MDS for individual differences

The models analyzed are based on a single proximity matrix obtained as synthesis of all the individual evaluations expressed by the interviewees.

This is equivalent to the hypothesis that the differences between the subjects are random.

With the MDS for individual differences we treat these differences in a systematic way: the distances between points in a space with common to all subjects dimensions may be systematically different, due to different weights given by each subject to the dimensions (attributes)

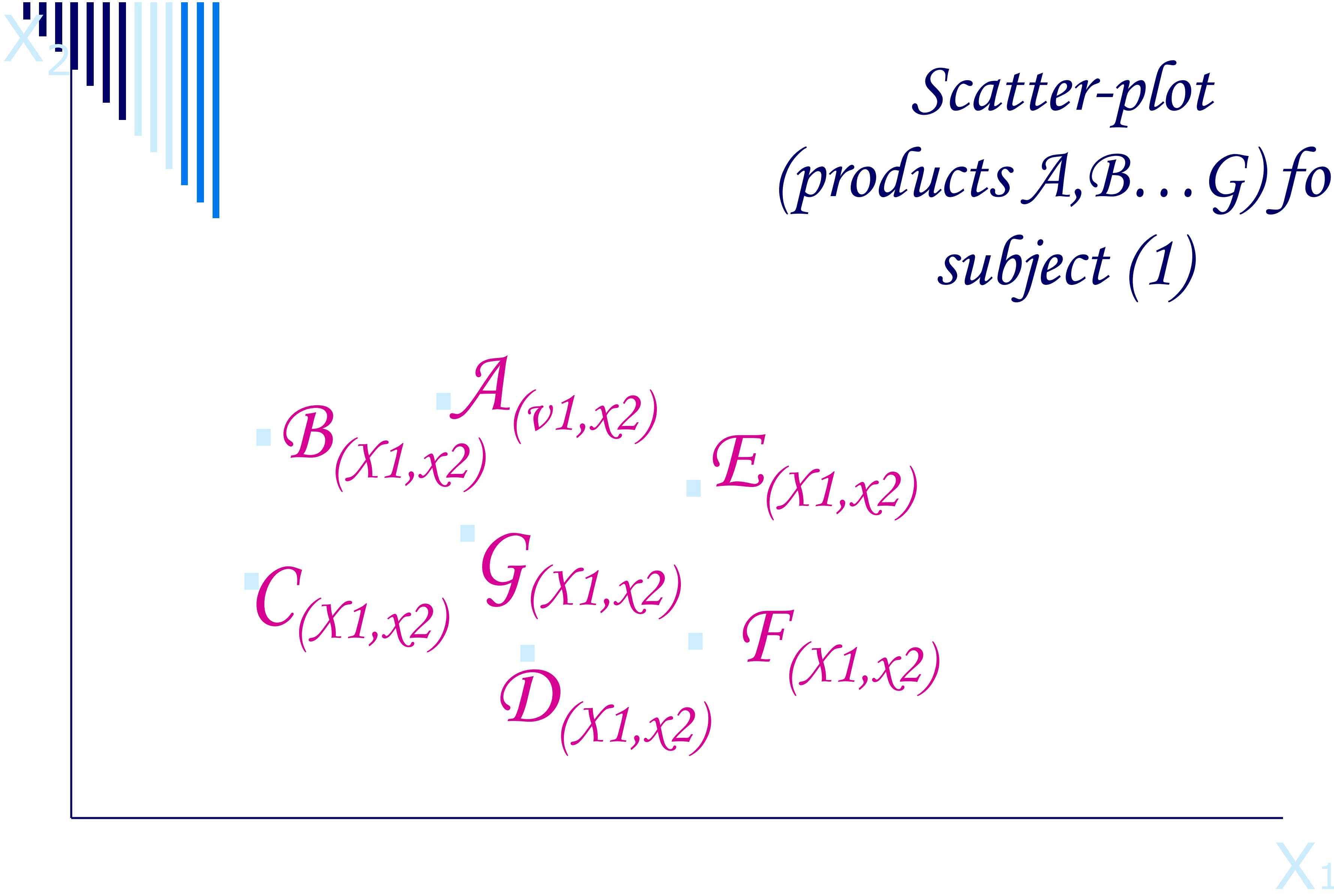
MDS for individual differences

The objective is to perceive some change in market trend

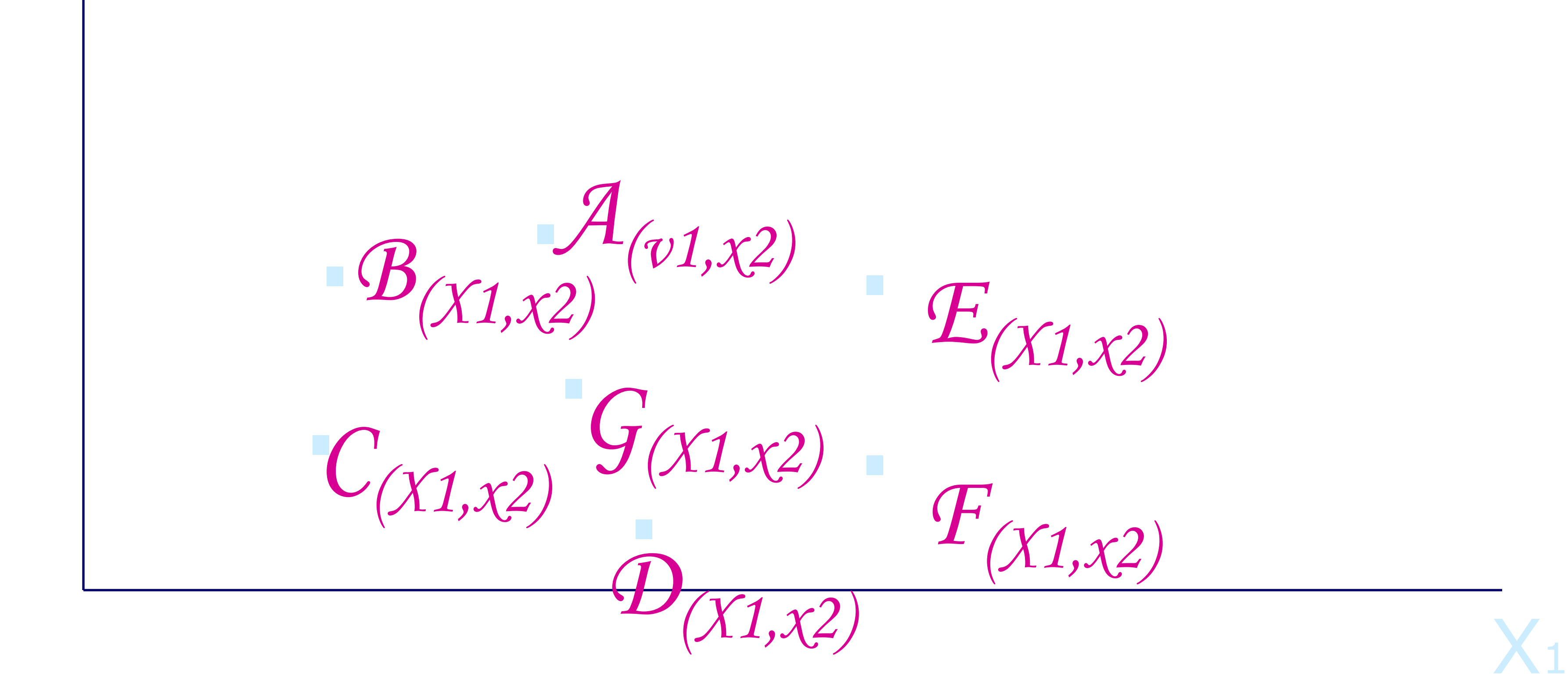
We interview a small number of people, able to receive any changes that will occur on the market

Typically, market analysts or consumption psychologists are chosen!

