Conjoint analysis, : principles, applications and new developments

Gilbert Saporta
Chaire de Statistique Appliquée
Conservatoire National des Arts et Métiers
292 rue Saint Martin, 75003 Paris
saporta@cnam.fr
http://cedric.cnam.fr/~saporta

Outline

1. Introduction
2. Data collection
3. Data analysis and simulation
4. Missing ranks
5. Designing scenarios
6. Choice based conjoint and related methods
7. Conclusions and recommandations

1. Introduction

- Conjoint analysis was introduced in the 70's to quantify consumer tradeoffs
- Both:
  - A particular form of multiple **regression** with an ordinal response and explanatory categorical variables (nominal or ordinal)
  - A complete **survey methodology** including data collection based on **experimental designs**, a data analysis phase with parameter **estimation**, a **simulation** phase

- A product is defined by a **combination** of p attributes
- The **perfect** product is generally **unrealistic** like a car with high speed, comfort, security and low price
- A **compensatory** model: the consumer makes a “Trade off” between attributes by putting into balance advantages and inconveniences.
- Conjoint analysis decomposes preferences according to an **additive utility** model, **specific** to each interviewee

2. Data collection

2.1 Full profiles :

- each respondent evaluates a set of K products described by all p attributes

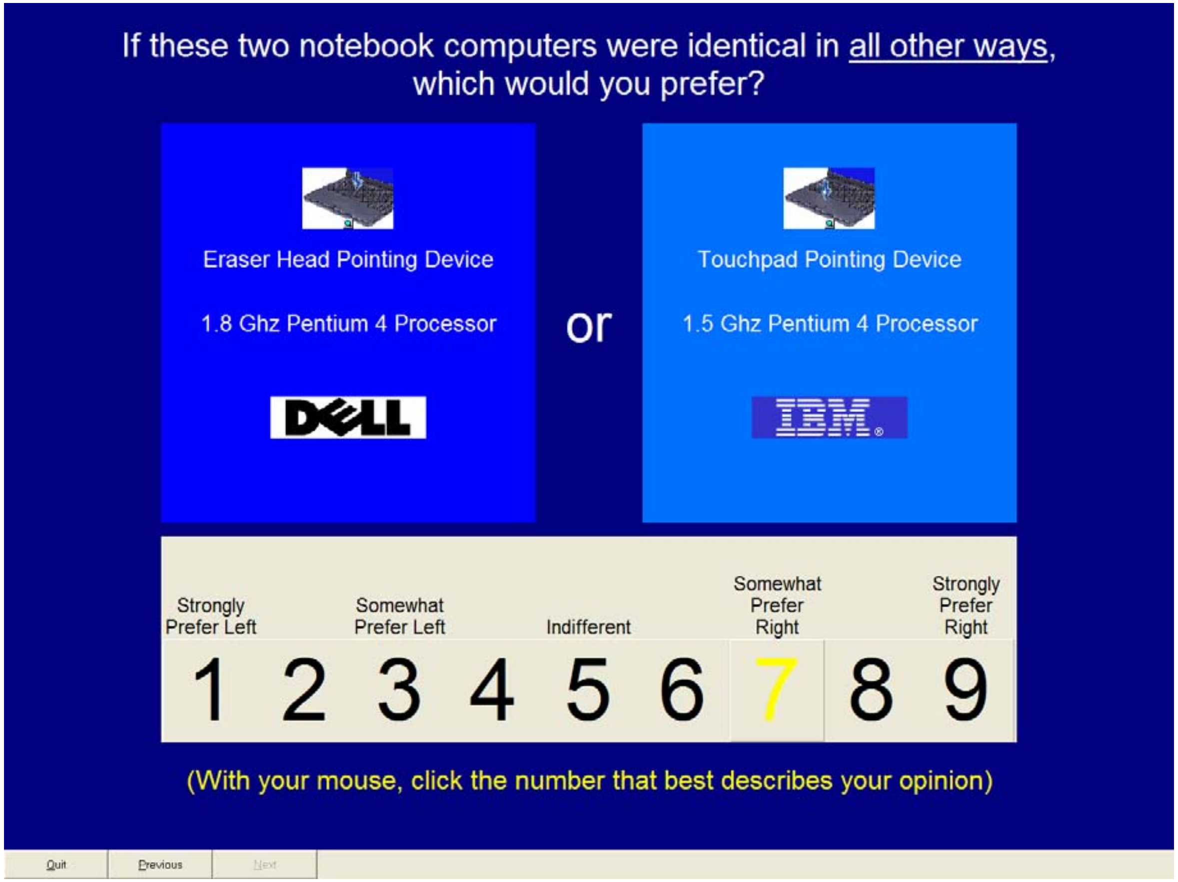
Frozen Diet Entrees

Obs	Ingredient	Fat	Price	Calories
1	Turkey	5 Grams	$1.99	350
2	Turkey	8 Grams	$2.29	350
3	Chicken	8 Grams	$1.99	350
4	Turkey	2 Grams	$2.59	250
5	Beef	8 Grams	$2.59	350
6	Beef	2 Grams	$1.99	350
7	Beef	5 Grams	$2.29	350
8	Beef	5 Grams	$2.29	250
9	Chicken	2 Grams	$2.29	350
10	Beef	8 Grams	$2.59	250
11	Turkey	8 Grams	$2.29	250
12	Chicken	5 Grams	$2.59	350
13	Chicken	5 Grams	$2.59	250
14	Chicken	2 Grams	$2.29	250
15	Turkey	5 Grams	$1.99	250
16	Turkey	2 Grams	$2.59	350
17	Beef	2 Grams	$1.99	250
18	Chicken	8 Grams	$1.99	250

(Kuhfeld, 2005)

- Ranking or ratings?
  - Rating expresses intensity of preferences, but: no comparison between products, problems with comparability of scales across respondents, risk of ties

