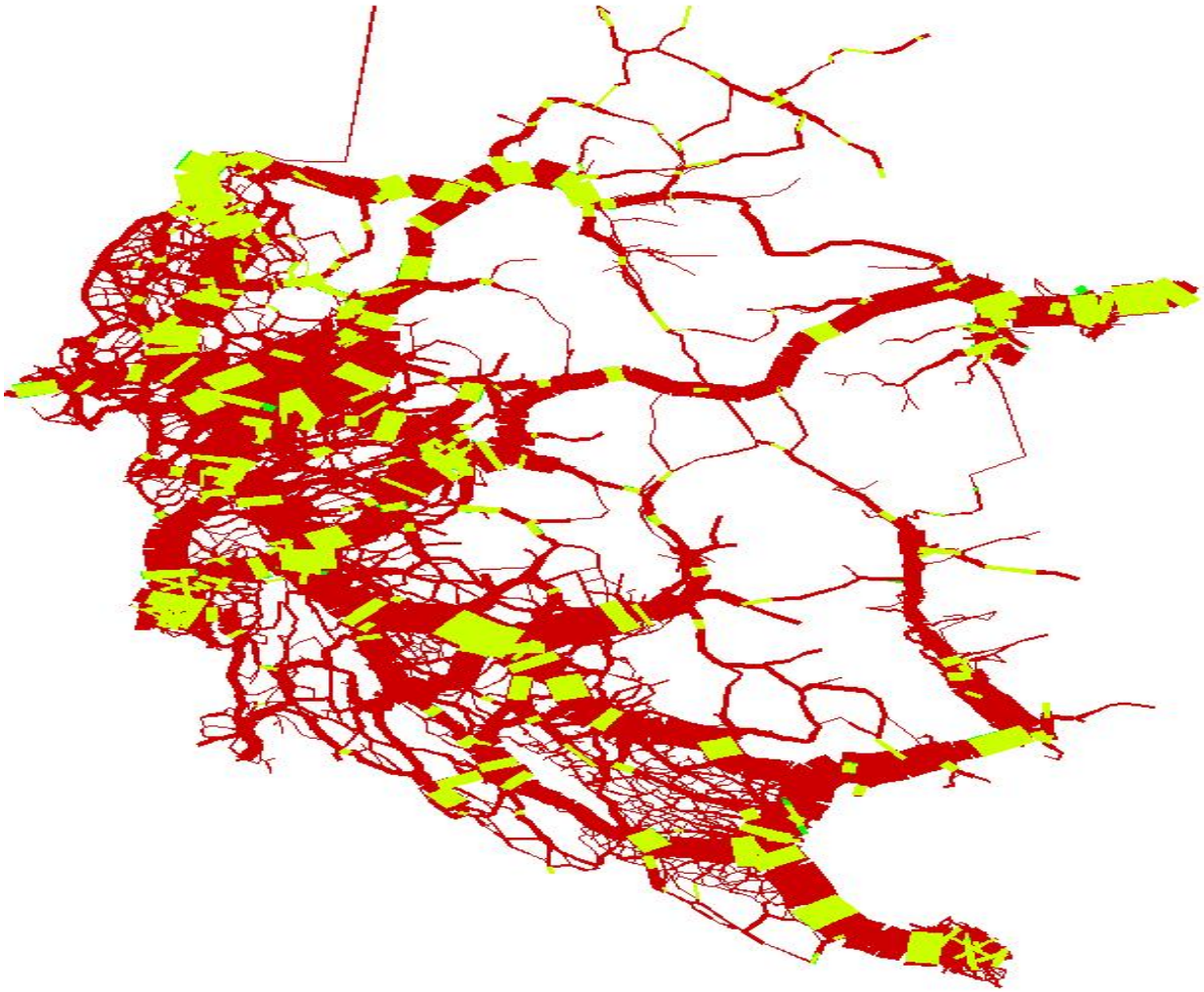

Forecast based on different data types: A before and after study (Stated preference Mode choice)

M Vrtic

Travel Survey Metadata Series **4**

Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Travel Survey Metadata Series

Forecast based on different data types: A before and after study (Stated preference Mode choice)

