Young Researchers Seminar 2013

Lyon, France , June 5-7 2013

Laboratoire d'Economie des Transports
Unité Mixte de Recherche du CNRS n° 5593
Université Lumière Lyon 2 - ENTPE

Reliability of quantitative travel surveys: a qualitative check through the Focus Group approach

Valentina Rappazzo

Politecnico di Torino
Interuniversity Dept. of Regional and Urban Studies and Planning

ECTRI
European Conference of Transport Research Institutes

FERSI
Road Safety Research

1. Context

- Interest towards travellers' segmentations performed thanks to multivariate statistical analyses based on people's behaviour, attitudes, habits and opinions;
- Doubts concerning the full reliability of quantitative data → people's opinions 'change' according to the scale used (odd/even, long/short)

Objective :

→ Attempt to check the reliability of a travellers' segmentation by using qualitative data (Focus Groups and interviews) collected on the same sample.

2. Methodology

- Design of an ad hoc survey (Torino metropolitan area) :
  - Questionnaire (quantitative data) → n=358 ;
  - Focus Groups (qualitative data) → addressed to a sub-sample, n=118



- Availability of both ‘qnt’ and ‘qlt’ data on travel behaviour and personality traits for a satisfying sample size (n=118)

Other studies based on a mixed approach :

- Mainly quantitative → very small sample for the qualitative survey ;
- Mainly qualitative → very basic quantitative data (socio-economic)

2. Methodology – The questionnaire

- Recruitment method :
  - On specific sites (Universities, unemployees' centres, disabled people association, elderly people recreational centres, …) ;
  - Snowball sampling technique (mainly for employed people)
- Administration method :
  - Paper and pencil OR electronic form (respondents' preference)
- Questionnaire respondents :
  - 1st stage → FG participants (n=118)
  - 2nd stage → wider sample (n=240÷270)

→ Sub-sample → n = 118

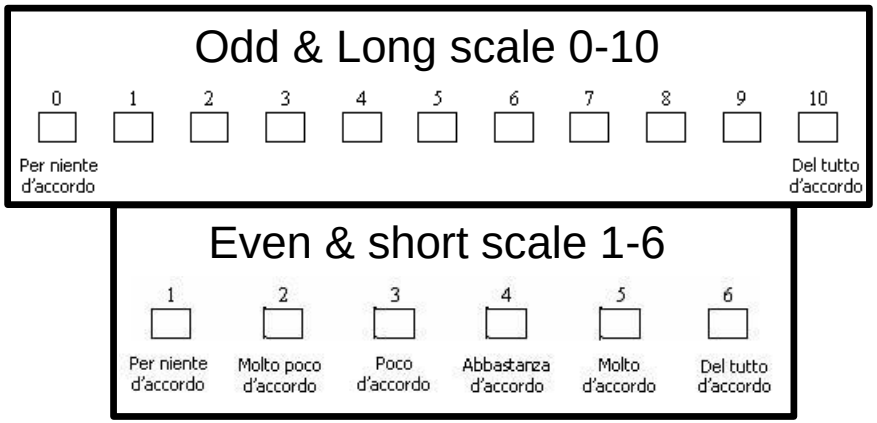
→ Whole sample → n = 358

2. Methodology – The questionnaire

Sections:

1. Principal travel habits;
2. Availability of transport modes;
3. Use of transport modes;
4. Attitudinal variables concerning:
   A. one's own way of being;
   B. the use of transport modes and environmental problems;
   C. life style;
   D. travel habits.
5. Behaviour and life style;
6. Willingness To Pay (WTP);
7. Socio-economical data.

