The Wald Statistics of Production is 0.335 means it is insignificant.
The Wald Statistics of Shopping is 1.442 means it is insignificant.

Multi-Dimensional Scaling

Objective

Perform **MDS** on the given data.

Commands

- Select **ANALYZE** from the SPSS menu bar.
- Click **SCALE** and then MULTIDIMENSIONAL SCALING (ALSCAL).
- Move ‘Aqua-Fresh [var00001]’, ‘Crest [var00002]’, ‘Colgate [var00003]’, ‘Aim [var00004]’, ‘Gleem [var00005]’, ‘Plus White [var00006]’, ‘Ultra-Brite [var00007]’, ‘Close-Up [var00008]’, ‘Pepsodent [var00009]’ and ‘Sensodyne [var00010]’ into the VARIABLES box.
- In the **DISTANCES** box check DATA ARE DISTANCES. SHAPE should be SQUARE SYMMETRIC (default).
- Click **MODEL**. In the pop-up window, in the LEVEL OF MEASUREMENT box, check INTERVAL. In the SCALING MODEL box, check EUCLIDEAN DISTANCE. In the CONDITIONALITY box, check MATRIX. Click **CONTINUE**.
- Click **OPTIONS**. In the pop-up window, in the DISPLAY box, check GROUP PLOTS, DATA MATRIX and MODEL AND OPTIONS SUMMARY. Click **CONTINUE**.
- Click **OK**

Case Processing Summary[a]					
Cases					
Valid		Missing		Total	
N	Percent	N	Percent	N	Percent
10	100.0%	0	0.0%	10	100.0%

There is no missing value in this dataset.

Alscal

Alscal Procedure Options

Data Options:-

Number of Rows (Observations/Matrix).	10
Number of Columns (Variables) . . .	10
Number of Matrices	1
Measurement Level	Interval
Data Matrix Shape	Symmetric
Type	Dissimilarity
Approach to Ties	Leave Tied
Conditionality	Matrix
Data Cutoff at	0.000000

There are 10 rows and 10 columns in our dataset.

Model Options :-

Model	**Euclid**
Maximum Dimensionality	2
Minimum Dimensionality	2
Negative Weights	Not Permitted
In our Model we use Euclidian Distance.	

Maximum dimensionality of our model is 2 while minimum dimensionality is 2.

Algorithmic Options :-

Maximum Iterations	30
Convergence Criterion	.00100
Minimum S-stress	.00500
Missing Data Estimated by	Ulbounds

Iteration history for the 2 dimensional solution (in squared distances)

Young's S-stress formula 1 is used.		
Iteration	**S-stress**	**Improvement**
1	0.25120	
2	0.22286	0.02834
3	0.22090	0.00196
4	0.22063	0.00027

Iterations stopped because S-stress improvement is less than **0.001000**

For Matrix
Stress = 0.20170
RSQ = 0.77827

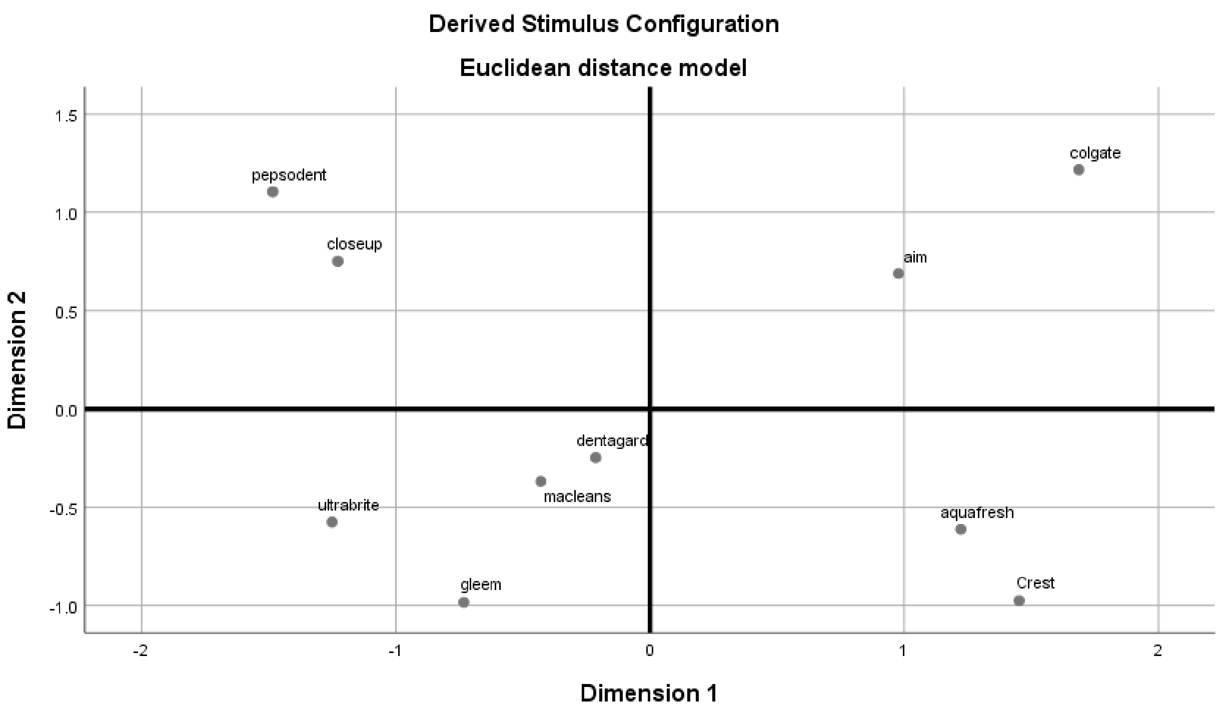
- The value of Stress is **0.20170** which means that our model is good fitted because lesser values of stress indicates the better fits.
- The value of RSQ (R^2) is **0.78** implies that our model is good fitted because higher values of R^2 indicates the better fits.