M Vrtic
IVT
ETH Zürich
Zürich

Phone: 01-633 31 07
Fax: 01-633 10 57

vrtic@ivt.baug.ethz.ch

Abstract

In addition to adequate data, the formulation of transport forecasts relies upon a knowledge of the relationships between the demand for transport and those factors which influence it. These relationships are described with mathematical functions and the their model parameters. The parameters derived from revealed preference (RP) data are often subject to too many impon-derables and are thus of only limited value in many cases. The main cause of this uncertainty is data which is either insufficiently detailed or unsuitable for estimating transport demand functions. For this reason, an earlier SVI study concerning the sensitivity of passenger transport to sup-ply- side and price changes recommended that research should be conducted in parallel with major transport infrastructure investments. In this way this recommended project would be able to check and validate the quality of the findings from alternative or supplementary stated preference (SP) (Vrtic et al., 2000), as both forecast and actual changes in demand would be known to it. The launch of intercity tilting-trains (known as ICNs) in 2001 and further supply-side im-provements to road and rail transport supply offered an opportunity to conduct ex ante/ex post surveys in order to verify the forecasting approaches in a defined period. This mix of qualita-tive and quantitative changes is a particular challenge for forecasting and the attendant data collection process. Nonetheless, it is a challenge that must be overcome again and again in day-to-day practice. As the supply-side changes are generally small, it was expected that changes on the demand side will be concentrated at the level of mode and route choice. The principal aim of this research remit was to verify and identify the limits and possibilities of the three data sources for forecasting by means of an ex ante/ex post analysis. However, this study also offered an opportunity to analyse other aspects of importance to transport fore-casting. Here, in addition to the study methodology, the quality and accessibility of the avail-able bodies of data proved to be crucial factors in modelling transport movements and events. In a first stage the study estimated a detailed public transport route choice model and cali-brated national network models for both road and rail demand. This is an essential preliminary stage to the calculation of modal shifts in demand for transport and the subsequent review of the different forecasts. In the case of mode choice changes, the three most common ap-proaches to forecasting were to be tested: • Direct elasticity, known from previous studies • RP models, i.e. model parameters based on RP data • SP models, i.e. model parameters based on SP data The primary benefits of this study can be summarized as follows: - It sets out the opportunities and limits, as well as the advantages and disadvan-tages, of the three data sources for forecasting under review. - This the first study to provide models of route and mode choice which have been estimated from SP data. The model parameters estimated using this data provide the basis for the practical application of

mode and route choice models following supply-side transport changes. - The estimated model parameters, current figures and the relative valuations dem-onstrate the importance of individual variables to mode and route choice. They were estimated for each trip purpose. - The study showed that in this case the forecasts derived from the SP-data were more consistent and more precise than either the estimates from the direct elastic-ities or the RP data. - Verifying transport forecasts shows how and where further improvements can or must still be made with regard to both data bases and methodology. - The study describes the possibilities and methodical foundation for common SP/RP estimates of model parameters.

Keywords

Route and mode choice

Preferred citation style

Vrtic, M. (2004) Forecast based on different data types: A before and after study (Stated preference Mode choice) , *Travel Survey Metadata Series*, **4**, Institute for Transport Planning and Systems (IVT); ETH Zürich, Zürich.

1.0 Document Description

Citation

Title:	Forecast based on different data types: A before and after study (Stated preference Mode choice)
Identification Number:	KEP SP MC
Authoring Entity:	Institute for Transport planning and Systems (IVT) (ETH Zurich)
Other identifications and acknowledgements:	Vtric M.
Producer:	Vtric M.
	Prof. Axhausen K.W.
Copyright:	Institute for Transport planning and Systems (IVT), ETH Zurich
Date of Production:	2003-06-17
Software used in Production:	Nesstar Publisher

2.0 Study Description

Citation

Title:	Forecast based on different data types: A before and after study (Stated preference Mode choice)
Identification Number:	KEP SP MC
Authoring Entity:	Mr. M. Vrtic (Institute for Transport Planning and Systems, ETH Zurich)
	Prof. KW Axhausen (Institute for Transport Planning and Systems, ETH Zurich)
Producer:	Mr. M. Vrtic
	Prof. KW Axhausen
Date of Production:	2003-06-17
Software used in Production:	Nesstar Publisher
Funding Agency/Sponsor:	Federal Office for Spatial Development
Funding Agency/Sponsor:	Swiss Federal Railway
Grant Number:	ARE, Bern
Grant Number:	SBB, Bern
Distributor:	Federal Office for Spatial Development
Distributor:	Swiss Federal Railway