Check consistency of the respondents' opinions



The same topics were also tackled during 12 Focus Group discussions (n=118)

2. Methodology - QNT data analysis design

Identifying the latent constructs: Exploratory Factor Analysis (EFA)

- Principal Axis Factoring Extraction method
→ to define latent constructs and to get a parsimonious representation of the structure of measured variables

No distributional assumptions

- Kaiser criterion (and scree test) → to identify n. of significant factors
- Oblique rotation of factors → allow correlated factors, better interpretation

Users' segmentation: K-means Cluster Analysis

Not unique solution : tests 3, 4, 5, 6, 7 clusters

- How to identify the best clustering?
  - ✓ Kruskal-Wallis non-parametric test statistics (categorical variables)
  - ✓ T-tests (continuous variables)
  - → to test the distributions of variables within the different groups
- K-means Cluster Analysis on the sub-sample (n= 118)

3. Results : Exploratory Factor Analysis

Variables' loadings on factors

Var. codes \ Factors	1. *Environmental responsibility*	2. *Travel pleasure*	3. *Pleasure to travel by car*	4. *Interest in travelling peacefully*	5. *Busy timetable*
packag16	0.712	0.000	0.000	0.000	0.000
lscr1_6	0.669	0.000	0.000	0.000	0.000
renv1_6	0.520	0.000	0.000	0.257	0.000
nwpl1_6	0.000	0.712	0.000	0.000	0.000
adve1_6	0.000	0.630	0.000	0.000	0.000
land1_6	0.000	0.510	0.000	0.000	-0.288
nren1_6	-0.280	0.000	0.549	0.000	0.000
nowc1_6	0.000	0.000	0.527	0.000	0.000
PTwo1_6	0.000	0.000	0.509	0.000	0.000
safe1_6	0.000	0.000	0.000	0.690	0.000
comf1_6	0.000	0.000	0.000	0.646	0.000
rela1_6	0.000	0.000	0.000	0.513	0.000
Atec1_6	0.000	0.000	0.000	0.000	0.567
acti1_6	0.000	0.000	0.000	0.000	0.425
Apub1_6	0.000	0.298	0.000	0.000	0.392
Aspo1_6	0.000	0.000	0.000	0.000	0.383
quic1_6	0.000	0.000	0.000	0.000	0.368
clhm1_6	0.000	-0.361	0.000	0.251	0.000
natd1_6	0.434	0.000	0.000	0.254	0.000
Abio1_6	0.431	0.000	0.000	0.000	0.000
bike1_6	0.327	0.000	-0.263	0.000	0.000
hBKd1_6	0.000	0.000	0.335	0.000	0.000
lkdr1_6	0.000	0.447	0.474	0.000	0.000
paper1_6	0.485	0.000	0.000	0.000	0.000
nwrd1_6	0.000	0.000	0.363	0.000	0.000
litter16	0.461	0.000	0.000	0.000	0.000
nece1_6	0.272	-0.385	0.277	0.000	0.000
water1_6	0.427	0.000	0.000	0.000	0.000
hPTd1_6	0.000	0.000	0.439	0.000	0.000
kwpl1_6	0.000	0.340	0.000	0.000	0.000
dral1_6	0.000	0.467	0.362	0.000	0.000
Variance expl. in data space	**27.19%**	**20.49%**	**13.12%**	**9.12%**	**7.36%**
→ cumulative	**27.19%**	**47.68%**	**60.80%**	**69.92%**	**77.28%**

The identified Factors:

1. Environmental responsibility
   9 variables; var. expl. 27.2%
2. Travel pleasure
   7 variables; var. expl. 20.5%
3. Pleasure to travel by car
   7 variables; var. expl. 13.1%
4. Interest in travelling peacefully
   3 variables; var. expl. 9.1%
5. Busy timetable
   5 variables; var. expl. 7.4%