Configuration derived in 2 dimensions

Stimulus Coordinates			
		Dimension	
Stimulus Number	Stimulus Name	1	2
1	aquafresh	1.2219	-0.6121
2	Crest	1.4520	-0.9751
3	colgate	1.6866	1.2168
4	aim	0.9772	0.6893
5	gleem	-0.7324	-0.9835
6	macleans	-0.4296	-0.3684
7	ultrabri	-1.2506	-0.5749
8	closeup	-1.2276	0.7508
9	pepsoden	-1.4837	1.1048
10	dentagar	-0.2138	-0.2477

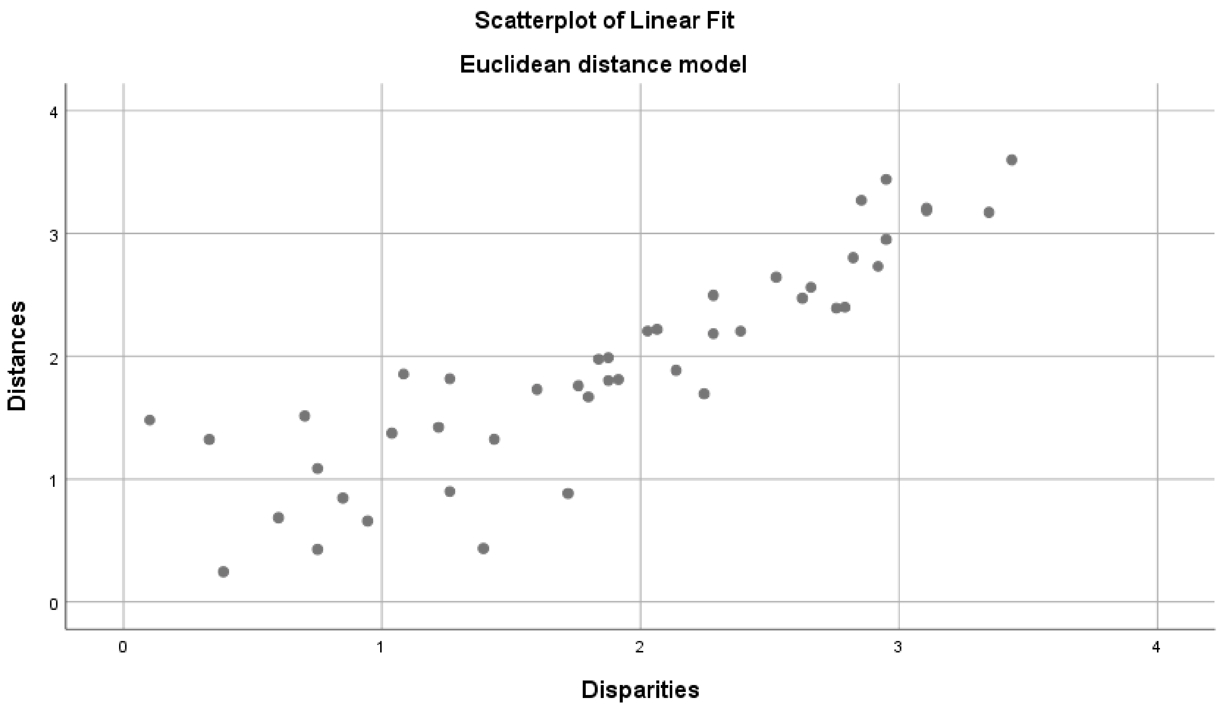
We can see that the values of **aquafresh, crest, colgate and aim** are positive so these factor are dimension 1 while **colgate** and **pepsodent** are positive so these factors are dimension 2.

