NON METRIC MDS (Shepard, Kruskal,64)

Monotonic relationship between distance and dissimilarity (prossimity measured on ordinal scale):

$$d_{ij}=f(\delta_{ij})$$

NON METRIC MDS

The order of distances in the configuration must reflect the order of dissimilarity indices

After calculating the Euclidean distances between pairs of points:

$$d_{ij} = [\sum_{s=1}^{k} (x_{is} - x_{js})^2]^{1/2}$$

One of the following conditions must be at least respected:

$\delta_{ij} < \delta_{lh} \Rightarrow d_{ij} \leq d_{lh}$ ⟶ Weak Monotonicity

$\delta_{ij} < \delta_{lh} \Rightarrow d_{ij} < d_{lh}$ ⟶ Strong Monotonicity

STEPWISE PROCEDURE

1. create an initial configuration, random or derived from a previous metric MDS
2. calculate the standardized distances (Euclidean)
3. compare the order of dissimilarities and of Euclidean distances calculated; if the monotonic order is violated, it is necessary to calculate the disparity $\hat{d}_{ij}$ (close to the distances and weakly monotonous with respect to dissimilarities)
4. estimate a goodness of fit measure (STRESS)

Comparison of distance and dissimilarity order

Building a chart of dissimilarities δ_{ij} with distances d_{ij}, it can occur:

A perfect monotonic correlation if: the smallest value of dissimilarity corresponds the smallest distance value, the second smallest of dissimilarity corresponds the second smallest distance ...and so on...

We use the Spearman's rank correlation

DISPARITY (Kruskal,1964)

The correlation between distances and dissimilarities is often low, it can be useful to calculate increasing monotonic transformations of dissimilarities $F'(\delta_{ij})$ (e.g. linear, logarithmic, exponential), defined "disparities":

$$F'(\delta_{ij}) = \hat{d}_{ij}$$

F is chosen respecting the following condition

$$\delta_{ij} < \delta_{lh} \Rightarrow \hat{d}_{ij} \leq \hat{d}_{lh}$$

DISPARITY

The disparities are defined with a criterion of weaker monotonicity than original dissimilarities, in fact, their estimation is based on the average of distances of groups of points that do not satisfy the monotonicity constraint with the original dissimilarities.

The estimation is done by a method called “monotone regression"

Let A, B, C be three brands, if the pairwise order, based on dissimilarity, is

δ_{AB}	Less dissimilar (1)
δ_{AC}	More dissimilar (3)
δ_{BC}	Quite dissimilar (2)

The distance between A and B has rank 1, that between B and C has rank 2 and the distance between A and C has rank 3.

Dissimilarity matrix

	A	B	C	D	E
A	0				
B	1	0			
C	2	4	0		
D	5	8	3	0	
E	7	9	6	10	0

Distance matrix

	A	B	C	D	E
A	0				
B	0.15	0			
C	0.28	0.42	0		
D	0.27	0.12	0.55	0	
E	2.52	2.67	2.24	2.79	0

Dissimilarities

1	2	3	4	5	6	7	8	9	10
0.15	0.28	0.55	0.42	0.27	2.24	2.52	0.12	2.67	2.79

Distances

Substituting values which do not respect monotonicity condition with their mean:

Dissimilarities

1	2	3	4	5	6	7	8	9	10
0.15	0.28	0.55	0.42	0.27	2.24	2.52	0.12	2.67	2.79

Distances

1	2	3	4	5	6	7	8	9	10
0.15	0.28	0.41	0.41	0.41	2.24	1.32	1.32	2.67	2.79

This sequence is not monotonic with respect to dissimilarities, the algorithm does not stop

1	2	3	4	5	6	7	8	9	10
0.15	0.28	0.41	0.41	0.41	1.63	1.63	1.63	2.67	2.79

These values are disparities!

Configuration check

To verify the differences between disparities and distances we can use the S index, which assumes value 0 when the distances are perfectly disparities (dissimilarity) and that increases with increasing differences between the two quantities:

$$\boxed{S = \sum_{i<j} \sum_{j=2}^{n} (d_{ij} - f(\delta_{ij}))^2 = \sum_{i<j} \sum_{j=2}^{n} (d_{ij} - \hat{d}_{ij})^2}$$

The objective is to find a configuration that minimizes S for a given number of dimensions!