Study Scope

Keywords: Route and mode choice , Stated preference , Revealed preference , Public transport , Tilting trains , Instituute for Transport Planning and Systems

Topic Classification: Revealed prefernce , Stated preference

Abstract: In addition to adequate data, the formulation of transport forecasts relies upon a knowledge of the relationships between the demand for transport and those factors which influence it. These relationships are described with mathematical functions and the their model parameters. The parameters derived from revealed preference (RP) data are often subject to too many impon-derables and are thus of only limited value in many cases. The main cause of this uncertainty is data which is either insufficiently detailed or unsuitable for estimating transport demand functions. For this reason, an earlier SVI study concerning the sensitivity of passenger transport to sup-ply- side and price changes recommended that research should be conducted in parallel with major transport infrastructure investments. In this way this recommended project would be able to check and validate the quality of the findings from alternative or supplementary stated preference (SP) (Vrtic et al., 2000), as both forecast and actual changes in demand would be known to it. The launch of intercity tilting-trains (known as ICNs) in 2001 and further supply-side im-provements to road and rail transport supply offered an opportunity to conduct ex ante/ex post surveys in order to verify the forecasting approaches in a defined period. This mix of qualita-tive and quantitative changes is a particular challenge for forecasting and the attendant data collection process. Nonetheless, it is a challenge that must be overcome again and again in day-to-day practice. As the supply-side changes are generally small, it was expected that changes on the demand side will be concentrated at the level of mode and route choice. The principal aim of this research remit was to verify and identify the limits and possibilities of the three data sources for forecasting by means of an ex ante/ex post analysis. However, this study also offered an opportunity to analyse other aspects of importance to transport fore-casting. Here, in addition to the study methodology, the quality and accessibility of the avail-able bodies of data proved to be crucial factors in modelling transport movements and events. In a first stage the study estimated a detailed public transport route choice model and cali-brated national network models for both road and rail demand. This is an essential preliminary stage to the calculation of modal shifts in demand for transport and the subsequent review of the different forecasts. In the case of mode choice changes, the three most common ap-proaches to forecasting were to be tested: • Direct elasticity, known from previous studies •

	RP models, i.e. model parameters based on RP data • SP models, i.e. model parameters based on SP data The primary benefits of this study can be summarized as follows: - It sets out the opportunities and limits, as well as the advantages and disadvan-tages, of the three data sources for forecasting under review. - This the first study to provide models of route and mode choice which have been estimated from SP data. The model parameters estimated using this data provide the basis for the practical application of mode and route choice models following supply-side transport changes. - The estimated model parameters, current figures and the relative valuations dem-onstrate the importance of individual variables to mode and route choice. They were estimated for each trip purpose. - The study showed that in this case the forecasts derived from the SP-data were more consistent and more precise than either the estimates from the direct elastic-ities or the RP data. - Verifying transport forecasts shows how and where further improvements can or must still be made with regard to both data bases and methodology. - The study describes the possibilities and methodical foundation for common SP/RP estimates of model parameters.
Country:	Switzerland
Geographic Coverage:	Switzerland
Unit of Analysis:	Person
Universe:	All the individuals permanantly residing in Switzerland.

Methodology and Processing

Time Method:	9 months (January 2001 to September 2001)
Sampling Procedure:	Sample frame: All the individuals of age 15-84, residing permanantly in Switzerland Sample unit: Person or Individual Sampling technique: Random sampling
Mode of Data Collection:	Self-Administrative and structured written interview technique was implemented to capture the preference on pre-defined situations.

Sources Statement

Weighting: No weighting was done.

Other Study Description Materials

Related Materials

Citation

Title:	Stated prefernce mode choice questionnaire
Holdings Information:	www.ivt.baug.ethz.ch/ethtda/kep/sp/mc/questionnaire.pdf

3.0 File Description

File: Stated Preference - Mode Choice.NSDstat

- Number of cases: 9028
- No. of variables per record: 23
- Type of File: NSDstat 200203

4.0 Variable Description

List of Variables:

