

This document is an historical remnant. It belongs to the collection Skeptron Web Archive (included in Donald Broady's archive) that mirrors parts of the public Skeptron web site as it appeared on 31 December 2019, containing material from the research group Sociology of Education and Culture (SEC) and the research programme Digital Literature (DL). The contents and file names are unchanged while character and layout encoding of older pages has been updated for technical reasons. Most links are dead. A number of documents of negligible historical interest as well as the collaborators' personal pages are omitted.

The site's internet address was since Summer 1993 www.nada.kth.se/~broady/ and since 2006 www.skeptron.uu.se/broady/sec/.

V — Structured Data Analysis

Reference:

B. Le Roux, *L'analyse géométrique des données multidimensionnelles*, Dunod 2014, Chapter 9.

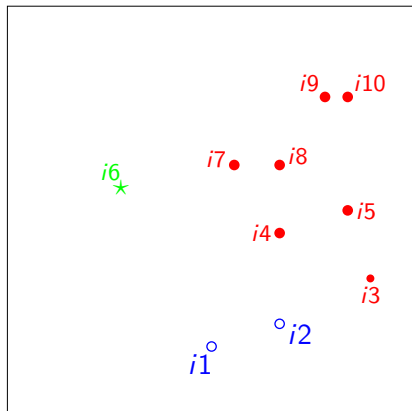
V.1. Partition of a Cloud: Between- and Within-variance

- Subclouds

A: subcloud of 2 points (dipole)
 $\{i1, i2\}$

B: subcloud of 1 point
 $\{i6\}$

C: subcloud of 7 points
 $\{i3, i4, i5, i7, i8, i9, i10\}$



Partition of a cloud into 3 subclouds: A , B and C .

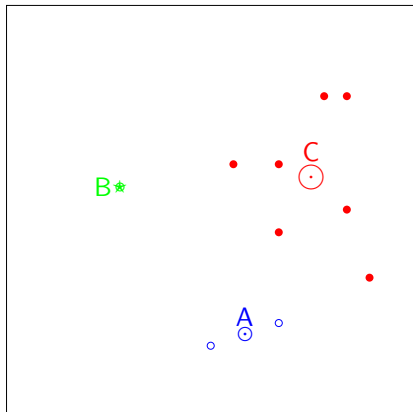
3 mean points A , B , C with weights 2, 1, 7.

By grouping:

— points “average up”

— weights add up

	weights	Coordinates		variances
		x_1	x_2	
A	$n_A = 2$	3	-11	10
B	$n_B = 1$	-8	2	0
C	$n_C = 7$	8.857	2.857	46.57
	$n = 10$	$\bar{x}_1 = 6$	$\bar{x}_2 = 0$	34.6



The mean of the variances of subclouds defines the *within-variance*

Between-cloud

The 3 mean points (A,2), (B,1) et (C,7) define the *between-cloud*.

The between-cloud is a weighted cloud;

- its total weight is $n = 10$;
- its mean point is G;
- its variance, called *between–variance*, is the variance of the mean points

$$\frac{2}{10}(GA)^2 + \frac{1}{10}(GB)^2 + \frac{7}{10}(GC)^2 = 57.4$$

Contributions of a subcloud

The *contribution of a subcloud* is the sum of the contributions of its points.

The *within-contribution* of a subcloud is the product of its weight by its variance and divided by V_{cloud} .

— *Example*: subcloud \mathcal{A}

$$\text{Ctr}_{i_1} = \frac{\frac{1}{10}(GM^{i_1})^2}{92} = \frac{\frac{1}{10} \times 180}{92} = \frac{18}{92}; \quad \text{Ctr}_{i_2} = \frac{\frac{1}{10}(GM^{i_2})^2}{92} = \frac{\frac{1}{10} \times 100}{92} = \frac{10}{92}$$

- contribution of the *subcloud*: $\text{Ctr}_{\mathcal{A}} = \frac{18}{92} + \frac{10}{92} = \frac{28}{92}$
- contribution of the *mean point*: $\text{Ctr}_{\mathcal{A}} = \frac{\frac{2}{10} \times 130}{92} = \frac{26}{92}$
- *within-contribution*: $\frac{\frac{2}{10} \times 10}{92} = \frac{2}{92}$

Huyghens theorem

The contribution of a subcloud is the sum of the contribution of its mean point and of its within-contribution.

