HARISH SAIFI
GL6134 23DSMSA112

LAB 3 – MULTIDIMENSIONAL SCALING

Objective

Find the distance matrix among all the objects. It can be obtained by simply ranking of distances between an object and all other objects by a consumer. This matrix can be obtained by providing the consumer a card containing pair of objects written on it, and the candidate needs to specify a number indicating the difference between the two objects on any numerical scale which can represent distance between the two objects. This process is repeated for all pairs of brands being included in the study. In this process, no attributes are identified on which the consumer is asked to decide on the difference. The distance measure so obtained for all the pairs of objects can be compiled into a matrix as shown in Table. This distance matrix serves the input data for multidimensional scaling.

Data Overview

	Alto	Estilo	WagonR	Swift	Santro	i10	FordFigo	TataIndica
Alto	0	1	3	7	4	2	4	1
Estilo	1	0	4	5	6	1	1	5
WagonR	3	4	0	2	1	5	6	7
Swift	7	5	2	0	1	5	7	6
Santro	4	6	1	1	0	4	5	4
i10	2	1	5	5	4	0	1	6
FordFigo	4	1	6	7	5	1	0	3
TataIndica	1	5	7	6	4	6	3	0

The data contains feedback about different cars and their similarity ratings. The diagonal elements are set to 0 because the rating of car itself does not make any sense, so the diagonal is kept empty for all the cars. The matrix/dataset shows that upper triangular matrix is equal to the lower triangular matrix which corresponds the same thing or the rating between the pairs inversely are also same.

3-dimensional Iterations for stress values and RSquared and their Coordinates.

Young's S-stress formula 1 is used.

Iteration	S-stress	Improvement
1	.14535	
2	.12004	.02531
3	.11372	.00632
4	.11188	.00184
5	.11126	.00062

Iterations stopped because S-stress improvement is less than .001000

Stress Values and Iterations:

- The stress values are a measure of how well the distances between points in the low-dimensional space represent the original dissimilarities.
- The stress begins at **0.14535** in the first iteration and decreases to **0.11126** in the fifth iteration.
- The iteration process stopped after 5 steps since the improvement was less than the threshold of **0.001.**

Stress and squared correlation (RSQ) in distances
For matrix

Stress = **.07911** RSQ = **.92211**

Stress and R-Squared (RSQ):

- Final **stress** = **0.07911,** which indicates a good fit (lower values indicate a better fit).
- **RSQ** = **0.92211,** meaning that 92.2% of the variance in the original dissimilarities is explained by the 3D solution.

Configuration derived in 3 dimensions(Spacious(1), Fuel Economy(2) and Stylish(3))

Stimulus Coordinates

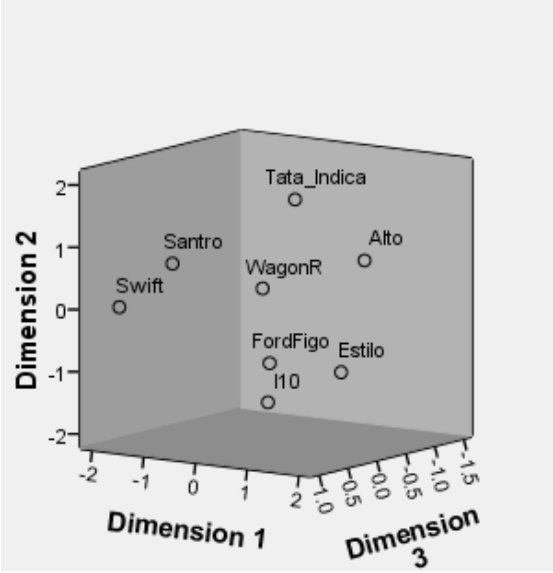
Stimulus Number	Stimulus Name	Dimension 1	2	3
1	Alto	0.8774	0.6086	-0.9932
2	Estilo	0.9917	-1.0586	-0.4867
3	WagonR	-1.3459	-0.1183	-1.2193
4	Swift	-1.8536	-0.0029	0.8010

5	Santro	-1.4590	0.6055	0.2352
6	I10	0.6751	-1.3468	0.4928
7	FordFigo	1.2702	-0.5423	0.9944
8	Tata_Ind	0.8441	1.8548	0.1759

The configuration in 3 dimensions shows where each car (stimulus) is in geometric space. For example:

- Alto is located at (0.8774, 0.6086, -0.9932).
- Ford Figo is located at (1.2702, -0.5423, 0.9944).

After plotting the graph of above values, we can get the overview of the similarity and dissimilarity between all the cars.



The 3D graph appears very clumsy to visualize so we will derive it into 2D by using Euclidean distance for the attributes.

2-dimensional Iterations for stress values and RSquared and their Coordinates(Spacious(1) and Fuel Economy(2))

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

Young's S-stress formula 1 is used.

Iteration	S-stress	Improvement
1	.22054	
2	.19234	.02820
3	.17623	.01611
4	.16411	.01211
5	.15461	.00950
6	.14791	.00670
7	.14367	.00424
8	.14159	.00208
9	.14139	.00020

Iterations stopped because S-stress improvement is less than .001000

Stress Values and Iterations:

- The stress begins at **0.22054** in the first iteration and drops to **0.14139**

by the ninth iteration.

- The improvement became less than **0.001,** so the process was stopped after 9 iterations.

Stress and squared correlation (RSQ) in distances

RSQ values are the proportion of variance of the scaled data (disparities) in t he partition (row, matrix, or entire data) whichis accounted for by their corre sponding distances.