Tot. Var. expl. in data space: **77.3%**

Loadings < 0.25 are set to 0

vel surveys:
us Group approach

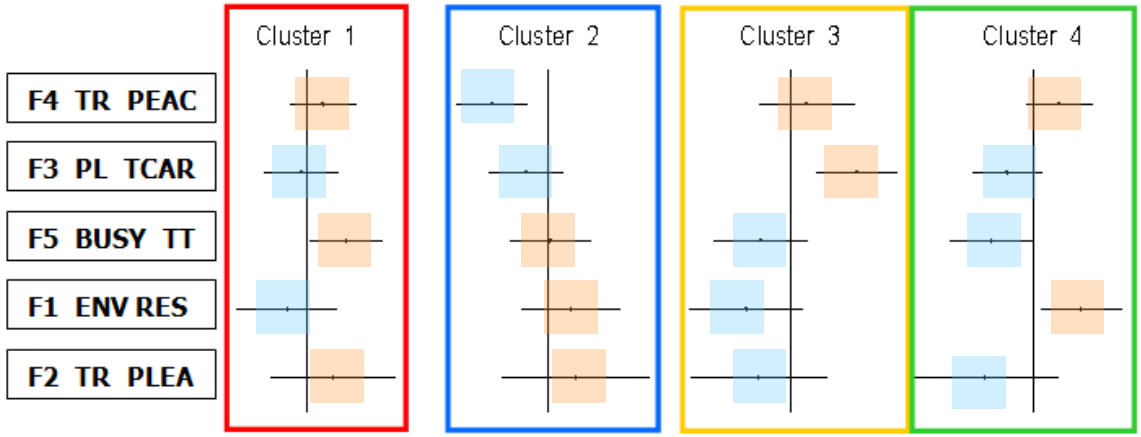
3. Results : Segmentation of the whole sample

K-means Cluster Analysis based on the scores of the 5 factors

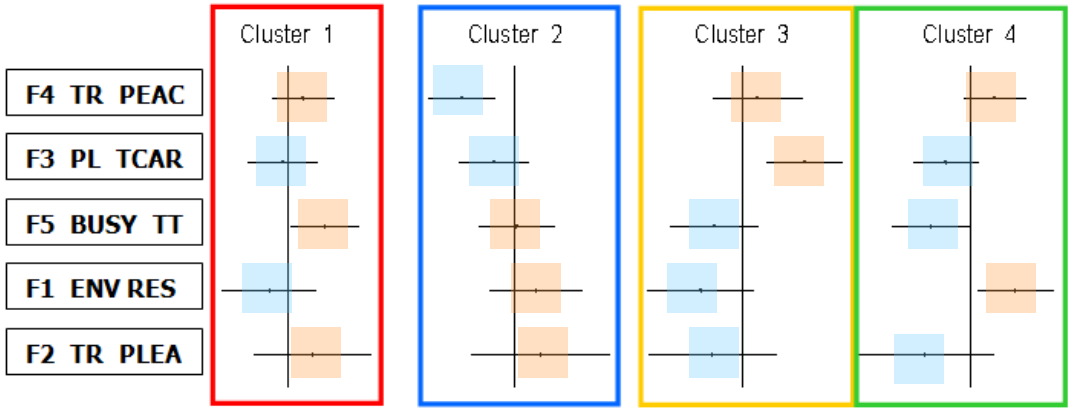
The 4 clusters obtained from the whole database

Cluster Means						
	Size	1. Environm. responsibility	2. Travel pleasure	3. Pleasure to travel by car	4. Interest in travelling peacefully	5. Busy timetable
1	121	4.1398	3.8881	2.5390	5.1377	4.5417
2	86	4.7068	3.8840	2.2203	4.0426	3.8965
3	66	3.7969	3.2703	3.8264	5.1263	3.3667
4	73	5.0628	3.1020	2.1663	5.2740	3.1803
Grand mean		4.4073	3.6071	2.6356	4.8921	3.8719
F-Ratio		51.985	26.540	90.737	93.957	76.246

The clusters' profiles



3. Results : the 4 clusters' profiles (whole sample)



1. Hedonist Urban Travellers (HUT)
   121 IDs → 35%:
   - are interested in travelling peacefully
   - have a busy timetable
   - enjoy travelling

2. BackPackers (BP)
   86 IDs → 24.8%
   - have a busy timetable
   - are environmentally responsible
   - enjoy travelling

3. Lazy TimeServers (LTS)
   66 IDs → 19.1%
   - are interested in travelling peacefully
   - appreciate travelling by car (only group !)