Abbreviated Name	Extended Name
aquafres	**aquafresh**
dentagar	**dentagard**
pepsoden	**pepsodent**
ultrabri	**ultrabrite**



- From this perceptual map we can say that **Pepsodent** and **closeup** are similar / closer to each other than others.
- **Colgate** and **aim** are similar / closer to each other than others.
- **Aquafresh** and **Crest** are similar / closer to each other than others.
- **Dentagard** and **Macleans** are similar & **Macleans** and **Ultrabrite** & **Macleans** and **Gleem** to each other than others.

From Stimulus Co-ordinate table we can infer that Aquafresh, Crest, Colgate and Aim is highly loaded on dimension 1 and Pepsodent, Close-Up, Ultrabright and Gleem is lowly loaded on Dimension 1 which is also evident from the perceptual or spatial map. Whereas, for Dimension 2 Aim, Close-Up, Colgate and Pepsodent is highly loaded. On the other hand, Ultrabrite , Gleem, Crest and Aquafresh is lowly loaded on the Dimension 2.



Disparities and **Distances** are linearly correlated to each other.

Clustering

Objective

Perform **Clustering** on the given data

Commands

1. Select **ANALYZE** from the SPSS menu bar.
2. Click **CLASSIFY** and then **HIERARCHICAL CLUSTER.**
3. Move ‘Fun [v1]’, ‘Bad for Budget [v2]’, ‘Eating Out [v3]’, ‘Best Buys [v4]’, ‘Don’t Care [v5]’ and ‘Compare Prices [v6]’ into the VARIABLES box.
4. In the **CLUSTER** box check CASES (default option). In the DISPLAY box check STATISTICS and PLOTS (default options).
5. Click **STATISTICS**. In the pop-up window, check AGGLOMERATION SCHEDULE. In the CLUSTER MEMBERSHIP box check RANGE OF SOLUTIONS. Then for MINIMUM NUMBER OF CLUSTERS enter ‘2’, and for MAXIMUM NUMBER OF CLUSTERS enter ‘4’. Click **CONTINUE**.
6. Click **PLOTS**. In the pop-up window, check DENDROGRAM. In the ICICLE box check ALL CLUSTERS (default). In the ORIENTATION box, check VERTICAL. Click **CONTINUE**.
7. 7 Click **METHOD**. For CLUSTER METHOD select WARD’S METHOD. In the MEASURE box check INTERVAL and select SQUARED EUCLIDEAN DISTANCE. Click **CONTINUE**.
8. Click **OK**

Case Processing Summary[a,b]

Cases					
Valid		**Missing**		**Total**	
N	Percent	N	Percent	N	Percent
20	100.0	0	.0	20	100.0

There is no missing value in our dataset.

a. Squared Euclidean Distance used	
b. Ward Linkage	

We use **Squared Euclidean Distance** and **Ward Linkage** method to perform clustering.

Ward Linkage

Agglomeration Schedule						
Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	14	16	1.000	0	0	6
2	6	7	2.000	0	0	7
3	2	13	3.500	0	0	15
4	5	11	5.000	0	0	11
5	3	8	6.500	0	0	16
6	10	14	8.167	0	1	9
7	6	12	10.500	2	0	10
8	9	20	13.000	0	0	11
9	4	10	15.583	0	6	12
10	1	6	18.500	0	7	13
11	5	9	23.000	4	8	15
12	4	19	27.750	9	0	17
13	1	17	33.100	10	0	14
14	1	15	41.333	13	0	16
15	2	5	51.833	3	11	18
16	1	3	64.500	14	5	19
17	4	18	79.667	12	0	18
18	2	4	172.667	15	17	19
19	1	2	328.600	16	18	0