For matrix with just two dimensions included for MDS scaling(i.e., Spacious and Fuel Economy)

Stress = .16611 RSQ = .87594

Configuration derived in 2 dimensions

		Stimulus Coordinates Dimension	
Stimulus Number	Stimulus Name	1	2
1	Alto	1.0933	.7542
2	Estilo	0.9651	-.7312
3	WagonR	-1.4408	.1261
4	Swift	-1.4492	.2257
5	Santro	-1.4133	.2052
6	I10	.9121	-.8085
7	FordFigo	.6121	-1.3142
8	Tata_Ind	.7206	1.5429

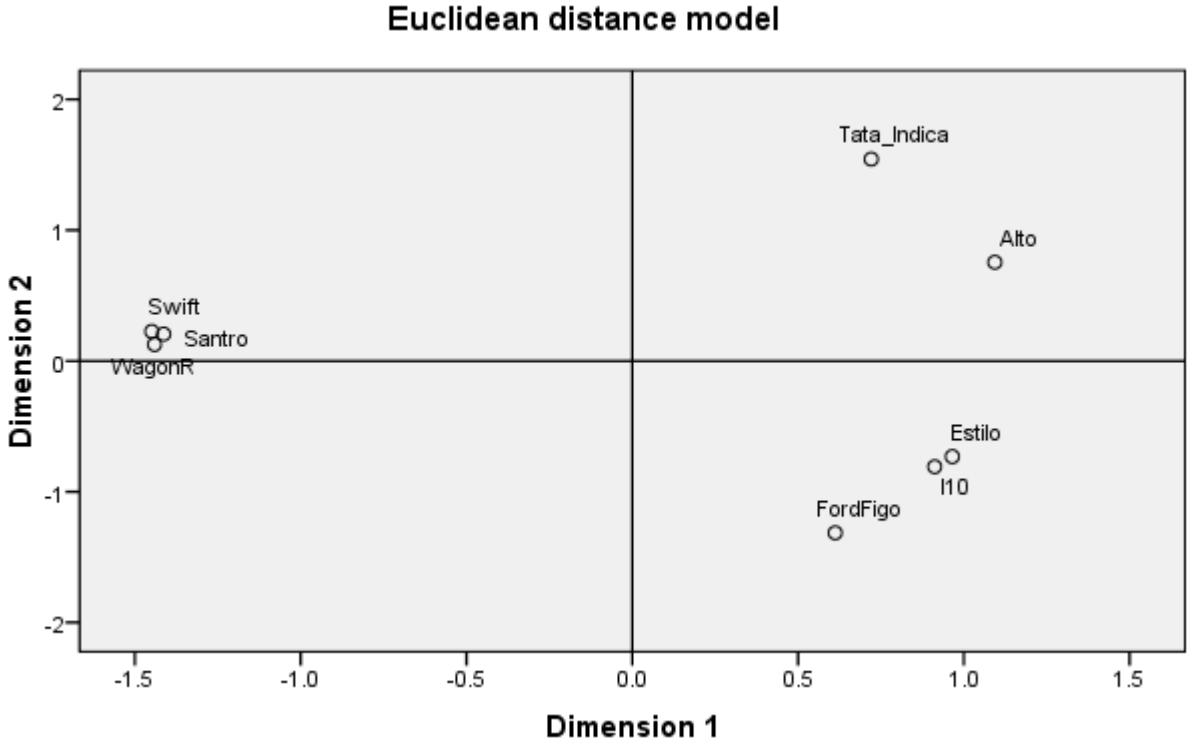
On comparing the result for MDS for 2D (Spacious and Fuel Economy) and 3D(Spacious, Fuel Economy and Stylish)
We find that 3D scaling is showing better results as compared to 2D.
For 3D we are getting about 92.2% of RSquare value (retention of information after MDS scaling) being explained and stress value of 0.0079 while for 2D we are getting 87.6% of RSquare value and stress value close to 0.17 which in respect to 3D (Spacious, Fuel Economy and Stylish) MDS is poorer.

Perception Mapping

Below is the perception map of the data provided for the customers feedback on the various dimension (Spacious, Fuel Economy and Stylish) regarding different brands of car.

The Euclidean distance is used to measure the similarity between all cars brand after scaling the provided multidimension data.

Based on the variable being studied we can get the perception about the consumer on the cars and the attributes they percept about different models/brands of car.
We can see that Swift+Santro+WagonR are spacious cars as compared to other, and Tata_Indica+Alto are top in Fuel Economy and Estilo+i10+FordFigo are considered stylish.
While choosing which axis corresponds to which attribute is subjective depends upon the researcher's prior knowledge and familiarity about the problem statement.



Here the –x(left) axis can be labeled for the more spacious cars and +x(right) axis as the less spacious cars.
While +y(top) can be labeled as more Fuel Economy cars and –y(bottom) axis as the less Fuel Economy cars.

After comparing the Perception graph following conclusion can be drawn:

- Swift¥Santro¥WagonR are more spacious than other cars but have less fuel economy and are less stylish.
- Tata Indica¥Alto have more Fuel Economy than others but are less spacious and stylish.
- Estilo¥i10¥FordFigo are more stylish but have less fuel economy and less spacious.

Although it feels very strange to put more spacious label to –x axis as it suggests less spacious but in real life after looking at the car's models and by prior knowledge, they appear more spacious so putting Spacious for –x axis is more logical than it appeals.