4. Chary and Wise Ecologists (CWE)
   73 IDs → 21.1%
   - are interested in travelling peacefully
   - are environmentally responsible

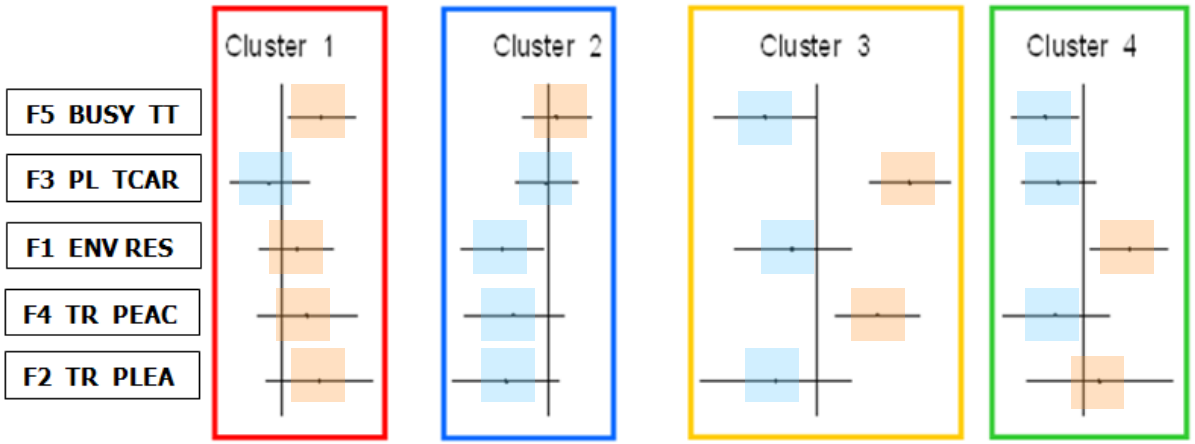
3. Results : Segmentation of the sub-sample

K-means Cluster Analysis based on the scores of the 5 factors

The 4 clusters obtained from the sub-sample of FG participants

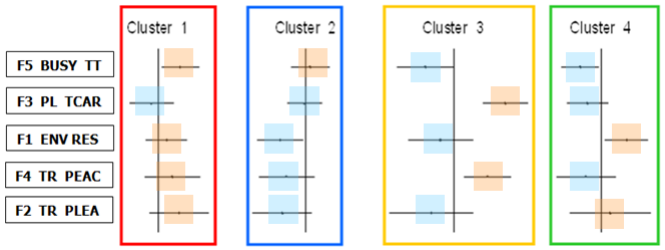
Cluster Means						
	Size	1. Environm. responsibility	2. Travel pleasure	3. Pleasure to travel by car	4. Interest in travelling peacefully	5. Busy timetable
1	40	4.4729	4.0929	2.4250	5.1833	4.5900
2	33	3.5278	3.2727	2.5848	4.5152	4.1152
3	15	3.8741	3.2967	4.3367	5.5778	3.2533
4	28	4.9444	3.8724	2.1786	4.5952	3.4500
Grand mean		4.2446	3.7105	2.6441	4.9023	4.0069
F-Ratio		27.328	12.732	35.180	19.458	36.934

The clusters' profiles



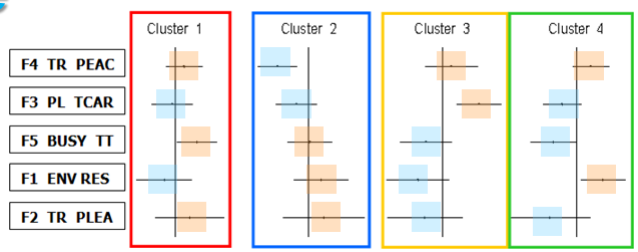
3. Results : Comparison of clusters' profiles

FG participants



Any correnspondence between the 2 clustering ?

Whole sample



1. Ecologist Urban Travellers (EUT) - 40 IDs → 34.5%:
   - have a busy timetable
   - are environmentally responsible
   - are interested in travelling peacefully
   - enjoy travelling

→ Hedonic Urban Travellers (HUTs)

2. Careless and Sedentary (CS) - 33 IDs → 28.5%
   - have a busy timetable

→ ?