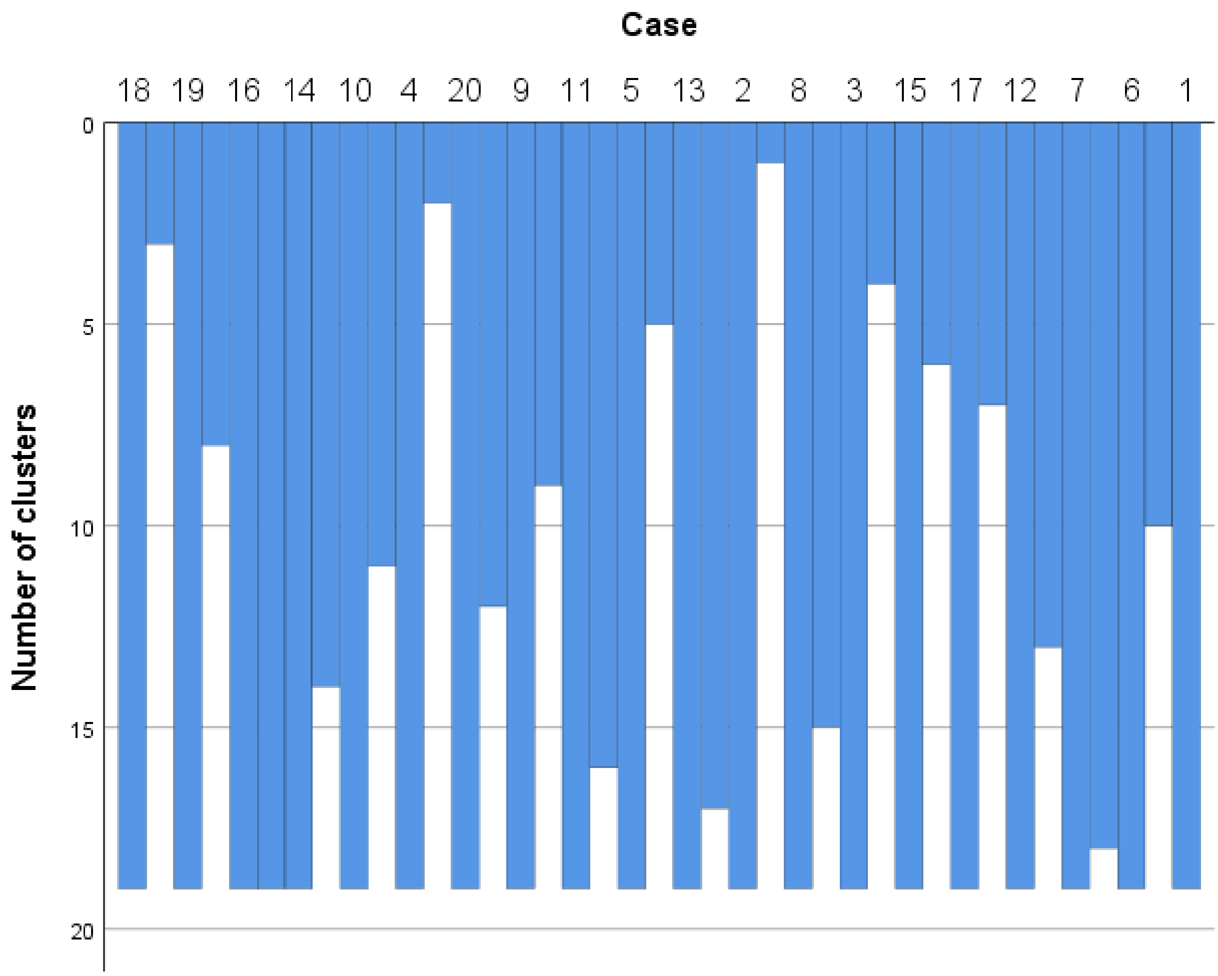
14, 16, 4, 19, 18, and **2** are similar to each other and they all are in the same cluster.

Table Agglomerative Schedule can be interpreted as at stage 1 observation 6 is clustered with observation 7 in Cluster A (say), now as next stage column suggested go to stage 6 where observation 6 is clustered with observation 12, implies now Cluster A contains three observations 6, 7 and 12. Now at stage 6 the next stage column is 9 so we go to stage 9 where observation 6 is clustered with observation 1, hence observation 1 is also added in Cluster A. Likewise, we move forward till be get all the observation in either of the two clusters which can further be verified by the Cluster membership table and Dendrogram.