2.2 Paired profiles comparisons

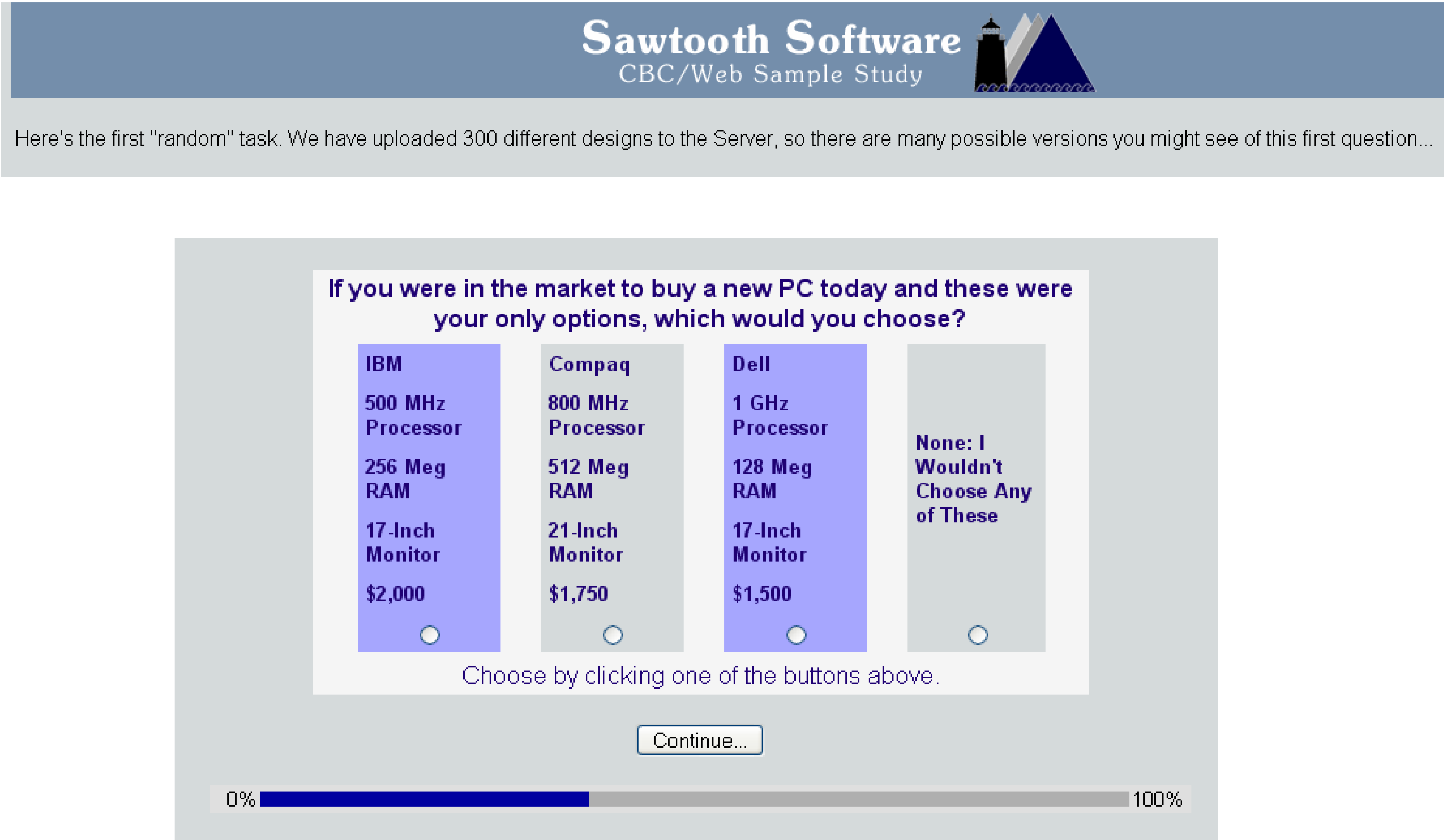


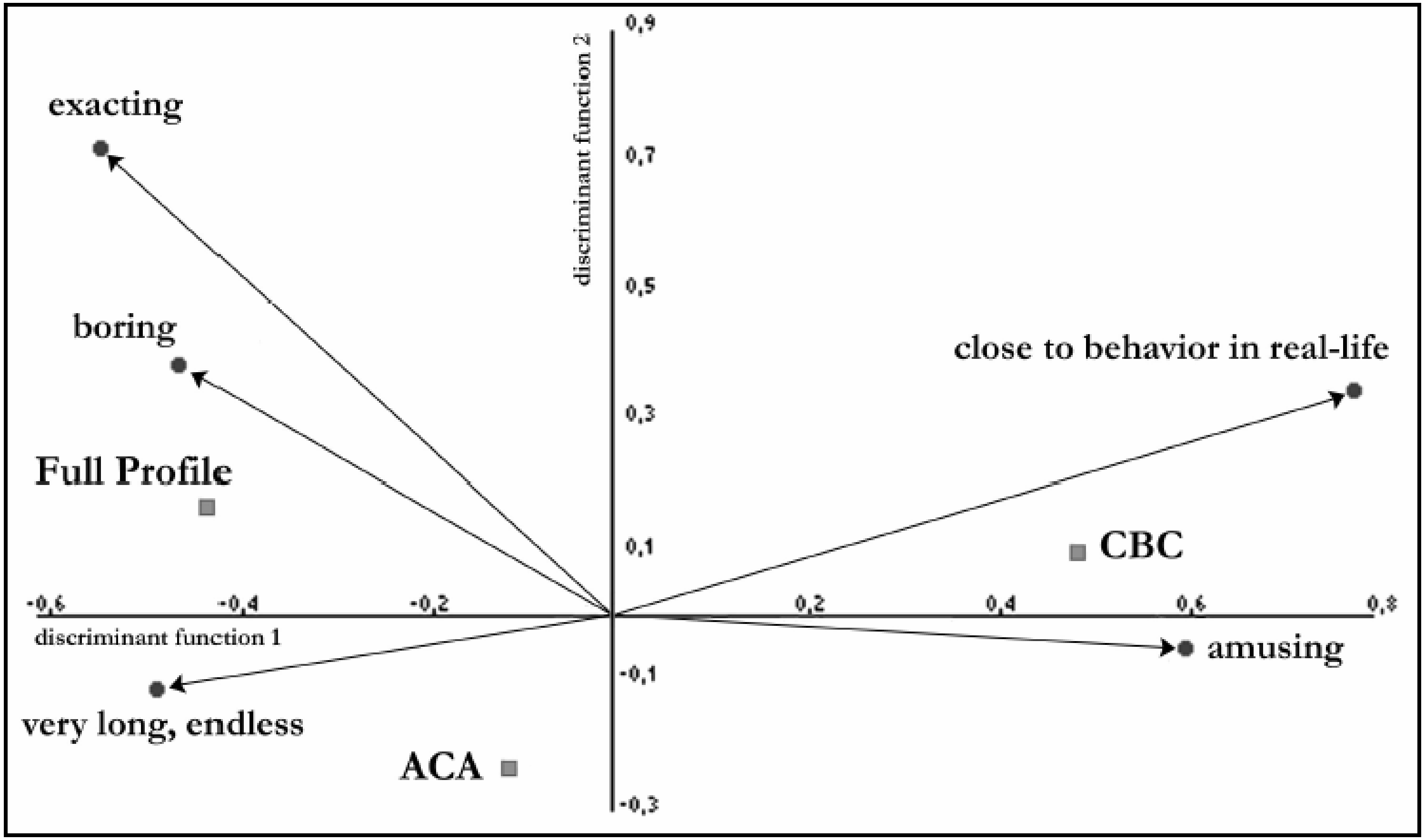
(From Sawtooth Software, ACA)

2.3 full or paired profiles?

- Full profiles : empirical bound (burden for respondents) for the number of comparisons : K ≤ 16 profiles
- Estimation bound : $K > \sum_{i=1}^{p} m_i - p$
- Paired comparisons for large p
- Otherwise « bridging » 2 or more conjoint analysis thanks to common variables

2.4 choice based conjoint





Furlan, Corradetti, 2004

3. Data analysis

- **3.1 A case study:** tariff choice for a cellular phone operator (1996)

  7 attributes

X1 device price	0 F, 700 F
X2 subscription fees	0 F, 200 F
X3 Peak hours definition	p1, p2
X4 Monthly subscription price	0F, 30F, 60F
X5 Duration of subscription	6m, 24m
X6 Price/minute peak hours	3 F, 5 F, 6 F
X7 Price/minute off-peak hours	0.5 F, 0.75 F, 1 F

- 12 products

Obs	Equip	FMS	Duree	Plage	Abt	HP	HC
1	2	2	2	2	3	1	1
2	2	2	1	1	3	3	2
3	2	2	1	1	1	2	3
4	2	1	2	2	1	2	2
5	2	1	2	1	2	1	3
6	2	1	1	2	2	3	1
7	1	2	2	2	1	3	3
8	1	2	2	1	2	2	1
9	1	2	1	2	2	1	2
10	1	1	2	1	3	3	2
11	1	1	1	2	3	2	3
12	1	1	1	1	1	1	1