Example: Subcloud \mathcal{A}

$$\text{Ctr}_{\mathcal{A}} = \text{Ctr}_{\mathcal{A}} + \text{within-contribution}$$

$$\frac{28}{92} = \frac{26}{92} + \frac{2}{92}$$

Between-within decomposition of variance

	Ctr \times V_{cloud}		subclouds
	mean points	within	
A	26.0	2.0	28
B	20.0	0	20
C	11.4	32.6	44
Total	57.4	34.6	92
Variance	between	within	total

Within-variance

= sum of within-contributions $\times V_{\text{cloud}}$

= weighted mean of variances of subclouds ($\frac{2}{10} \times 10 + 0 + \frac{7}{10} \times 46.6$)

= 34.6

Total variance = between-variance + within-variance

$$\eta^2 = \frac{\text{between-variance}}{\text{total variance}} \quad (\text{eta-square})$$

V.2. Cognitive Tests and Education

Research on metacognitive factors in scientific problem-solving strategies
(P. Rozenchwajg)

Individuals: 12-13-year old seventh graders from two middle schools in the metropolitan Paris area.

Schools a_1 underprivileged socioeconomic environment with 5 boys and 9 girls;

a_2 : medium-level socioeconomic environment with 17 boys and 11 girls.

Variables: 6 cognitive tests

- General intelligence test (g -factor test)
- Numerical test
- Verbal test
- Spatial test
- FDI (“field dependence–independence”) test
- RI (reflective–impulsive) cognitive test

multivariate numerical data (table Students \times Cognitive tests)
two structuring factors: *Gender* and *Status* (socioeconomic environment).

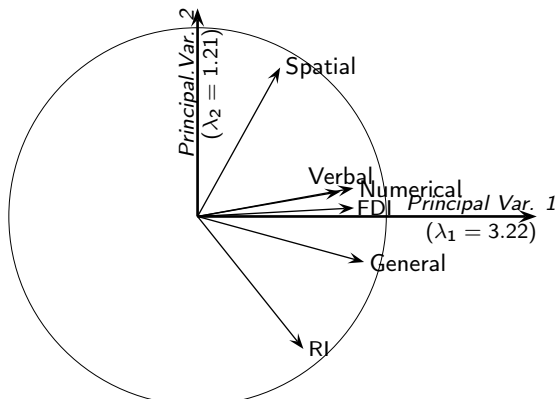
The aim of the study is to figure out to what extent *Status* and *Gender* explain the position of students in the cognitive space.

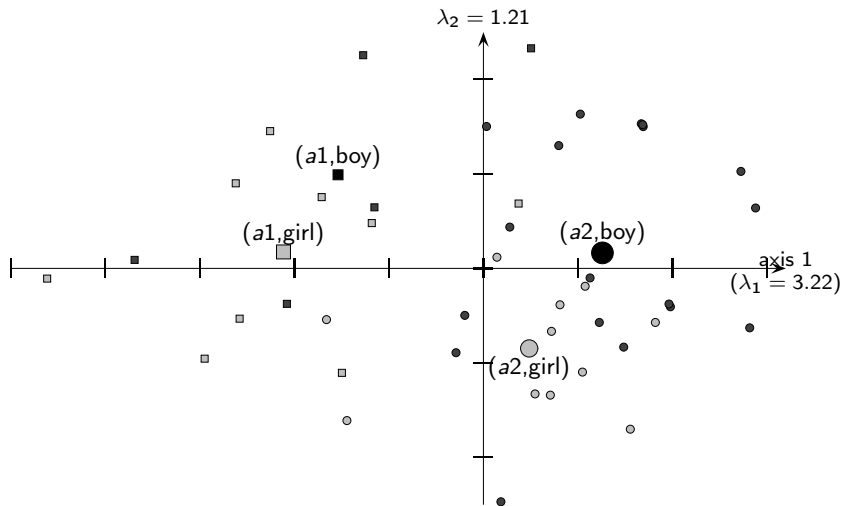
PCA: Construction of the cognitive space
Structured Data Analysis