3. Very Lazy TimeServers (VLTS) - 15 IDs → 13.9%
   - appreciate travelling by car → only group !!
   - are interested in travelling peacefully

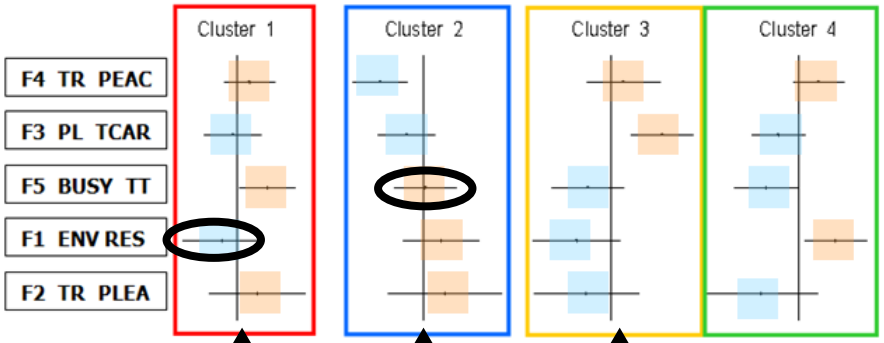
→ Lazy TimeServers (LTSs)

4. Quiet BackPackers (QBP) - 28 IDs → 24.1%
   - are environmentally responsible
   - enjoy travelling

→ BackPackers (BPs)

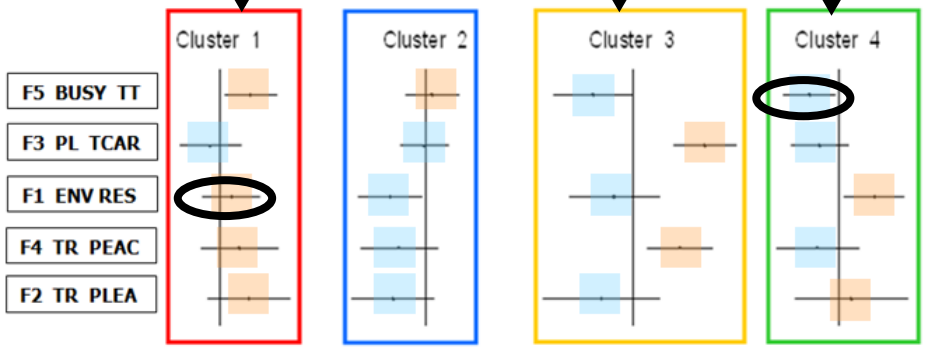
3. Results : Identifing the correspondences

whole sample



1. Hedonist Urban Travellers (n=121; 35%)
2. BackPackers (n=86; 24.8%)
3. Lazy TimeServers (n=66; 19.1%)
4. Chary and Wise Ecologists (n=73; 21.1%)

FG participants



1. Ecologist Urban Travellers (n=40; 34.5%)
2. Careless and Sedentary (n=33; 28.5%)
3. Very Lazy TimeServers (n=15;12.9%)
4. Quiet BackPackers (n=28; 24.1%)

EUTs: 85% (34/40) belonged to HUTs

VLTSs: 93.3% (14/15) belonged to LTSs

QBPs: 50% (14/28) belonged to BPs

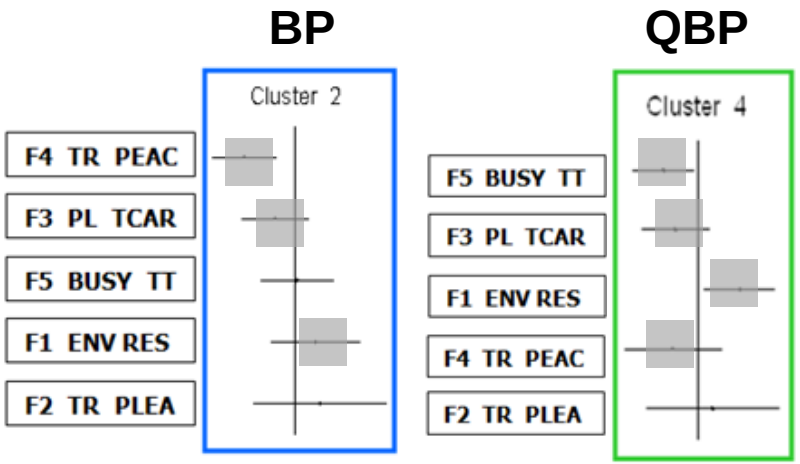
	HUTs	BPs	LTSs	CWEs
EUTs	**85.0 %**	7.5 %	0 %	7.5 %
CSs	**42.4 %**	**36.4 %**	15.1 %	6.1 %
VLTSs	0 %	0 %	**93.3 %**	6.7 %
QBPs	10.7 %	**50 %**	3.6 %	**35.7 %**