- Interview Number
- Person and Situation number (key)
- Week of the year
- Mode choice
- Stated preference situation number
- Comfort
- Train station access time (in minutes)
- Travel time with car (in minutes)
- Travel time with train (in minutes)
- Number of transfers
- Headway (in minues)
- Travel cost with train (in CHF)
- Travel cost with car (in CHF)
- Reliability with train (delay in minutes)
- Reliability with cars (dealy in minutes)
- Employed
- Car availability
- Age (in years)
- Gender
- Number of km in '000 travelled by car in the last year
- Season ticket availability
- Trip purpose
- Language

Variables

Variable: **Interview Number**

Location: **Summary Statistics:**

Width: 11 *Variable Format*: character

Variable: **Person and Situation number (key)**

Location: **Summary Statistics:**

Width: 11 *Variable Format*: character

Variable: **Week of the year**

Location:

Width: 11

Value	Label	Frequency
13 .		295
14 .		267
15 .		431
16 .		374
17 .		579
18 .		334
19 .		344
20 .		479
21 .		245
22 .		675
23 .		521
24 .		281
25 .		160
26 .		343
27 .		246
28 .		396
30 .		254
31 .		283
32 .		587
33 .		319
34 .		222
35 .		383
36 .		220
37 .		350
38 .		243
Sysmiss .		197

Range of Valid Data Values: 13 to 38

Total Responses: Summation of listed categories: 9028

Summary Statistics:

Minimum : 13

Maximum : 38

Mean : 24.7

Variable Format: numeric

Variable: **Mode choice**

Location:

Width: 11

Value	Label	Frequency
1 .	Car	6125
2 .	Train	2903

Range of Valid Data Values: 1 to 2

Total Responses: Summation of listed categories: 9028

Summary Statistics:

Minimum : 1

Maximum : 2

Mean : 1.322

Standard deviation : 0.467

Variable Format: numeric

Variable: **Stated preference situation number**

Location:

Width: 11

Value	Label	Frequency
1 .		1140
2 .		1134
3 .		1127
4 .		1131
5 .		1134
6 .		1125
7 .		1121
8 .		1116

Range of Valid Data Values: 1 to 8

Total Responses: Summation of listed categories: 9028

Summary Statistics:

Variable Format: numeric

Variable: **Comfort**

Location:

Width: 11

Value	Label	Frequency
1 .	Regional train	2286
2 .	Inter regional train (double decker)	2259
3 .	Inter-city	2254
4 .	Tilting trains	2229

Range of Valid Data Values: 1 to 4

Total Responses: Summation of listed categories: 9028

Summary Statistics:

Minimum : 1

Maximum : 4

Mean : 2.49

Standard deviation : 1.118

Variable Format: numeric

Variable: **Train station access time (in minutes)**

Location:

Width: 11

Value	Label	Frequency
10 .		61
11 .		156
12 .		101
13 .		88
14 .		267
15 .		493
16 .		544
17 .		639
18 .		460
19 .		522
20 .		2010
21 .		672
22 .		209
23 .		400
24 .		477
25 .		734
26 .		222
27 .		232
28 .		71
29 .		107
30 .		502
35 .		61

Range of Valid Data Values: 10 to 35

Total Responses: Summation of listed categories: 9028

Summary Statistics:

Minimum : 10

Maximum : 35

Mean : 20.582

Standard deviation : 4.621

Variable Format: numeric

Variable: **Travel time with car (in minutes)**

Location:

Width: 11

Value	Label	Frequency
3 .		1
4 .		4
5 .		93
6 .		10
7 .		25
8 .		22
9 .		57
10 .		307
11 .		65
12 .		83
13 .		88
14 .		32
15 .		188
16 .		74
17 .		76
18 .		138
19 .		44
20 .		401
21 .		78
22 .		98
23 .		70
24 .		30
25 .		201
26 .		97
27 .		90
28 .		70
29 .		32
30 .		444
35 .		209

40 .	576
45 .	222
50 .	513
55 .	202
60 .	233
65 .	369
70 .	319
75 .	104
80 .	285
85 .	125
90 .	249
95 .	95
100 .	217
105 .	90
110 .	202
115 .	93
120 .	108
125 .	148
130 .	213
135 .	73
140 .	157
145 .	59
150 .	153
155 .	35
160 .	124
165 .	21
170 .	114
175 .	41
180 .	53
185 .	103
190 .	86