Cognitive space: PCA

variance	$\lambda_1 = 3.219$	$\lambda_2 = 1.213$	$\lambda_3 = 0.590$	$\lambda_4 = 0.478$	$\lambda_5 = 0.314$	$\lambda_6 = 0.186$
Variance rate	$\tau_1 = .537$	$\tau_2 = .202$	$\tau_3 = .098$	$\tau_4 = .080$	$\tau_5 = .052$	$\tau_6 = .031$

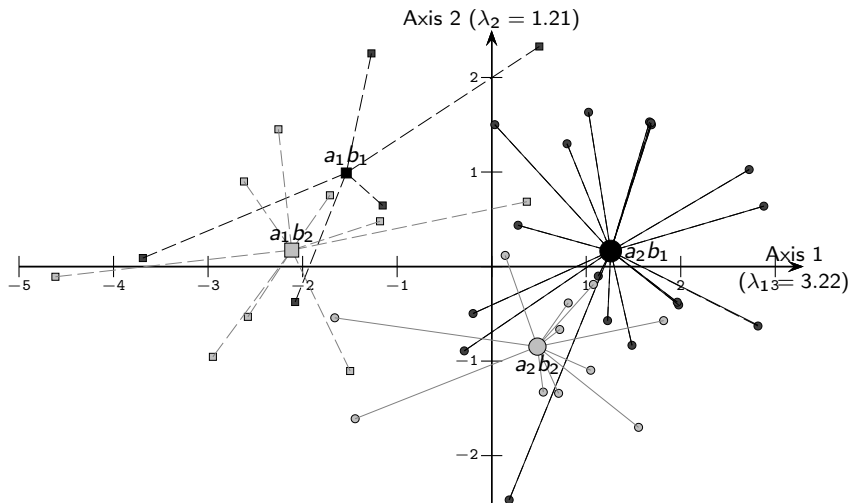
Correlations		<i>General</i>	<i>Numerical</i>	<i>Verbal</i>	<i>Spatial</i>	<i>FDI</i>	<i>RI</i>
Axis 1	$r_{\ell 1}$	0.881	0.825	0.757	0.437	0.828	0.560
Axis 2	$r_{\ell 2}$	-0.241	0.150	0.132	0.788	0.045	-0.701
Plane 1-2	R_{1-2}	0.913	0.838	0.768	0.901	0.829	0.897





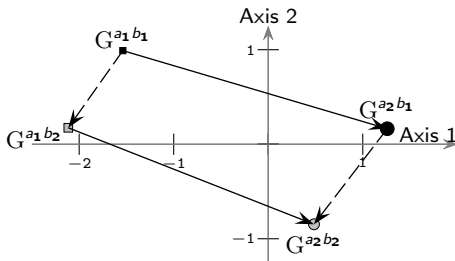
Axis 1 is an axis of *general cognitive abilities*.

Axis 2 is an axis of *processing speed*.