4. Discussion : comparing ‘qnt’ & ‘qlt’ data

- Intuitive associations from FG data confirmed by the K-means CA !
- FG data allowed to discover doubtful allocations of the K-means CA

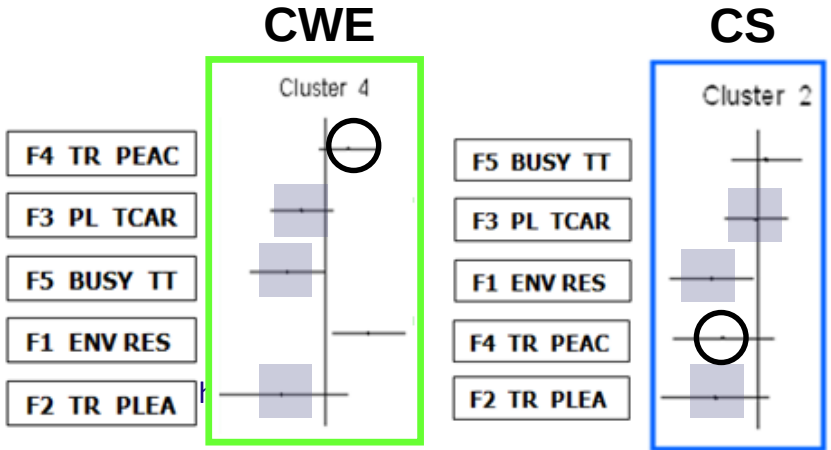
Francesca Bs. – ID74 → BP & QBP

She is not very careful towards the environment (F1), she loves travelling (F2) but hates PT.
She definitely enjoys travelling by car (F3) because she is interested in travelling peacefully and in feeling comfortable (F4). Furthermore, she chairs a voluntary association which became her new job (she is retired), so she is very busy and has an overall very active life style (F5).



Adelina – ID76 → CWE & CS

Adelina is rather carefuls as regards environmental concerns (F1). Although being 65, she travels a lot (F2) and has a very active life style (F5). She enjoys travelling by car (F3) because she wants to feel comfortable (F4), but her driving style is sometimes reckless.



4. Discussion : Face to face interviews

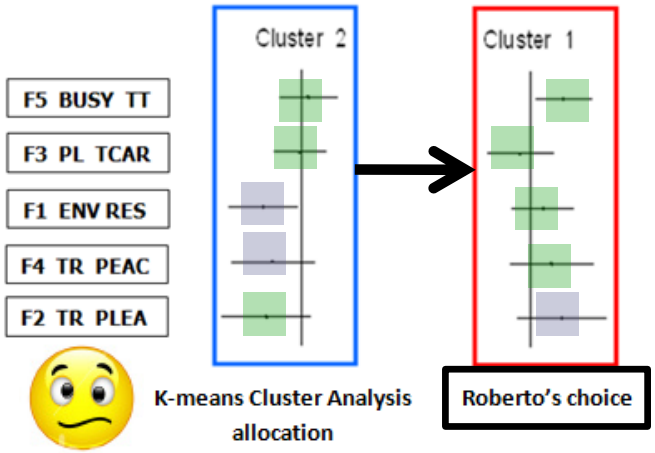
Do the interviewees recognize themselves in their cluster profile ?