Unit ID	Ranks												Would buy											
10107000006	2	5	6	7	4	8	9	10	3	11	12	1	3	3	3	3	3	3	3	3	3	3	3	3
10109000004	11	6	10	9	12	7	5	2	8	1	4	3	5	2	5	4	4	2	2	1	3	1	2	2
10109000006	5	2	3	4	10	6	7	8	9	1	11	12	1	2	2	1	4	3	3	3	3	1	4	4
10109000008	1	2	3	4	5	6	8	7	9	10	11	12	1	1	1	3	3	2	3	3	3	4	4	4
10109000009	1	3	2	4	5	6	7	8	9	10	11	12	1	1	1	1	1	3	3	1	3	4	2	1
10109000011	1	2	3	4	5	6	7	8	9	10	11	12	1	2	4	5	4	4	4	4	4	1	4	5
10109000014	1	2	3	4	5	6	7	8	9	10	11	12	2	1	1	1	2	2	2	2	2	2	3	4
10111000007	11	12	9	3	4	10	8	6	5	2	7	1	4	4	4	2	2	4	3	3	2	2	3	2
10119000004	9	12	10	8	7	11	5	4	2	6	3	1	4	4	4	4	4	4	4	4	1	4	1	1
10119000007	6	7	10	4	8	9	11	2	5	12	3	1	4	4	4	4	4	4	4	3	4	4	3	1
10121000003	9	12	8	7	10	11	5	2	3	4	6	1	4	4	4	2	4	4	2	2	2	2	2	2
10125000005	5	11	10	6	2	8	12	7	3	9	1	4	2	3	3	3	2	3	3	3	2	3	2	2
10125000009	8	1	2	7	11	3	9	10	4	12	5	6	3	3	3	3	3	3	3	3	3	3	3	3
11000000007	12	11	8	7	10	9	3	5	4	2	6	1	4	4	4	4	4	4	4	4	4	3	4	3
11003000007	12	11	9	7	8	10	5	1	3	4	6	2	4	4	4	4	4	4	1	1	1	1	1	1
11008000001	12	10	11	8	9	7	6	5	4	2	3	1	4	4	4	4	4	4	3	3	3	2	3	1

3.2 Estimation

3.2.1 Ordinary least squares

- $\min \| \mathbf{y} - \mathbf{Xb} \|^2$ $\mathbf{y} = \mathbf{Xb} + \mathbf{e}$
- $\mathbf{y}$ vector of ranks

$$
\begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ \\ y_{12} \end{pmatrix} = \begin{pmatrix} 01|01|01|.....|100 \\ 01|01|10|.....|010 \\ . \\ . \\ . \\ 10|10|10|.....|100 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ . \\ \\ \\ \\ \xi_3 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ . \\ \\ \\ e_{12} \end{pmatrix}
$$

- Model not of full rank: constraints on utility coefficients. The most popular constraint:

$$\alpha_1 + \alpha_2 + \alpha_3 = 0$$

- Only differences $\alpha_1 - \alpha_2$ are estimable
- Low degree of freedom: 12 scenarios, 10 independent coefficients!
- Criticism to OLS : ranks are not quantitative variables

3.2.2 Monotonous regression

- fit T(**y**) instead of **y** where T is a monotonous transformation of ranks:
- minimize $\| T(\mathbf{y}) - \mathbf{Xb} \|^2$ over T and b

$$y = 1 \qquad T(1) = b_1$$

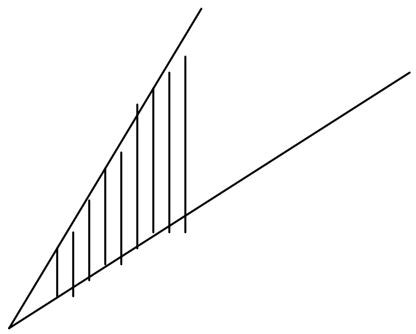
$$y = 2 \qquad T(2) = T(1) + b_2 = b_1 + b_2$$

...

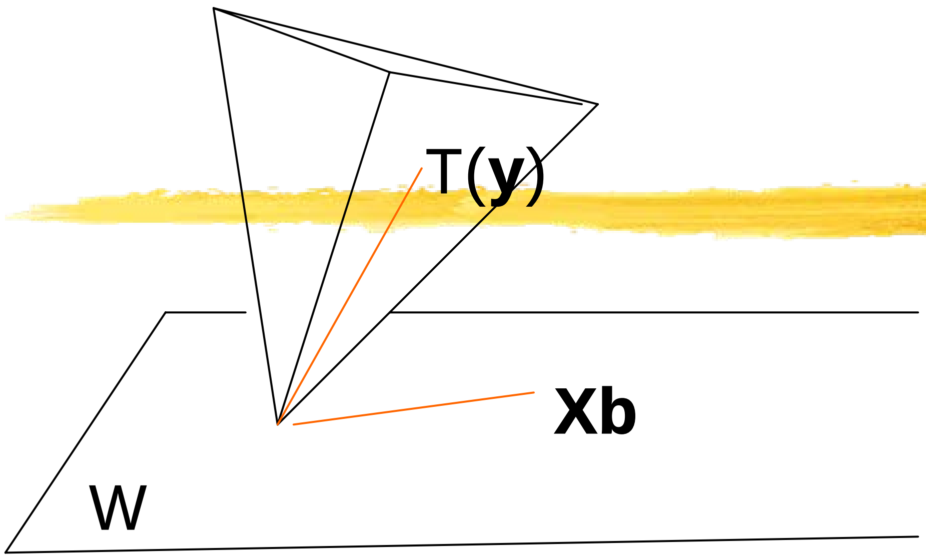
$$y = K \quad T(K) = T(K-1) + b_K = b_1 + b_2 + \ldots + b_K$$

$$b_2 \ldots b_K \geq 0$$

$$T(\mathbf{y}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \\ b_K \end{pmatrix} = b_1 \mathbf{1} + \sum_{j=2}^{K} b_j \mathbf{x}_j$$



Linear combination of K-1 vectors with positive coefficients: Polyedral cone



- a canonical analysis between a cone and the vector sub-space W spanned by the X's columns. Solved by a method of successive projections (alternated least squares)
- But **risk of overfitting .** OLS more robust.

3.3 Goodness of fit

- Measure the agreement between initial ordering and the estimated one:
    - **R^2** or **Kendall's** τ

Threshold → data elimination, but:

- Incoherence or ill-posed problem?

