

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

Brigitte.LeRoux@math-info.univ-paris5.fr
rouanet@math-info.univ-paris5.fr

www.math-info.univ-paris5.fr/~lerb/
www.math-info.univ-paris5.fr/~rouanet/

1 Structuring factors

The basic sets for MCA (active individuals and variables) may be built from structuring factors.

Example : EPGY (Educational Program for Gifted Youth)

Two factors:

1. Topics: Integers, Fractions, Geometry, Logic, Measurement
2. Type of variables: error rates, latencies, number of exercises

Hence $5 \times 3 = 15$ *active Variables*.

Structuring factors on *active individuals*: Number of hours on computer, gender, age.

2 Conventional techniques for handling structured data

Analysis of variance: ANOVA, MANOVA

Regression

Structured Data Analysis integrates ANOVA and Regression into GDA.

3 From experimental to observational data

- *Experimental data*: factors or independent variables vs dependent variables.

Effects on factors on dependent variables

- *Observational data*: example of educational study

Structuring factors (I): age, gender, ...: predictors

positions of individuals in space: “dependent” or to be predicted

Structuring factor (II): final exam

position in space: predictors

success to exam to be predicted

Supplementary variables vs structuring factors

Recall property:

one-one correspondence between modalities (in the cloud of modalities) and mean–point modalities (in cloud of individuals).

Example: age classes

Limitation of supplementary modalities: amounts to confining to mean–point modalities

The case of Bourdieu's *La Distinction*.

4 Breakdown of variance

Structuring factors induce a “*factorial design*”, hence

- main effects,
- between and within effects,
- interaction effects

Warning: non-orthogonality of structuring factors

Breakdown of variance for a partition of individuals

Between–cloud, between–variance, within–variance

double breakdown of variance according to partition and axes

Culture Example: Age and principal axes

Age	Abs.	Variances		
	freq.	Axis 1	Axis 2	Axis 3
18-25	449	.1931	.1884	.1938
26-35	574	.2102	.2400	.2157
36-45	520	.2057	.2331	.2101
46-55	394	.2730	.2013	.2183
56-65	317	.2626	.1910	.2074
> 65	466	.2789	.2145	.1793
within		.2335	.2146	.2042
between		.0591	.0270	.0207
total (λ)		.2925	.2415	.2248

Regression of axis 1 to 3 on Age

Regression of axis 1 to 3 on Gender

5 Concentration ellipses

Culture example: Age (see III).

Concentration vs confidence (see VI).

REFERENCES

— LE ROUX B., & ROUANET H. (2004). *Geometric Data Analysis; From Correspondence Analysis to Structured Analysis*. Dordrecht: Kluwer (chapter 6, p.251-258)

— LE ROUX B. & ROUANET H. (2003). Geometric Analysis of Individual Differences in Mathematical Performance for EPGY Students in the Third Grade. www-epgy.stanford.edu/research/.