Roberto: CS and LTS

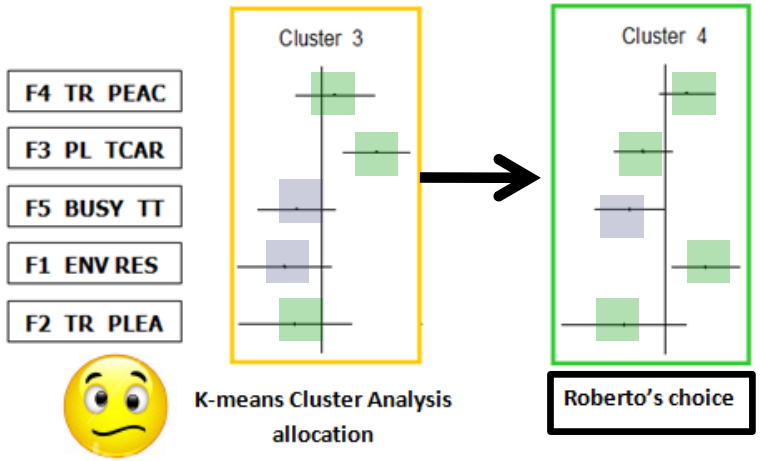
	Scores	1. Env. resp.	2. Trav. pl.	3. Pl. trav. by car	4. Int. trav. peac.	5. Busy timetable
QNT	Individual	4.44	2.00	2.43	5.67	4.20
QNT	CS_FG_2	3.53	3.27	2.58	4.52	4.12
QNT	LTS_ALL_3	3.80	3.27	3.83	5.13	3.37
QLT	Perceived during the FG	M-H	N.A.	H	H	H
QLT	Stated during the interview	H **M**	L	M **L-M**	H	H except Apub1_6



Clustering FG participants



Clustering all cases



5. Discussion and Conclusion

Blindly trust on quantitative data and on travellers' segmentations based on them may entail some risks

Endemic problems affecting quantitative data collection:

- Fatigue and annoyance effects jeopardize data quality and reliability
- Low response rate
- Missing data

} Sample representativeness is compromised

Quantitative surveys using scales to investigate people's opinions:

- Any certitude about the meaning that different people attribute to the same value
- People's opinions may vary a lot according to the used scale

5. Discussion and Conclusion

People's opinion on the same variables: differences according to the scale used

Attitudinal variables concerning one's own way of being

Odd scale 0-10

	quic0_10	land0_10	comf0_10	safe0_10	rela0_10	lkdr0_10	p	inc0_10	nece0_10	acti0_10	clhm0_10	adve0_10	nwpl0_10	kwpl0_10	
0	32	21	1	1		47		0	51	39	83	68	14	9	0
1	6	16	1	1	1	14		1	22	15	33	23	9	7	1
2	8	40	4	2	1	21		4	34	26	48	26	16	11	2
3	32	21	13	7	7	25		4	26	32	36	31	17	13	3
4	15	30	19	6	8	20		9	30	18	19	20	19	19	4
5	39	57	48	21		49		21	41	57	43	44	38	52	5
6	36	39	50	28	22	33		27	34	31	31	43	41	44	6
7	35	36	46	47	55	29		32	27	42	21	31	44	70	7
8	54	40	64	58	55	37		68	33	35	17	29	53	52	8
9	23	12	26	45	56	24		43	15	21	2	13	38	32	9
10	73	37	78	137		48		146	38	34	17	22	63	43	10
*	5	9	8	5	4	11		3	7	8	8	8	6	6	*

Even scale 1-6

	quic1_6	land1_6	comf1_6	safe1_6	rela1_6	lkdr1_6	p	inc1_6	nece1_6	acti1_6	clhm1_6	adve1_6	nwpl1_6	kwpl1_6	
1	37	23	2	0	0	60		0	43	44	86	75	14	9	1
2	19	51	7	5	4	39		6	64	37	77	44	31	15	2
3	47	72	22	13	9	59		21	74	77	76	73	45	44	3
4	105	119	146	72		95		74	75	99	73	92	115	135	4
5	79	59	107	116	122	50		94	66	63	23	42	96	104	5
6	66	27	70	147	126	46		158	29	30	16	24	52	45	6
*	5	7	4	5	4	9		5	7	8	7	8	5	6	*

5. Conclusion

As the **objective** was « Attempt to check the reliability of a travellers' segmentation by using qualitative data (Focus Groups and interviews) collected on the same sample », the main **conclusions** are :

- When the availability of both quantitative and qualitative data on the same people allows to check the results of a segmentation performed through multivariate statistical analysis, some inconsistencies are always found.
- Risk of trusting only quantitative data to perform a users' segmentation.
- An accurate travellers' segmentation is crucial to define ad hoc transport policies to be adressed to specific target groups