STRESS

If the configuration is stretched or contracted (multiplying or dividing the coordinates by a k factor) S is influenced, also if the relationship between distance and dissimilarity/disparity does not change

For this reason other objective functions ("STRESS") are then selected

STRESS (Kruskal,1964)

STRESS (STandardised REsidual Sum of Squares) is given by:

$$\text{STRESS} = \left\{ \frac{\sum_{i=1}^{n-1} \sum_{j=2}^{n} (d_{ij} - \hat{d}_{ij})^2}{\sum_{i=1}^{n-1} \sum_{j=2}^{n} (d_{ij})^2} \right\}^{1/2}$$ TO MINIMIZE !!!

STRESS=0	perfect monotonicity
STRESS < =0,05	excellent fit
0,05 < STRESS <0,10	good fit
0,10 < STRESS < 0,20	fairly good fit
STRESS > = 0,20	insufficient fit

S-STRESS (Takane et al.,1977)

STRESS decreases with an increasing number of dimensions k [k<=(n-1)/4]

$$\text{S-STRESS} = \left\{ \frac{\sum_{i=1}^{n-1} \sum_{j=2}^{n} (d^2_{ij} - \hat{d}^2_{ij})^2}{\sum_{i=1}^{n-1} \sum_{j=2}^{n} (d^2_{ij})^2} \right\}^{1/2}$$

Where distances and disparities are squared!

The points on the map are reallocated to minimize the S-STRESS

RSQ-STRESS (SPSS)

Another fitting measure is the square of the correlation coefficient between distances and disparities

$$\mathrm{RSQ} = \left\{ \frac{\sum_{i=1}^{n-1} \sum_{j=2}^{n} (d_{ij} - d_{..})(\hat{d}_{ij} - \hat{d}_{..})}{\left[\sum_{i=1}^{n-1} \sum_{j=2}^{n} (d_{ij} - d_{..})^2 \sum_{i=1}^{n-1} \sum_{j=2}^{n} (\hat{d}_{ij} - \hat{d}_{..})^2 \right]^{1/2}} \right\}^2$$

If STRESS, S-STRESS and RSQ do not give an overall judgment of fitting, a graphical representation can be useful (Shepard) for which the behavior of disparities, distances and dissimilarities can be represented by a straight line.

Objective: minimum STRESS

In general, the algorithms start from an initial random configuration and move* more or less the points in order to reduce iteratively the STRESS and repeat the procedure when it is no longer possible to obtain an improvement of the STRESS (S-STRESS) value. In particular, the algorithm stops if…

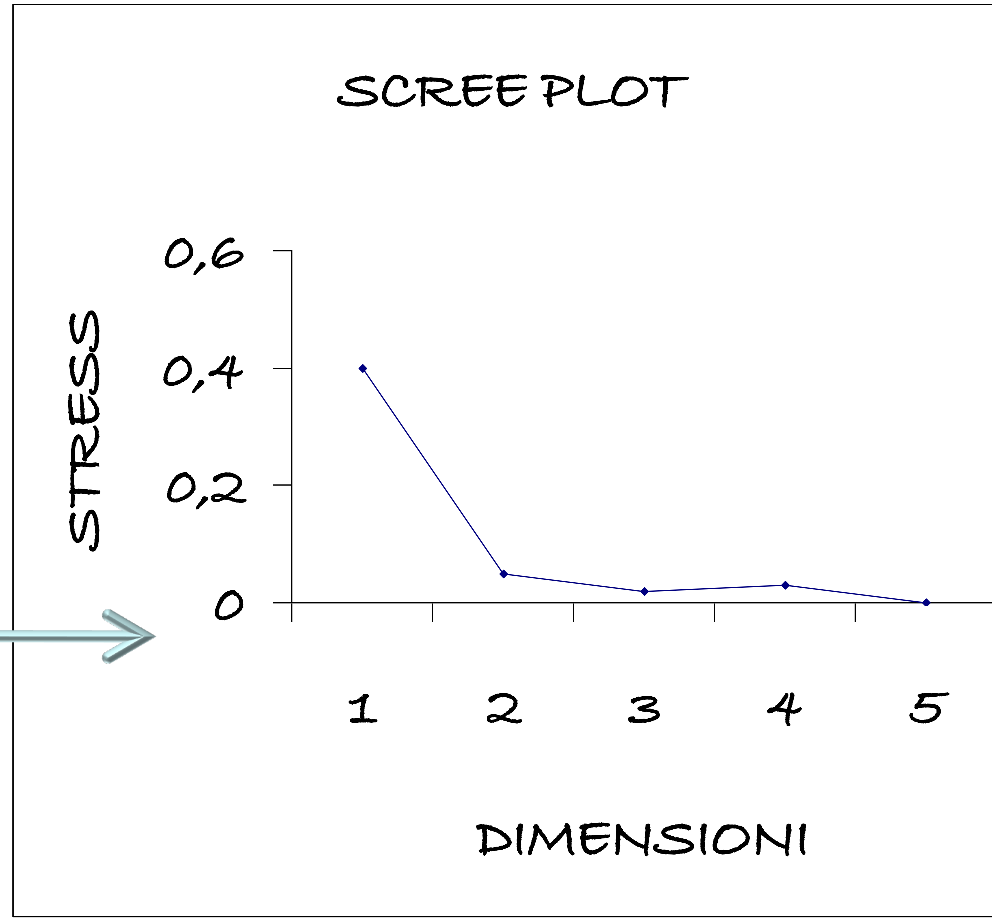
*xi and xj points are approached if $f(\delta_{ij}) < d_{ij}$ and outdistanced in the opposite case