3.4 Attribute importance

- Hierarchy of factors
  - anova , but often non-orthogonal designs
  - Empirical importance: **% of utility range**
- Could be misleading if the category range is unrealistic such as a too large price range

```
Title 'Analyse du trade-off Italienne';

data scen;
infile 'c:\gilbert\papiers et présentations\analyse
conjointe\plital.dat'  ;
input  Equip $ FMS $  Duree $  Plage $ Abt $  HP $ HC $ ;
label Equip='Equip:'
      FMS='FMS:'
      Duree='Duree:'
      Abt='Abt:'
      Plage='Plage:'
      Abt='Abonn:'
      HP='H.Pleines:'
      HC='H.Creuses:' ;
run;

proc print data=scen;
run;

data ranks;
infile 'c:\gilbert\papiers et présentations\analyse
conjointe\italcum.dat' missover ;
input id s1-s12 int1-int12;

proc means mean min max std maxdec=3 ; title2'rangs moyens des
12 scenarios';
var s1-s12;
run;

data rangs;
set ranks;
drop id int1-int12 ;
 proc transpose data=rangs out=ran1 prefix=sujet;
 run;

 data reunion;
 merge scen ran1;
drop _name_;
 run;
```

```
proc transreg data=reunion  utilities method=morals
              outtest=utils;
model linear(sujet:/reflect)=
class (Equip FMS  Duree Plage Abt HP HC/ zero=sum);

output dapproximations ireplace
       out=results(keep=_depend_ t_depend_ a_depend_
       _depvar_  Equip FMS Duree Plage Abt HP HC  );
run;


proc sort data=results out=sims;
by  _depvar_ ;
run;


proc transpose data =sims out=im prefix=scen;
by notsorted _depvar_;
var a_depend_;
label _depvar_='sujet';

proc means data=im n mean std maxdec=3;
title2 'scores moyens des 12 scenarios';
 run;

 data toto;
 set utils;
   if n(importance);
 _depvar_= scan(_depvar_,2);
 name=scan (label,1,' :');
 label=scan(label,1,':');
```

```
proc transpose data=toto out=toto(drop=_name_ _label_);
 id name;
 idlabel label;
 by notsorted _depvar_ ;
 var importance;
 label _depvar_ = 'sujet';
 run
proc means n mean min max std maxdec=3;
 title2 ' Importances moyennes des attributs';
 run;


 data tito;
 set utils;
   if n(coefficient);
 _depvar_= scan(_depvar_,2);
 name=scan (label,1,' :');
 label=scan(label,1,':');
 keep coefficient _depvar_ label ;
 run;


 proc transpose data=tito out=tito(drop= _name_ _label_) ;

 by notsorted _depvar_ ;
 var coefficient;
 label _depvar_ = 'sujet';
 run;


 proc means n mean min max std maxdec=3;
 title2 ' Utilites moyennes des attributs';
 run;

/* Exportation des importances et des utilites individuelles */

 data fich; merge ranks toto tito;
 file 'c:\gilbert\papiers et présentations\analyse conjointe\utital.dat';
 put id @+1 (Equip FMS Duree Plage Abonn H_Pleines H_Creuses col2-
col18)(7.3+1);
 run;
```

Output of Transreg for the first unit

Utilities Table Based on the Usual Degrees of Freedom

Label	Utility	Standard Error	Importance (% Utility Range)	Variable
Intercept	6.5000	0.28868		Intercept
Equip: 1	-1.1667	0.28868	14.286	Class.Equip1
Equip: 2	1.1667	0.28868		Class.Equip2
FMS: 1	-0.6667	0.28868	8.163	Class.FMS1
FMS: 2	0.6667	0.28868		Class.FMS2
Duree: 1	0.6667	0.28868	8.163	Class.Duree1
Duree: 2	-0.6667	0.28868		Class.Duree2
Plage: 1	0.3333	0.28868	4.082	Class.Plage1
Plage: 2	-0.3333	0.28868		Class.Plage2
Abonn: 1	1.6111	0.43033	15.306	Class.Abt1
Abonn: 2	-0.8889	0.43033		Class.Abt2
Abonn: 3	-0.7222	0.43033		Class.Abt3
H.Pleines: 1	4.1111	0.43033	39.796	Class.HP1
H.Pleines: 2	-2.3889	0.43033		Class.HP2
H.Pleines: 3	-1.7222	0.43033		Class.HP3
H.Creuses: 1	0.4444	0.43033	10.204	Class.HC1
H.Creuses: 2	0.6111	0.43033		Class.HC2
H.Creuses: 3	-1.0556	0.43033		Class.HC3

scores individuels des 12 scenarios

Obs	_DEPVAR_	_DEPEND_	T_DEPEND_	A_DEPEND_	Equip	FMS	Duree	Plage	Abt	HP	HC
1	Linear(sujet1)	2	11	11.1667	2	2	2	2	3	1	1
2	Linear(sujet1)	5	8	7.5000	2	2	1	1	3	3	2
3	Linear(sujet1)	6	7	7.5000	2	2	1	1	1	2	3
4	Linear(sujet1)	7	6	5.8333	2	1	2	2	1	2	2
5	Linear(sujet1)	4	9	8.8333	2	1	2	1	2	1	3
6	Linear(sujet1)	8	5	5.1667	2	1	1	2	2	3	1
7	Linear(sujet1)	9	4	3.8333	1	2	2	2	1	3	3
8	Linear(sujet1)	10	3	2.8333	1	2	2	1	2	2	1
9	Linear(sujet1)	3	10	10.1667	1	2	1	2	2	1	2
10	Linear(sujet1)	11	2	2.5000	1	1	2	1	3	3	2
11	Linear(sujet1)	12	1	0.8333	1	1	1	2	3	2	3
12	Linear(sujet1)	1	12	11.8333	1	1	1	1	1	1	1

The TRANSREG Procedure Hypothesis Tests for Linear(sujet1)

Univariate ANOVA Table Based on the Usual Degrees of Freedom

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	142.0000	14.20000	14.20	0.2039
Error	1	1.0000	1.00000		
Corrected Total	11	143.0000			

Root MSE	1.00000	R-Square	0.9930
Dependent Mean	6.50000	Adj R-Sq	0.9231
Coeff Var	15.38462		

Importances and part-worth utilities

	importances							utilities					
10107000006	14.286	8.163	8.163	4.082	15.306	39.796	10.204	-1.167	1.167	-0.667	0.667	0.667	-0.667
10109000004	32.990	6.186	2.062	10.309	7.216	23.711	17.526	2.667	-2.667	0.500	-0.500	0.167	-0.167
10109000006	18.621	10.345	8.276	6.207	12.069	24.483	20.000	-1.500	1.500	-0.833	0.833	-0.667	0.667
10109000008	49.315	24.658	10.959	0.000	6.393	5.023	3.653	-3.000	3.000	-1.500	1.500	-0.667	0.667
10109000009	50.233	25.116	11.163	2.791	6.047	2.326	2.326	-3.000	3.000	-1.500	1.500	-0.667	0.667
10109000011	50.000	25.000	11.111	2.778	7.407	1.852	1.852	-3.000	3.000	-1.500	1.500	-0.667	0.667
10109000014	50.000	25.000	11.111	2.778	7.407	1.852	1.852	-3.000	3.000	-1.500	1.500	-0.667	0.667
10111000007	17.241	20.690	8.621	8.621	15.517	15.517	13.793	1.667	-1.667	2.000	-2.000	-0.833	0.833
10119000004	48.649	8.108	0.000	2.703	8.108	29.730	2.703	3.000	-3.000	0.500	-0.500	-0.000	0.000
10119000007	11.905	4.762	9.524	2.381	3.571	40.476	27.381	0.833	-0.833	0.333	-0.333	0.667	-0.667

3.5 Averaging and clustering

- Individual analysis: one of the main advantages of CA
- Necessity of analysing sample heterogeneity before averaging: factor analysis and clustering on utilities

scores moyens des 12 scenarios

The MEANS Procedure

Variable	N	Mean	Std Dev
scen1	263	5.405	3.379
scen2	263	4.037	3.120
scen3	263	5.331	2.785
scen4	263	6.166	2.778
scen5	263	6.360	2.847
scen6	263	5.013	2.505
scen7	263	6.173	2.863
scen8	263	6.960	2.698
scen9	263	8.199	3.205
scen10	263	6.879	3.074
scen11	263	7.459	3.466
scen12	263	10.018	3.422