Scatter-plot (products A,B… G) for subject (1)



Scatter-plot (products A,B…G) for subject (2)



MDS for individual differences (Carrol, Chang)

PROXIMITY MEASURES (METRIC OR NON METRIC)

Basic assumptions: the geometric space (set of common dimensions k) shared by all individuals and evaluation differences (distances) are due to the weight given to the dimensions of the common area

A dissimilarity matrix for each interviewed subject

Weighted Euclidean Model (WEM)

Δ = proximity matrix of q subjects (n x n x q)

It is a summary of the evaluations of q subjects (we have q matrices)

X_t = coordinate matrix given by the t-th subject with respect to all dimensions with generic element:

x_{ist} = coordinate of i-th element given on s by t

Weighted Euclidean Model (WEM)

The generic coordinate of X_t can be expressed as:

$$x_{ist} = x_{is} w_t$$

where w_t is the weight given to s by t

The distance between two evaluations on brands/products (i and j) given by t on k dimensions [s=1,..k] is :

$$\delta_{ijt} = d_{ijt} = \left\{ \sum_{s=1}^{k} (x_{ist} - x_{jst})^2 \right\}^{1/2} = \left\{ \sum_{s=1}^{k} w_{st}^2 (x_{is} - x_{js})^2 \right\}^{1/2}$$

Weighted Euclidean Model (WEM)

In matrix notation

$$\delta_{ijt} = d_{ijt} = \left\{ (x_i - x_j)' W_t (x_i - x_j) \right\}^{1/2}$$

OUTPUT WEM

Includes

1) X (nxk) = coordinate matrix for all subjects in a common map,
2) W(kxq) = [w1,w2,…wt, …wq] weight matrix of q subjects for k dimensions;
3) Goodness of fit measure

WEM

We estimate the coordinates in the same way of metric MDS

$$Q = X W_t X' \qquad t=1,\dots,q$$

Q (symmetric) can be obtained after the estimation of Wt:

$$Wt = AMA'$$

This approach deals with Torgerson model, where the proximity is equal to distance

Modello Euclideo Ponderato (WEM)

Empirically, we can not assume the equality between distance and proximity, but we can impose a function:

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

All popular algorithms start from an initial configuration (or by Torgerson model), for which, known coordinates and distances, it is possible to estimate the weights with non-linear regression. Such weights allow to determine a new configuration (new coordinates) that minimizes the difference between distances and proximity (defined by the value of ε_{ij}) …

Modello Euclideo Ponderato (WEM)

The algorithms mostly used are INDSCAL (which maximizes RSQ) and ALSCAL (which minimizes the S-STRESS)

They are used for both metric and non-metric data

N.B. There is no orthogonal rotation of obtained dimensions!

INDSCAL(metric)

The objective is to maximize RSQ, the square correlation coefficient between distances and dissimilarities

$$\mathrm{RSQ} = \left\{ \frac{1}{q} \sum_{t=1}^{q} \frac{\sum_{i=1}^{n-1} \sum_{j=2}^{n} (d_{ijt} - d_{..t})(f(\delta_{ijt}) - f(\delta_{..t}))}{\left[\sum_{i=1}^{n-1} \sum_{j=2}^{n} (d_{ijt} - d_{..t})^2 \sum_{i=1}^{n-1} \sum_{j=2}^{n} (f(\delta_{ijt}) - f(\delta_{..t}))^2 \right]^{1/2}} \right\}^2$$

$$d_{ijt} = \sum_{s=1}^{k} x_{is} x_{js} w_{st}$$

INDSCAL(non metric)

It can use the square correlation coefficient between distances and disparities

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

ALSCAL(metric)

The objective is to minimize

$$\text{S-STRESS} = \left\{ \frac{1}{q} \sum_{t=1}^{q} \frac{\sum_{i=1}^{n-1} \sum_{j=2}^{n} \left(d_{ijt}^{2} - f(\delta_{ijt}^{2})\right)^{2}}{\sum_{i=1}^{n-1} \sum_{j=2}^{n} \left(d_{ijt}^{2}\right)^{2}} \right\}^{1/2}$$

ALSCAL (non metric)

Or minimizes the following, for non metric data

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

ALGORITHM

The procedure works for successive iterations, until reaching a convergence criterion:

For a given configuration X, determine the weights Wt and the coordinates of the system of q equations expressed in matrix form dt = XWtX', by the method of nonlinear least squares.