STOPPING RULE:

- The fitting measure states a perfect fitting
- The improvement of the fitting compared with the previous iteration is < threshold
- The maximum number of iterations has been reached
- if none of the rules is respected, they calculate the coordinates and again from step II

TO MINIMIZE THE STRESS

For different dimensions, we get different levels of STRESS.
After calculating the value of the STRESS for each k value, we can choose the optimal number of dimensions



POSITIONING_ANALYSIS OF SCOOTER

INTERVIEW WITH 300 BUYERS RANDOMLY SELECTED

1) 8 scooter models to divide in 2 groups

Dissimilarities = proportions of cases in which pairs of models were assigned into different groups

2) Survey on a interval scale 1 - 5 about 13 characteristics (performance, handling, shutter) for each scooter purchased

Modal values of satisfaction judgemnts expressed by buyers

	Aprilia	Gilera	Malaguti	MBK	Peugeot	Piaggio	Suzuki	Yamaha
Performance	1	2	2	3	4	3	3	3
Road holding	1	4	2	3	4	3	2	2
Brakes	4	4	4	4	5	3	4	4
Sprint	1	3	3	4	4	3	2	3
Maneuverability	4	4	4	4	4	4	4	4
Equipment	3	4	4	3	5	5	2	3
Comfort	4	4	3	3	3	3	4	3
Lighting	2	2	3	2	4	2	2	3
Mirrors	5	3	2	2	5	3	3	2
Noisiness	4	3	4	3	2	3	3	3
Polluting emissions	2	2	2	3	2	2	2	3
Antitheft	1	2	1	4	4	1	1	4
Maintenance	3	3	2	2	2	3	2	2