Between-groups cloud ($G^{A \times B}$)

Coordinates	weights	Axis 1	Axis 2
$G^{a_1 b_1}$	5	-1.538	0.990
$G^{a_1 b_2}$	9	-2.115	0.174
$G^{a_2 b_1}$	17	1.258	0.164
$G^{a_2 b_2}$	11	0.484	-0.847
$\text{Var } G^{A \times B}$		1.943	0.322



Overall variance = $3.219 + 1.213 = 4.432$;

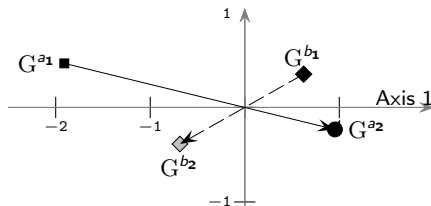
between-groups variance = $1.943 + 0.322 = 2.265$; $\eta^2 = 2.265/4.432 = 0.51$.

Descriptively, the global difference between the four groups is large.

Status main effect and Gender main effect

Coordinates	n	Axis 1	Axis 2
G^{a1}	14	-1.909	0.466
G^{a2}	28	0.954	-0.233
Var G^A	42	1.822	0.108

Coordinates	n	Axis 1	Axis 2
G^{b1}	22	0.623	0.352
G^{b2}	20	-0.686	-0.387
Var G^B	42	0.427	0.136

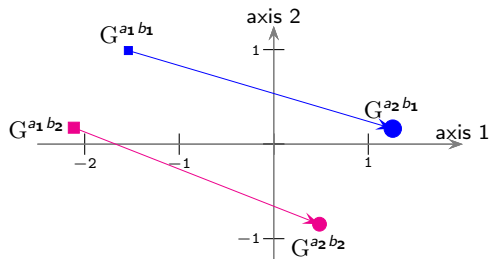


between-*Status* variance: $1.822 + 0.108 = 1.930$, hence 85% of the variance of the *Status* × *Gender* cloud; $\eta^2 = 0.44$ to 0.44 (quite a large value).

between-*Gender* variance: $0.427 + 0.136 = 0.563$, hence 25% of the variance of the *Status* × *Gender* cloud and $\eta^2 = 0.13$ (a large value).

Descriptively, the difference between the two socioeconomic statuses and that between boys and girls are large.

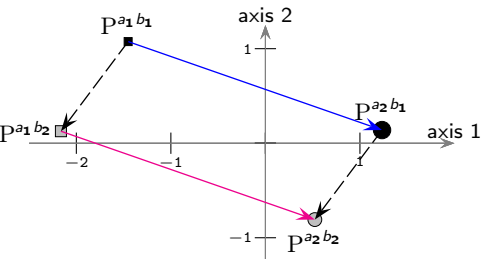
Effect of Status within–Gender



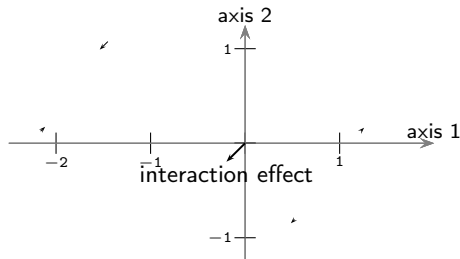
	weights	Axis 1	Axis 2	Plane 1-2
$\text{Var } G^{A/b_1}$	14	1.373	0.120	1.493
$\text{Var } G^{A/b_2}$	28	1.671	0.258	1.930
$\text{Var } G^{A_{\text{within}}B}$	42	1.515	0.186	1.701