These weights are used to determine a new configuration that minimizes S-STRESS (or maximizing the RSQ) with respect to that calculated for the previous configuration and iteratively, until it reaches the minimum value (or another stopping rule)

Interpreting the configuration

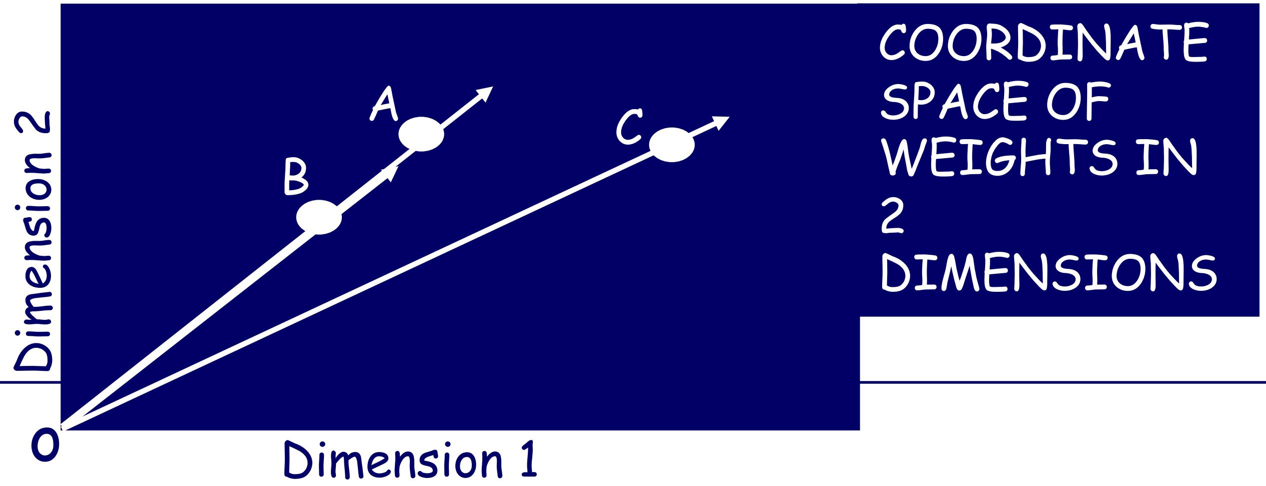
Simultaneous explanation of the maps relating to the common coordinates space (products/brands) and to the common space of weights (subjects)

It is how to evaluate the solutions of a single matrix of a metric MDS, taking into account the weights of the k dimensions given by the subjects

What configurations are more similar to each other?

Evaluating the weight space:

1. Join the points-space coordinates of the weights with the origin (vectors): all points that lie on the same vector have the same relative weights
2. Determine vector module (sum of the weight squares) to check the goodness of fit of the model with respect to dissimilarity indices



Interpreting the configuration

In order to measure the difference between weights we have to estimate the angular separation between the corresponding weight vectors