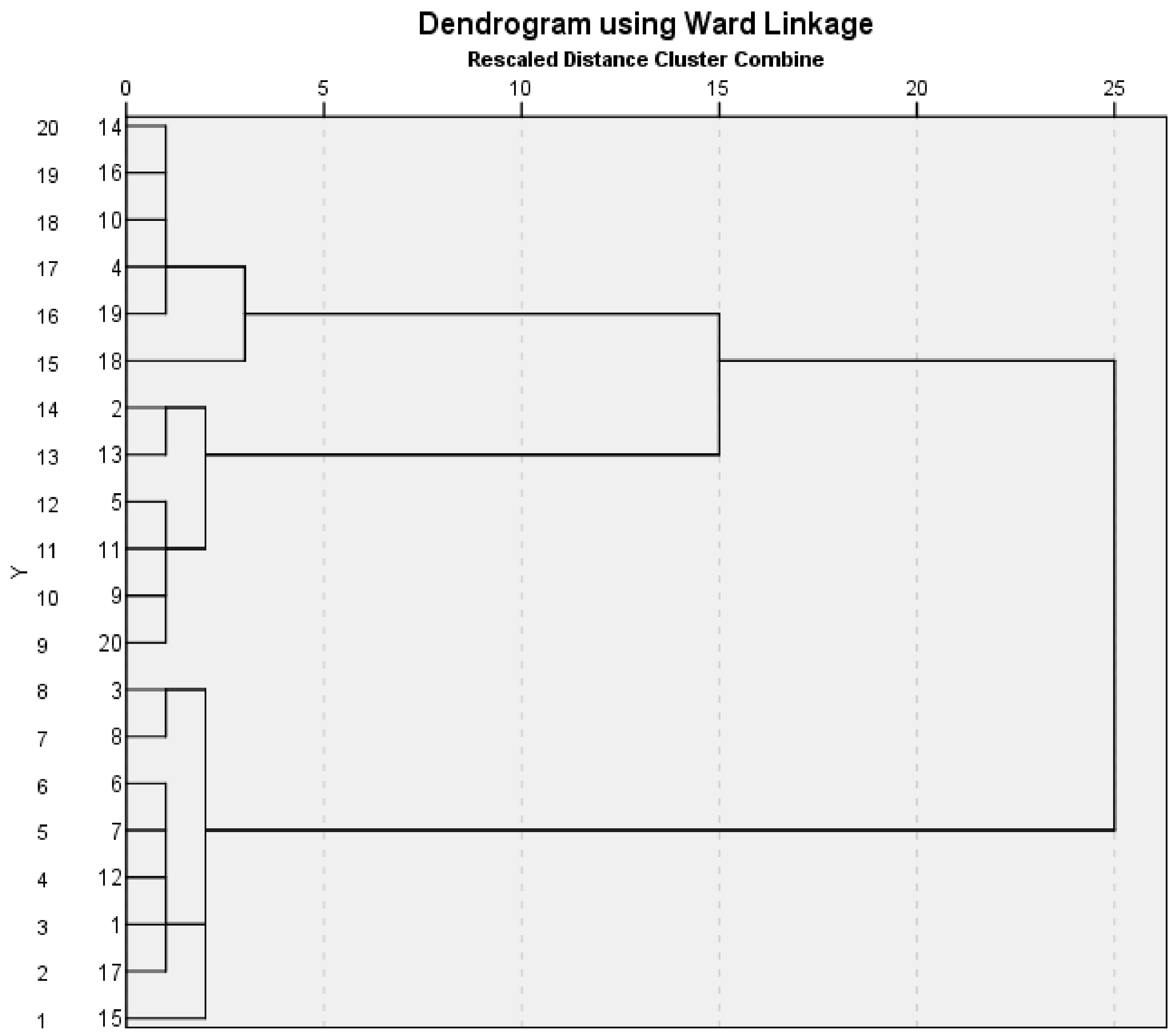
Cluster Membership			
Case	**4 Clusters**	**3 Clusters**	**2 Clusters**
1	1	1	1
2	2	2	2
3	1	1	1
4	3	3	2
5	2	2	2
6	1	1	1
7	1	1	1
8	1	1	1
9	2	2	2
10	3	3	2
11	2	2	2
12	1	1	1
13	2	2	2
14	3	3	2
15	1	1	1
16	3	3	2
17	1	1	1
18	4	3	2
19	3	3	2
20	2	2	2

If we have two clusters then **1, 3, 6, 7, 8, 12, 15**, and **17** are 1st cluster and the rest are in 2nd cluster.

If we have three clusters then **1, 3, 6, 7, 8, 12, 15, 17** are in 1st cluster, **2, 5, 9, 11, 13,** and 20 are in 2nd cluster rest are in 3rd cluster.



Case
18 19 16 14 10 4 20 9 11 5 13 2 8 3 15 17 12 7 6 1
Number of clusters
0
5
10
15
20



From the dendogram we can easily see that there are three clusters of our data:

14, 16, 10, 4, 19, and **18** are in the first cluster.

2, 13, 5, 11, 9, and **20** are in the second cluster

3, 8, 7, 12, 1, 17, and **15** are in third cluster