Additive cloud

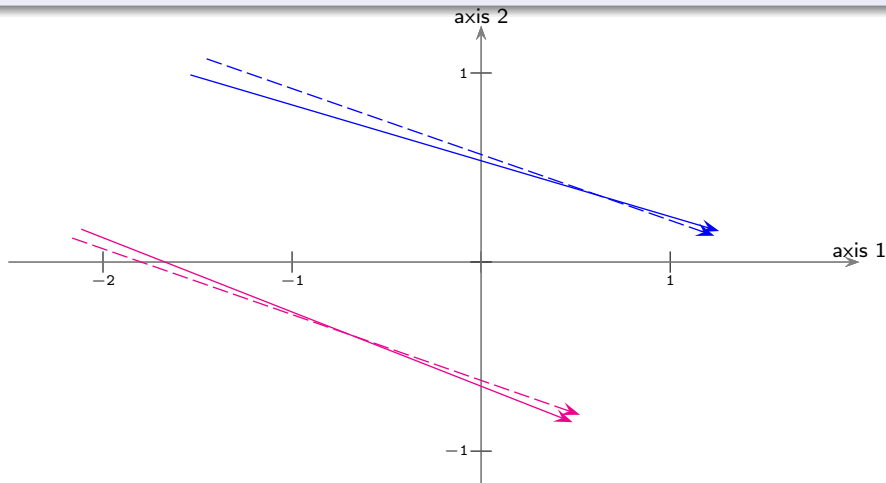
Coordinates	n	Axis 1	Axis 2
$P^{a_1 b_1}$	5	-1.452	1.075
$P^{a_1 b_2}$	9	-2.163	0.127
$P^{a_2 b_1}$	17	1.234	0.139
$P^{a_2 b_2}$	11	0.523	-0.808
$\text{Var } P^{A+B}$	42	1.941	0.320



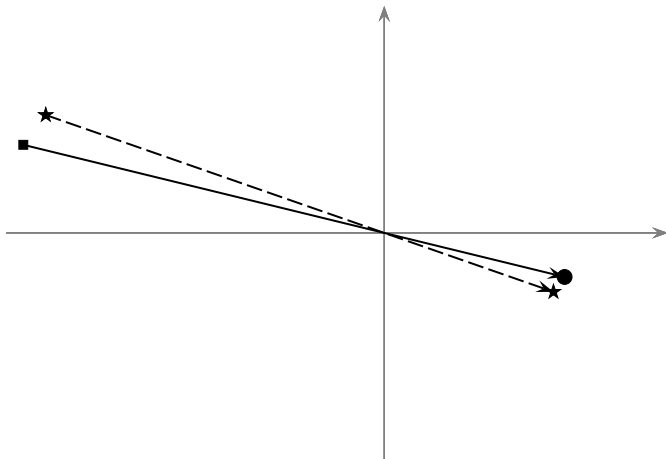
Coordinates	n	Axis 1	Axis 2
$G^{a_1 b_1} - P^{a_1 b_1}$	5	-0.0859	-0.0850
$G^{a_1 b_2} - P^{a_1 b_2}$	9	0.0477	0.0472
$G^{a_2 b_1} - P^{a_2 b_1}$	17	0.0253	0.0250
$G^{a_2 b_2} - P^{a_2 b_2}$	11	-0.0390	-0.086
Variations	42	0.00202	0.00198



In the first principal plane, the variance of the additive cloud ($1.941 + 0.320 = 2.260$) takes into account 99.8% of the variance of the *Status* × *Gender* cloud and 51% of the variance of the overall cloud (η^2 coefficient is equal to $2.260/4.432 = 0.51$, a value that is quite large).



Structure effect



Structure effect and interaction are two different things.

Decompositions of variance

Three additive decompositions of variances of the $Status \times Gender$ cloud:
 additive)+(interaction)

$Status + (Gender \text{ within-} Status)$

$Gender + (Status \text{ within-} Gender)$

	$Status \times$ $Gender$	additive	inter- action	$Gender$	$Status$ within- $Gender$	$Status$	$Gender$ within- $Status$
Axis 2	1.943	1.941	0.002	0.427	1.515	1.822	0.121
Axis 1	0.322	0.320	0.002	0.136	0.186	0.108	0.213
Plane 1-2	2.265	2.261	0.004	0.563	1.701	1.930	0.334

Descriptive findings

The geometric analysis (pca) shows that the structure of the cognitive space is mainly two-dimensional, pause and, by studying the cloud of students, it shows that, in the cognitive space,

- 1 the four groups are well differentiated;
- 2 the interaction effect between factors is nearly null, that is, the crossing of the two factors can be adjusted by an additive model;
- 3 the main effect of *Status* and that of *Gender* are both of large magnitude.