The smaller the more similar are evaluations assigned to dimensions

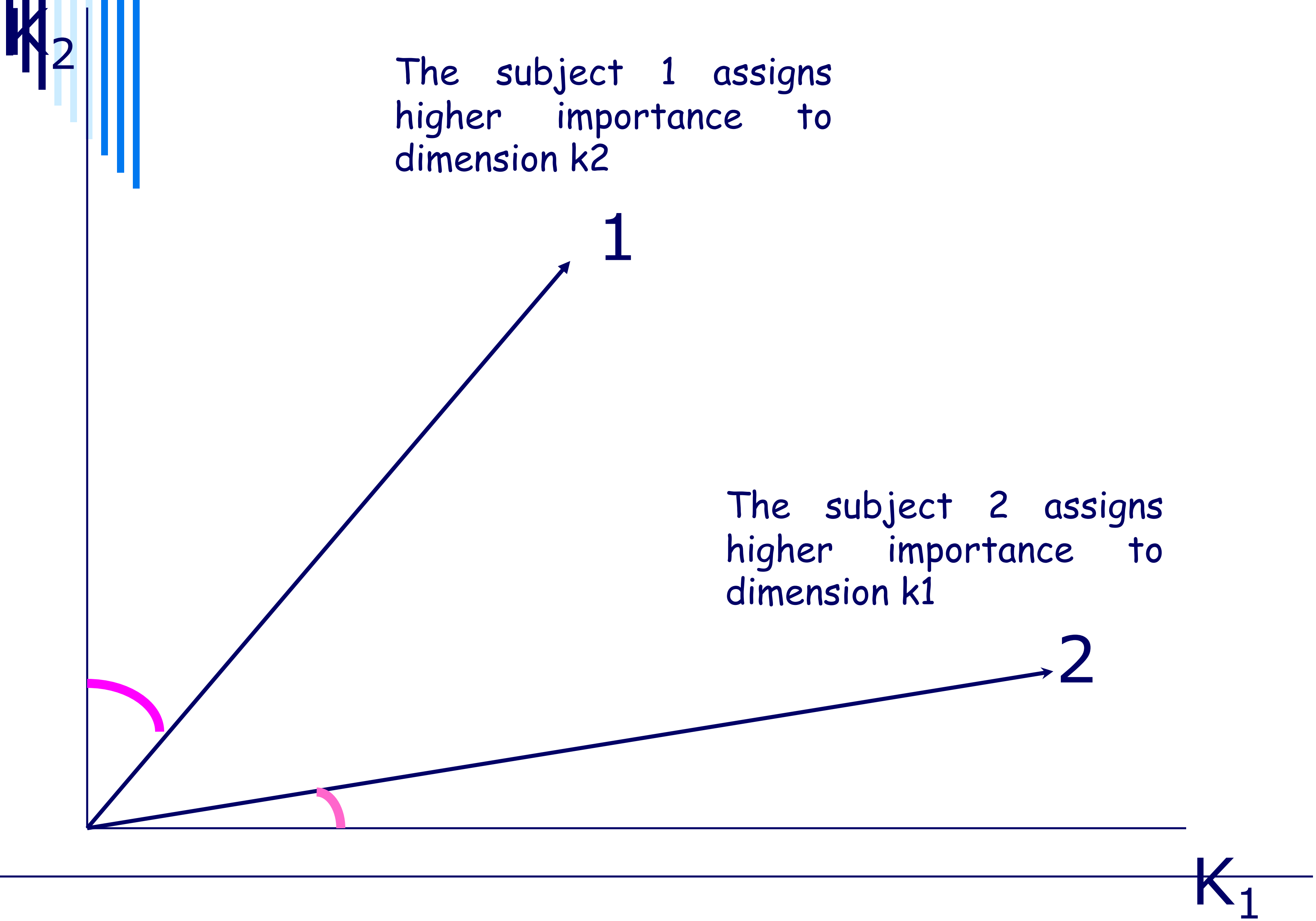
Validity of the comparison: same module (length of the vectors)

Interpreting the configuration

The importance attributed by each subject to a dimension is provided by the value of its coordinate on that dimension:

the higher the value, the greater the contribution of the subject to the determination of that dimension.

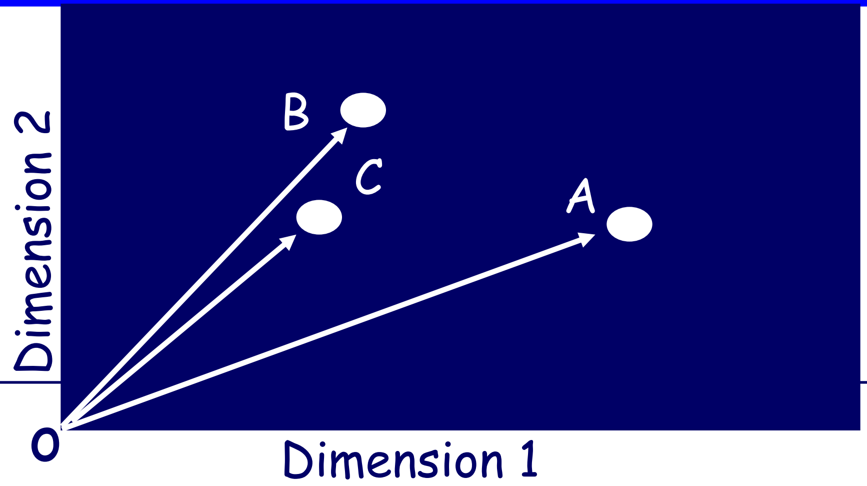
Minor results, therefore, the angle formed by the dimension and by the "weight" vector.