Utilites moyennes des attributs

Variable		N	Mean	Minimum	Maximum	Std Dev
constante		263	6.500	6.500	6.500	0.000
Equip	0F	263	1.115	-3.000	3.000	1.722
	700F	263	-1.115	-3.000	3.000	1.722
FMS	0F	263	0.482	-2.000	3.000	0.920
	200F	263	-0.482	-3.000	2.000	0.920
Duree	6m	263	0.176	-2.333	3.000	0.885
	24m	263	-0.176	-3.000	2.333	0.885
Plage	p1	263	0.098	-2.000	2.333	0.606
	P2	263	-0.098	-2.333	2.000	0.606
Abt	0 F	263	0.511	-2.611	4.111	1.178
	30 F	263	-0.151	-4.333	3.111	1.236
	60 F	263	-0.360	-3.944	4.111	1.452
HP	3 F	263	0.999	-3.111	4.444	1.521
	5 F	263	-0.079	-2.889	3.000	1.088
	6 F	263	-0.919	-4.111	2.389	1.316
HC	0.5 F	263	0.137	-3.556	3.889	1.067
	0.75 F	263	0.140	-2.778	3.111	0.945
	1 F	263	-0.277	-3.944	2.944	1.085

Analyse du trade-off Italienne
rangs moyens des 12 scenarios

The MEANS Procedure

Variable	Mean	Minimum	Maximum	Std Dev
s1	7.616	1.000	12.000	3.369
s2	8.901	1.000	12.000	3.207
s3	7.730	1.000	12.000	2.997
s4	6.814	1.000	12.000	2.813
s5	6.620	1.000	12.000	2.896
s6	8.008	1.000	12.000	2.547
s7	6.806	1.000	12.000	2.925
s8	6.019	1.000	12.000	2.764
s9	4.821	1.000	12.000	3.199
s10	6.183	1.000	12.000	3.493
s11	5.521	1.000	12.000	3.442
s12	2.962	1.000	12.000	3.453

Average Importances

Variable	N	Mean	Minimum	Maximum	Std Dev
Equip	263	24.743	0.000	54.545	16.355
FMS	263	11.183	0.000	46.154	8.492
Duree	263	8.448	0.000	46.154	8.110
Plage	263	5.745	0.000	34.783	5.046
Abonn	263	17.072	0.000	64.516	11.115
H_Pleines	263	20.030	1.852	65.333	13.409
H_Creuses	263	12.779	0.000	61.039	8.428

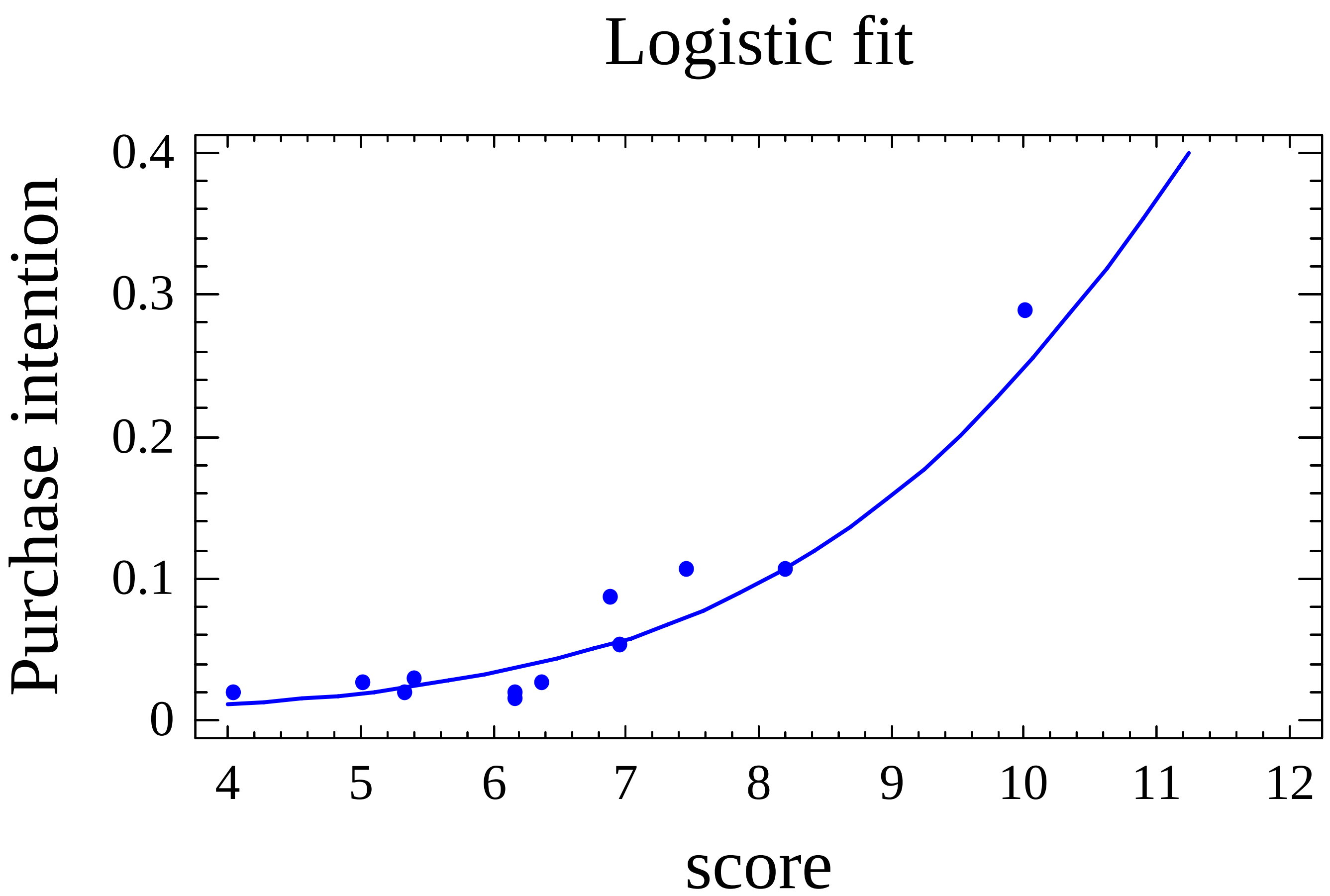
3.6 Simulation, market shares, potential

- **3.6.1 Acceptability and purchase intention curve.**

Respondents are obliged to ranks all K products but nothing indicates that the product classified in 8-th position would be bought by the consumer; it is so necessary to ask it explicitly.