DISSIMILARITIES

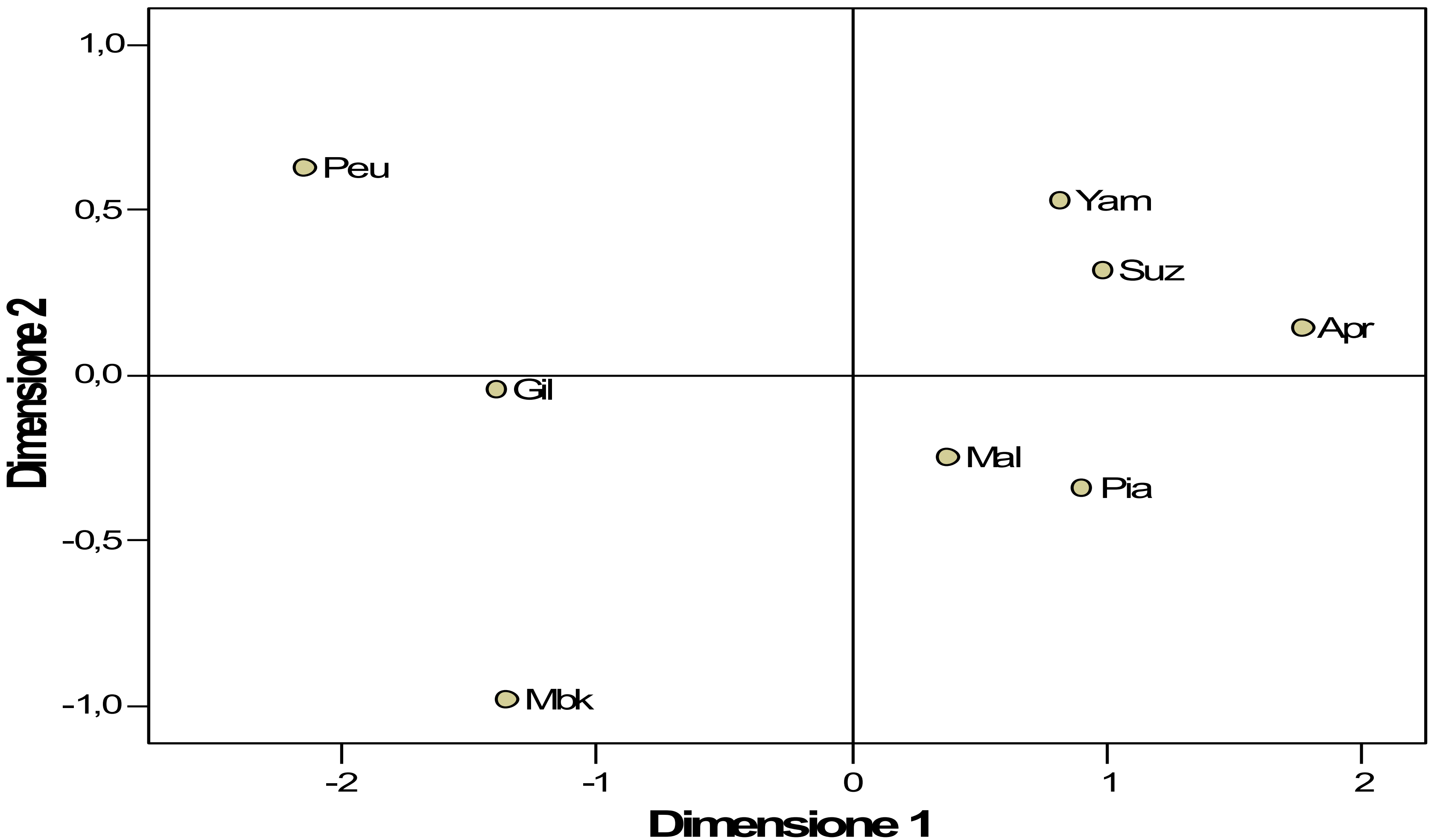
	Apr	Gil	Mal	Mbk	Peu	Pia	Suz	Yam
Apr	0,00							
Gil	0,72	0,00						
Mal	0,35	0,40	0,00					
Mbk	0,73	0,18	0,38	0,00				
Peu	0,85	0,28	0,62	0,38	0,00			
Pia	0,28	0,56	0,20	0,44	0,70	0,00		
Suz	0,26	0,48	0,24	0,58	0,72	0,30	0,00	
Yam	0,30	0,42	0,22	0,62	0,66	0,24	0,26	0,00

POSITIONING ANALYSIS OF SCOOTER

		DIMENSION 1	DIMENSION 2
1	Apr	1,7722	,1413
2	Gil	-1,3884	-,0452
3	Mal	,3828	-,2474
4	Mbk	-1,3462	-,9814
5	Peu	-2,1359	,6289
6	Pia	,9052	-,3426
7	Suz	,9873	,3189
8	Yam	,8231	,5275

STRESS = 0,060
RSQ = 0,9838

Perceptive Map



POSITIONING_ANALYSIS OF SCOOTER

From right to left scooter brands take positions corresponding to increasing modal values for performance, holding and sprint (see April and Gilera);

DIMENSION 1 CAN BE RELATED TO SPEED

From top to bottom brands show decreasing modal values for mirror and lighting

DIMENSION 2 CAN BE LINKED TO VISIBILITY