K2
The subject 1 assigns higher importance to dimension k2
1
The subject 2 assigns higher importance to dimension k1
2
K1

Weight matrix for subject A, B and C

	Dimension 1	Dimension 2
A	0,90	0,50
B	0,45	0,75
C	0,30	0,50

EXAMPLE: FOR A THE DIMENSION 1 IS MOST IMPORTANT! FOR B AND C THE DIMENSION 2 IS MORE IMPORTANT BUT C IS LESS SATISFACTORY THAN B! C IS NOT COMPARABLE!



CONDITIONAL PROXIMITY

IMPORTANT: TYPE OF DATA

The scores of two individuals can be equal in terms of subjective evaluations expressed on a scale from 1 to 10, even if they are numerically different

YOU MUST USE THE CONDITIONAL PROXIMITY

CONDITIONAL PROXIMITY

	Dimension 1	Dimension 2
A	0,90	0,50
B	0,45	0,75
C	0,30	0,50

For A dimension1 is more important of 1,8 (0,90/0,50) than dimension 2!

For B dimension 1 is 0,6 (0,45/0,75) times dimension 2

A assigns to dimension 1, relative to dimension 2, a weight 3 times (1,8/0,6) higher than that given by B

EXAMPLE (ordinal data)

GENDER		soccer	volley	ski	tennis	basket	Swim.	baseball
m	soccer	0,00		.	.	.	.	.
m	volley	5,99	0,00	.	.	.	.	.
m	ski	8,80	8,97	0,00	.	.	.	.
m	tennis	8,17	7,06	7,40	0,00	.	.	.
m	basket	5,22	4,47	8,72	8,01	0,00	.	.
m	Swim.	8,93	8,37	6,59	7,16	8,67	0,00	.
m	baseball	6,96	6,54	8,77	7,11	6,70	8,62	0,00
f	soccer	0,00	.	.	.	.	.	.
f	volley	5,35	0,00	.	.	.	.	.
f	ski	8,60	8,91	0,00	.	.	.	.
f	tennis	7,91	6,51	6,81	0,00	.	.	.
f	basket	4,87	4,01	8,52	7,30	0,00	.	.
f	swim	8,59	7,79	5,93	6,62	8,07	0,00	.
f	baseball	5,22	5,47	8,37	6,25	5,09	8,38	0,00