- For every product a "yes no" variable gives the percentage of acceptance *p*.
- Usually: good fit by a logistic function which allows the estimation of potential for not tested products.

$$p = \exp(a + bs)/(1 + \exp(a + bs))$$


Logistic fit
Purchase intention
0.4
0.3
0.2
0.1
0
4
5
6
7
8
9
10
11
12
score

```
Regression Analysis - Logistic model: Y = exp(a + b*X)/(1 + exp(a + b*X))
-----------------------------------------------------------------------------
Dependent variable: int/100
Independent variable: score
-----------------------------------------------------------------------------
                               Standard           T
Parameter        Estimate        Error        Statistic         P-Value
-----------------------------------------------------------------------------
Intercept          -6.762       0.626188      -10.7987           0.0000
Slope            0.565314      0.0938412       6.02415           0.0001
```

Correlation Coefficient = 0.885422
R-squared = 78.3973 percent

3.6.2 Market shares

- 3 competing products
- For each respondent i : U_i^1 ; U_i^2 ; U_i^3
- Several models for respondent choice:
  - Maximal utility
  - Probabilities proportional to U_i^j (Bradley-Terry-Luce)
  - Probabilities proportional to exp (U_i^j) (« logit »)

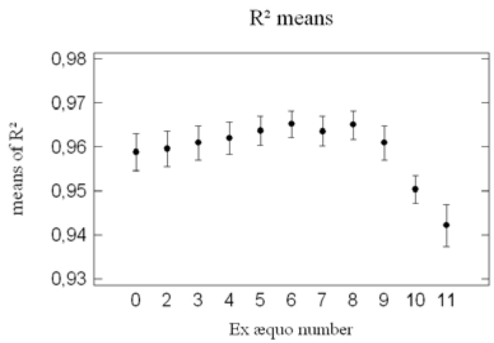
4. Missing ranks

- Ranking up to 16 profiles could be a hard job leading to :
  - Missing values
  - Unreliable rankings
- Often only a partial ordering is available

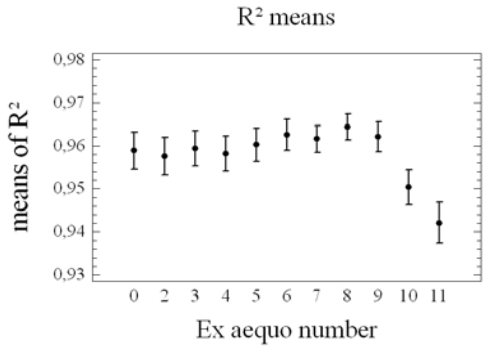
- Benammou & al. (2003) : a study of the influence and remedies to partial ordering with:
  - Three different missing values estimation methods
    - a) All the non ranked scenarios are at the rank of the last given rank plus 1 :   q+1.
    - b) All the non ranked scenarios are at the average (rounded to the nearest integer) of missing ranks.
    - c) All the non ranked scenarios are at the maximum rank K.
  - Simulation of an increasing number of ex-aequos
  - Same data set as before

4.1 Influence on goodness of fit

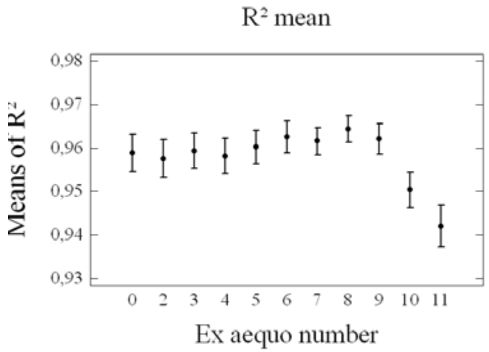
Missing values estimated by the last given rank +1



Missing values estimated by the average of missing ranks

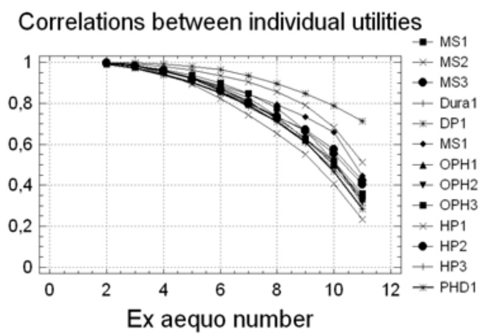


Missing values estimated by the maximum rank

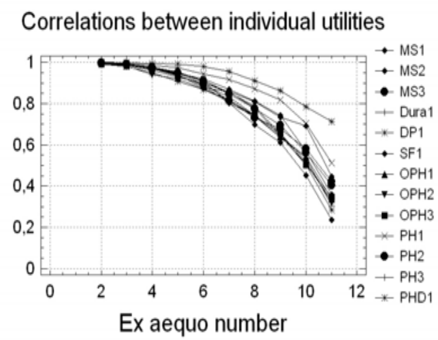


4.2 Influence on utilities

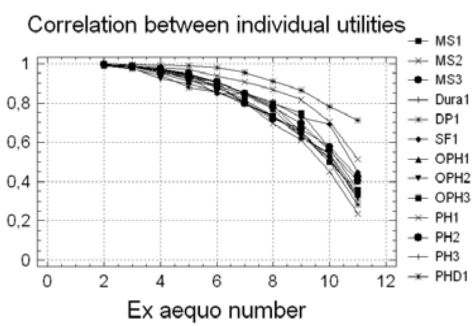
Missing values estimated by the last given rank +1



Missing values estimated by the average of missing ranks



Missing values estimated by the maximum rank



- Weak influence of the imputation method
- Robust up to K/2 missing ranks

5. Designing scenarios

- How to choose the set of products or the X matrix?
- In order to have optimal estimations of the utilities.

  $b = (X'X)^{-1} X'y$ or $(X'X)^{-1} X'T(y)$

  $V(b) = \sigma^2 (X'X)^{-1}$

  (X'X)-1 « small » or (X'X) « large »
- D - optimality: $| X'X |$ maximal
- A - optimality: trace(X'X) max

5.1 Orthogonal designs

- Orthogonal designs are optimal
  - For 2 categories attributes: fractional designs, Plackett-Burman are commonly used
  - For Attributes with the same number of categories: graeco-latin designs

- A few things about experimental designs
  - (-1; +1) or (1;2) are equivalent coding for 2-levels factors
  - (-1;+1) more convenient to check orthogonality
  - Design matrix ≠ model matrix

Run	T	C	Y
1	-1	-1	25
2	+1	-1	31
3	-1	+1	24
4	+1	+1	38

" design matrix " response

linear model $y = X\beta + e$

I	T	C
1	-1	-1
1	+1	-1
1	-1	+1
1	+1	+1

X model matrix, no interaction

I	T	C	T*C
+1	-1	-1	+1
+1	+1	-1	-1
+1	-1	+1	-1
+1	+1	+1	+1

X model matrix, with interaction

X orthogonal: $X'X=4I$

Towards fractional designs

	A	B	C
1	- 1	- 1	- 1
2	+1	- 1	- 1
3	- 1	+1	- 1
4	+1	+1	- 1
5	- 1	- 1	+1
6	+1	- 1	+1
7	- 1	+1	+1
8	+1	+1	+1

Full 2^3 factorial design

	I	A	B	C	A*B	B*C	A*C	A*B*C
1	+1	- 1	- 1	- 1	+1	+1	+1	- 1
2	+1	+1	- 1	- 1	- 1	+1	- 1	+1
3	+1	- 1	+1	- 1	- 1	- 1	+1	+1
4	+1	+1	+1	- 1	+1	- 1	- 1	- 1
5	+1	- 1	- 1	+1	+1	- 1	- 1	+1
6	+1	+1	- 1	+1	- 1	- 1	+1	- 1
7	+1	- 1	+1	+1	- 1	+1	- 1	- 1
8	+1	+1	+1	+1	+1	+1	+1	+1

Orthogonal model matrix with all interactions

If interactions can be ignored, additional factors may be added: D instead of A*B, E instead of B*C etc.