Example

BOTH MALES FEMALES EVALUATE MORE SIMILAR BASKETBALL-VOLLEYBALL AND MORE DISSIMILAR SKI-VOLLEY

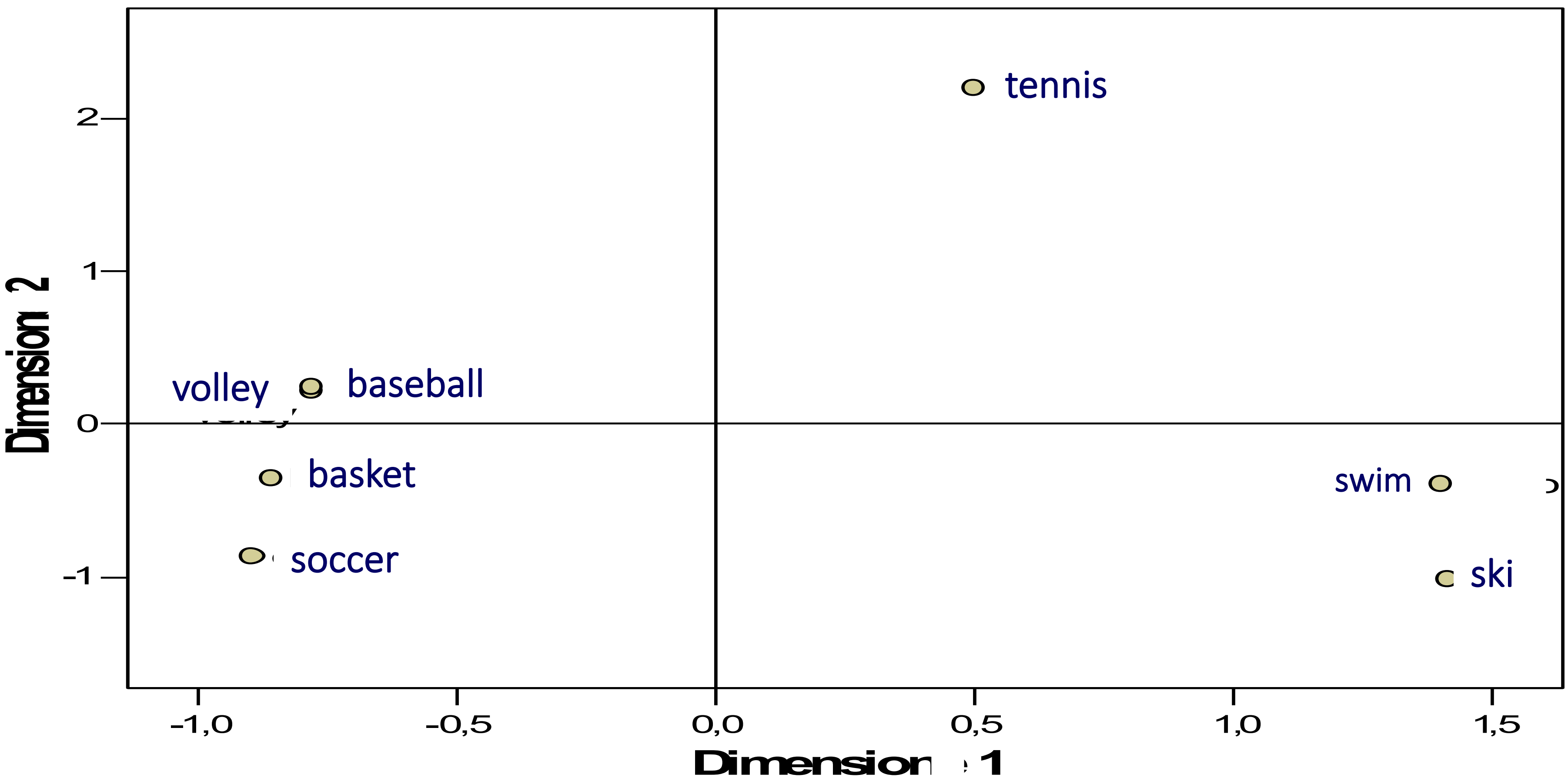
Matrix	Stress	RSQ
m	0,052	0,990
f	0,064	0,984

Minimum STRESS and RSQ close to 1

The values indicate a more satisfactory for males than for females !!!

	Dimension 1	Dimension 2
soccer	-,8973	-,8703
volley	-,7809	,2157
ski	1,4179	-1,0217
tennis	,4978	2,1933
basket	-,8592	-,3625
swimming	1,4052	-,3922
baseball	-,7836	,2377

Perceptive map



Subject Weights

	Dimension		
	Weirdness	1	2
m	,0330	,9640	,2453
f	,0317	,9547	,2689
Overall importance of each dimension:		,9204	,0662

Males have attributed to the dimension 1 a weight four times higher (3.93 = 0.964 / 0.2453) to that given to the dimension 2

The females have provided the dimension 1 a weight 3.55 times greater (0.9547 / 0.2689) than that given to the dimension 2

Comparing the two groups, the males have given a weight to the dimension 1 relative to dimension 2, 1.11 (= 3.93 / 3.55) times higher than that assigned by females!

Subject Weights

		Dimension	
	Weirdness	1	2
1=m	,0330	,9640	,2453
2=f	,0317	,9547	,2689
Overall importance of each dimension:		,9204	,0662

Weirdness index verifies if the weights given are typical:

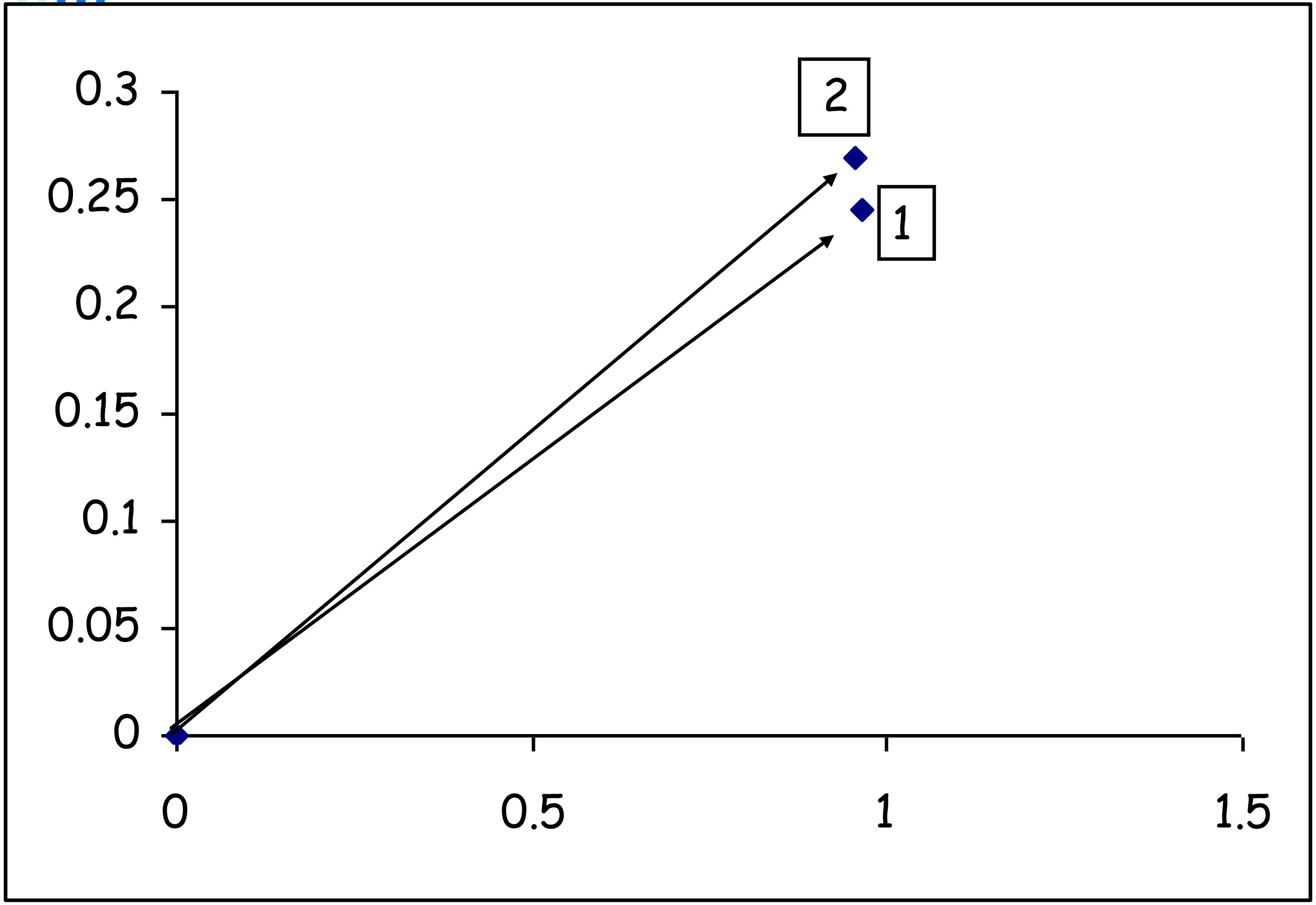
W=0 typical

W=1 atypical

In the table

Wm = 0,033 e Wf = 0,0317 the weights are typical

Weight space



Scatter-plot to fit a linear relationship