L_8 2^7 (Taguchi) or 2^{7-4} (Box-Hunter)

	A	B	C	D	E	F	G
1	1	1	1	2	2	2	1
2	2	1	1	1	1	2	2
3	1	2	1	1	2	1	2
4	2	2	1	2	1	1	1
5	1	1	2	2	1	1	2
6	2	1	2	1	2	1	1
7	1	2	2	1	1	2	1
8	2	2	2	2	2	2	2

- Next Box-Hunter designs have 16, 32 runs
- Other orthogonal designs where $n=4k$
  - If every 4-uple (1;1)(1;2)(2;1)(2;2) is present k times
  - Plackett-Burman designs n=12,20, 24

Plackett- Burman design L_{12} 2^{11}

A	B	C	D	E	F	G	H	I	J	K
1	2	1	2	2	2	1	1	1	2	1
1	1	2	1	2	2	2	1	1	1	2
2	1	1	2	1	2	2	2	1	1	1
1	2	1	1	2	1	2	2	2	1	1
1	1	2	1	1	2	1	2	2	2	1
1	1	1	2	1	1	2	1	2	2	2
2	1	1	1	2	1	1	2	1	2	2
2	2	1	1	1	2	1	1	2	1	2
2	2	2	1	1	1	2	1	1	2	1
1	2	2	2	1	1	1	2	1	1	2
2	1	2	2	2	1	1	1	2	1	1
2	2	2	2	2	2	2	2	2	2	2

- **Latin squares**
- Lessives: 3 facteurs à 3 modalités
- *Marque (m1,* m2 , m3)

  Conditionnement (classique, micro, liquide)

  Prix (p1 , p2 , p3)
- 27 combinaisons réduites à 9 scénarios

Conditioning \ Brand	m1	m2	m3
classic	p1	p2	p3
micro	p2	p3	p1
liquid	p3	p1	p2

- $k = 4$ carré latin 4^{3-1}

	B_1	B_2	B_3	B_4
A_1	C_1	C_2	C_3	C_4
A_2	C_2	C_3	C_4	C_1
A_3	C_3	C_4	C_1	C_2
A_4	C_4	C_1	C_2	C_3

- **Graeco-latin design for 4 attributes with 3 categories**

Cond\ Brand\Soft.

	m1	m2	m3
classic	p1 α	p2 β	p3 γ
micro	p2 γ	p3 α	p1 β
liquid	p3 β	p1 γ	p2 α

Graeco-latin design for 4 attributes with 4 categories

	1	2	3	4
1	A_α	B_β	$C\gamma$	D_δ
2	$B\ \gamma$	A_δ	D_α	C_β
3	C_δ	$D\ \gamma$	A_β	B_α
4	D_β	C_α	B_δ	$A\ \gamma$

5.2 General case

- Asymmetric designs difficult to find
- May not exist for small K
- A solution : D-optimal designs

Detergents (follows)

- Two price levels instead of 3. Smallest common multiple of 3x3 and 3x2 = 18!
- Solution: create a third fictitious price level to get a latin square with 9 scenarios, the replace price level p_3 by p_2. Orthogonality in broad sense : p_1 present 3 times, p_2 6 times: *collapsing*.

	m_1	m_2	m_3
classic	p_1	p_2	p_2
micro	p_2	p_2	p_1
liquid	p_2	p_1	p_2

- Combining factors:
  - Example 4 factors with 3, 2, 2, 4 levels. Smallest multiple of 3x2 , 2x2 4x2 et 4x3 = 24
  - Orthogonal design in 12 scenarios by using a 5 factors design with one 3-levels factor and 4 2-levels factors:

	A	B	C	D	E
1	1	1	1	1	1
2	1	1	1	2	2
3	1	2	2	1	1
4	1	2	2	2	2
5	2	1	1	1	2
6	2	1	2	2	1
7	2	2	1	2	1
8	2	2	2	1	2
9	3	1	2	1	1
10	3	1	2	2	2
11	3	2	1	1	2
12	3	2	1	2	1

- The 4-levels factor (a, b, c, d) : crossing of 2 2-levels factors (11 , 12 , 21, 22). Replace C and D par E

	A	B	C	E
1	1	1	a	1
2	1	1	b	2
3	1	2	c	1
4	1	2	d	2
5	2	1	a	2
6	2	1	d	1
7	2	2	b	1
8	2	2	c	2
9	3	1	c	1
10	3	1	d	2
11	3	2	a	2
12	3	2	b	1

Optimal designs

- Algorithmic search
  - N candidate design points
  - Choose K, evaluate criterium (eg |X'X|)
  - Exchange points to increase criterium etc.

A sample SAS program

```
proc plan ordered;
factors A=2 B=2 C=2 D=2 E=3 F=3 G=3/noprint;
output out=design;
run;
proc optex data=design;
class A B C D E F G;
model A B C D E F G;
generate n=12 iter=100 method=detmax;
output out=plan;
run;
```

Obs	A	B	C	D	E	F	G
1	2	2	2	1	2	3	2
2	2	2	1	2	3	3	1
3	2	2	1	1	1	1	3
4	2	1	2	2	2	2	3
5	2	1	2	2	1	1	1
6	2	1	1	1	3	2	2
7	1	2	2	2	3	1	2
8	1	2	2	1	1	2	1
9	1	2	1	2	2	2	3
10	1	1	2	1	3	3	3
11	1	1	1	2	1	3	2
12	1	1	1	1	2	1	1

6. Choice based conjoint and related methods

- Asks the respondent to choose one product among several ones

- Closer to real situations
- « No choice » option
- Direct estimation of market shares
- Discrete choice: multinomial logit model
- Major drawback: individual utilities rarely estimated for few choices are proposed.
- **Aggregated** model.

6.1 A preference model for binary choice

- Choice between pairs of products
- Example : 5 attributes
  - Product A x'(A)=10 1000 0001 10 10
  - Product B x'(B)=10 0100 0100 01 01
- $b=(b_1,b_2 \ldots.,b_{14})$: utilities vector
- Scores $s(A)=x'(A)b$ $s(B)=x'(B)b$

- A is preferred to B if s(A)-s(B)>0

  (x'(A)-x'(B))b>0

- In n is the number of binary choices (« duels »)

$$\underbrace{\begin{bmatrix} 00\ \ 1\text{-}100\ \ 0\text{-}101\ 1\text{-}1\ \ 1\text{-}1 \\ \\ \\ \\ \\ . \end{bmatrix} \begin{bmatrix} b_1 \\ . \\ . \\ b_q \end{bmatrix}}_{Xb} \qquad \underbrace{\begin{bmatrix} + \\ - \\ + \\ . \\ . \\ \\ - \end{bmatrix}}_{y}$$

- Estimation of b
  - Logit model (logistic regression) seems appropriate but many degeneracies
  - Fisher's linear discriminant function (or OLS regression)
- Choice of pairs by experimental design

Lancement de la carte Imagine-R

- Produit destiné aux jeunes (scolaires et étudiants)
  - Destiné à lutter contre la fraude
  - Enquête en 1977 auprès de 1200 sujets
  - 5 attributs:
    - Durée (2), Prix (4), Dézonage (4), Carte « avantage » (2), Communication (2)

- 128 combinaisons
- Un plan de Plackett et Burman en 12 scénarios était possible
- la RATP a préféré un modèle de choix binaire, pensant que ce serait plus facile pour des jeunes
- 20 « duels »

PROPOSITION A

CARTE MENSUELLE

PAS DE DÉZONAGE

PRIX 30% INFÉRIEUR

AU PRIX DE LA CARTE ORANGE ACTUELLE

PROPOSITION B

CARTE MENSUELLE

DÉZONAGE

LES WEEK-END, LES MERCREDIS

ET DURANT LES PETITES VACANCES

(HORS CONGÉS D'ÉTÉ)

PRIX 15% INFÉRIEUR

AU PRIX DE LA CARTE ORANGE ACTUELLE

Typologie en 4 groupes selon les utilités

Importances moyennes par groupe

TYPC=1

Variable	N	Mean	Std Error	Minimum	Maximum
DUREE	475	22.920	0.499	0.310	50.430
PRIX	475	40.492	0.505	5.310	65.310
DEZONAGE	475	22.139	0.382	6.259	59.910
CENTRALE	475	7.678	0.305	0.100	44.196
COMM	475	6.772	0.203	0.100	30.607

TYPC=2

Variable	N	Mean	Std Error	Minimum	Maximum
DUREE	345	12.533	0.487	0.260	40.176
PRIX	345	44.265	0.601	9.869	67.400
DEZONAGE	345	29.287	0.376	6.229	55.934
CENTRALE	345	8.142	0.398	0.120	41.774
COMM	345	5.773	0.197	0.000	18.682

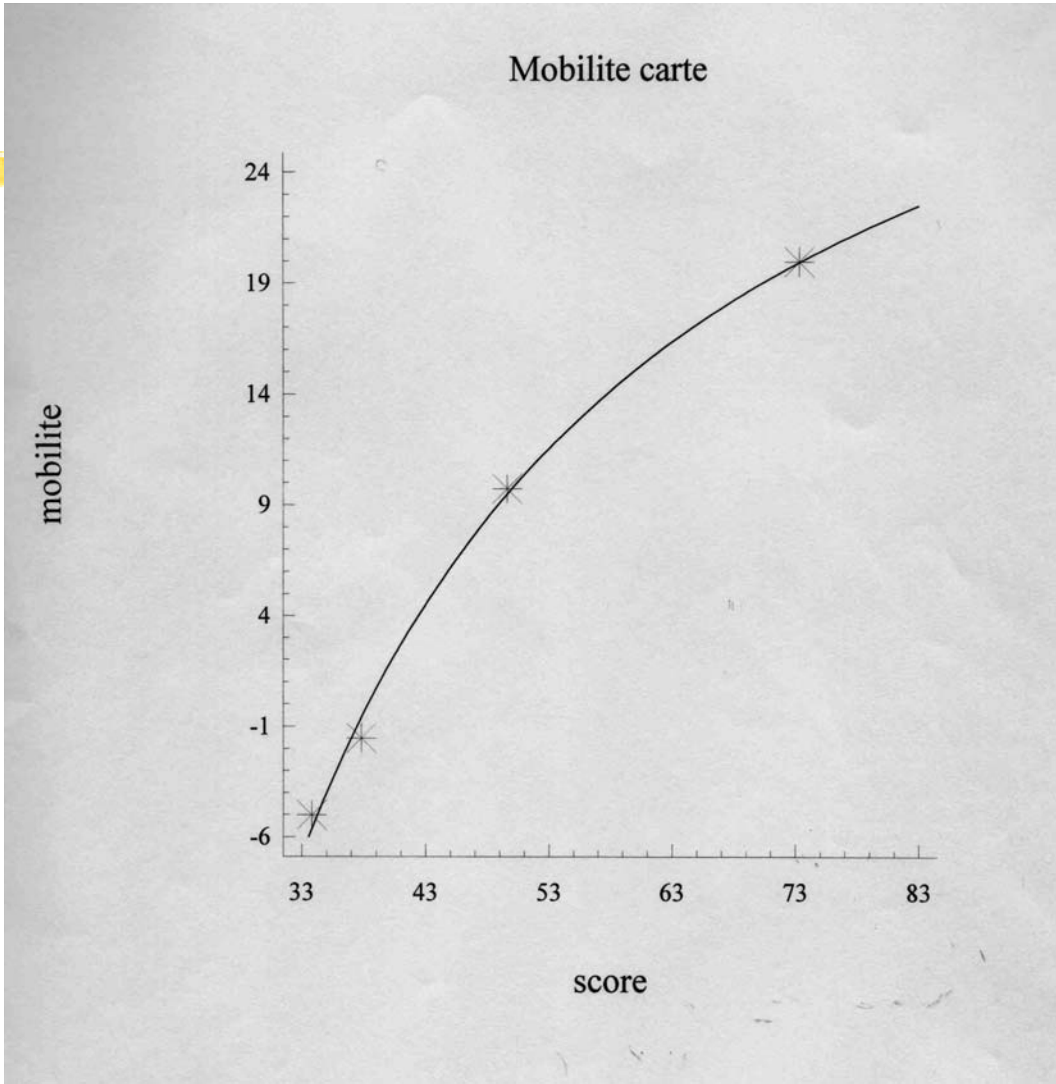
TYPC=3

Variable	N	Mean	Std Error	Minimum	Maximum
DUREE	255	13.888	0.628	0.110	48.730
PRIX	255	41.096	0.864	6.080	68.633
DEZONAGE	255	27.110	0.615	6.530	62.384
CENTRALE	255	9.466	0.441	0.220	43.756
COMM	255	8.439	0.366	0.080	47.520

TYPC=4

Variable	N	Mean	Std Error	Minimum	Maximum
DUREE	264	12.975	0.561	0.110	42.994
PRIX	264	27.529	0.667	4.400	63.326
DEZONAGE	264	40.753	0.672	11.529	69.440
CENTRALE	264	10.178	0.504	0.090	49.830
COMM	264	8.565	0.347	0.040	27.923

Induction de mobilité à partir de 4 scénarios


Mobilite carte
mobilite
24
19
14
9
4
-1
-6
33
43
53
63
73
83
score

6.2 Hierarchical Bayes

- Two levels:
  - At the *upper level*, individuals' vectors of part-worths are assumed to be drawn from a multivariate normal distribution: **b ~ N(a, C)**
  - at the *lower level*, a logit model for each individual is fitted
- Usually: estimation by MCMC
  - needs a large amount of computations, slow convergence

7. Conclusions and recommandations

- Full profile ranking with OLS estimation remains a reliable and robust method and is not a black box
- Analysis of individual preferences, preferable to aggregated models
- CA and data analysis should be done together
- Efficient if there is a real trade-off between attributes
- Necessity of realistic range for categories
- Avoid extrapolation
- However: practical limitations to additive utilities (no interactions)

References

- Benammou S., Harbi S., Saporta G. (2003), *Sur l'utilisation de l'analyse conjointe en cas de réponses incomplètes ou de non-réponses* - Revue de Statistique Appliquée 51, 31-55.
- Dussaix A.M., Saporta G. & al. (1998), *L'analyse conjointe, la statistique et le produit idéal*, CISIA
- Green P.E., Srinivasan V. (1990), *Conjoint analysis in marketing: new developments with implications for research and practice*, Journal of Marketing, 3-19
- Gustaffson A., Herrmann A., Hubert F. (2001), *Conjoint Measurement: Methods and Applications*, Springer
- Kuhfeld W. (2003), *Marketing research methods in SAS*, TS 689
- Lauro N.C., Giordano G. and Verde R. (1998), *A multidimensional approach to conjoint analysis*, Applied Stochastic Model and Data Analysis 14, 265-274.
- Louviere J. J. (1988), *Analyzing Decision Making – Metric Conjoint Analysis*, Sage University Papers.
- Ohannessian S. (2008) *L'option « zéro » en analyse conjointe ; une nouvelle spécification*, Doctorat CNAM
- http://www.sawtoothsoftware.com/techpap.shtml