195 .	34
200 .	53
205 .	18
210 .	48
215 .	19
220 .	60
225 .	16
230 .	19
235 .	12
240 .	12
245 .	22
250 .	41
255 .	10
260 .	21
265 .	4
270 .	10
275 .	10
280 .	5
285 .	4
290 .	10
305 .	7
310 .	13
315 .	2
330 .	7
340 .	4
350 .	10
355 .	5
380 .	6
385 .	1
390 .	5
395 .	2

400 .	1
420 .	3
435 .	5
440 .	5
455 .	5
475 .	5
495 .	5

Range of Valid Data Values: 3 to 495

Total Responses: Summation of listed categories: 9028

Summary Statistics:

Minimum : 3

Maximum : 495

Mean : 75.922

Standard deviation : 66.183

Variable Format: numeric

Variable: **Travel time with train (in minutes)**

Location:

Width: 11

Value	Label	Frequency
3 .		13
4 .		17
5 .		251
6 .		66
7 .		51
8 .		67
9 .		132
10 .		445
11 .		86
12 .		70
13 .		72
14 .		84
15 .		174
16 .		74
17 .		70
18 .		116
19 .		81
20 .		218
21 .		74
22 .		71
23 .		54
24 .		68
25 .		153
26 .		67
27 .		82
28 .		45
29 .		57
30 .		290
35 .		235

40 .	442
45 .	153
50 .	364
55 .	205
60 .	204
65 .	293
70 .	271
75 .	130
80 .	224
85 .	83
90 .	246
95 .	117
100 .	216
105 .	65
110 .	199
115 .	60
120 .	132
125 .	164
130 .	206
135 .	83
140 .	136
145 .	62
150 .	80
155 .	75
160 .	95
165 .	45
170 .	129
175 .	49
180 .	72
185 .	84
190 .	105

195 .	25
200 .	72
205 .	39
210 .	94
215 .	21
220 .	69
225 .	35
230 .	52
235 .	48
240 .	35
245 .	21
250 .	47
255 .	22
260 .	59
265 .	15
270 .	30
275 .	13
280 .	39
285 .	13
290 .	36
295 .	11
300 .	10
305 .	10
310 .	30
315 .	1
320 .	5
325 .	10
330 .	30
335 .	5
340 .	9
345 .	4

350 .	10
365 .	9
370 .	10
375 .	3
380 .	10
400 .	4

Range of Valid Data Values: 3 to 400

Total Responses: Summation of listed categories: 9028

Summary Statistics:

Minimum : 3

Maximum : 400

Mean : 83.893

Standard deviation : 76.992

Variable Format: numeric

Variable: **Number of transfers**

Location:

Width: 11

Value	Label	Frequency
0 .	None	4479
1 .	One	2256
2 .	Two	2293

Range of Valid Data Values: 0 to 2

Total Responses: Summation of listed categories: 9028

Summary Statistics:

Minimum : 0

Maximum : 2

Mean : 0.758

Standard deviation : 0.832

Variable Format: numeric

Variable: **Headway (in minues)**

Location:

Width: 11

Value	Label	Frequency
15 .		2221
30 .		2273
60 .		4534

Range of Valid Data Values: 15 to 60

Total Responses: Summation of listed categories: 9028

Summary Statistics:

Mean : 41.376

Standard deviation : 19.441

Variable Format: numeric

Variable: **Travel cost with train (in CHF)**

Location:

Width: 11

Value	Label	Frequency
2 .		1531
3 .		622
4 .		555
5 .		518
6 .		473
7 .		258
8 .		322
9 .		256
10 .		227
11 .		254
12 .		192
13 .		256
14 .		218
15 .		182
16 .		160
17 .		149
18 .		133
19 .		140
20 .		138
21 .		122
22 .		106
23 .		137
24 .		147
25 .		96
26 .		125
27 .		100
28 .		99
29 .		94
30 .		92

31 .	83
32 .	92
33 .	54
34 .	83
35 .	69
36 .	52
37 .	43
38 .	61
39 .	55
40 .	59
41 .	41
42 .	60
43 .	44
44 .	39
45 .	50
46 .	20
47 .	35
48 .	20
49 .	18
50 .	8
51 .	41
52 .	27
53 .	26
54 .	29
55 .	16
56 .	46
57 .	15
58 .	15
59 .	34
60 .	15
61 .	6

62 .	5
63 .	3
64 .	2
66 .	9
68 .	3
70 .	5
71 .	3
72 .	8
74 .	6
75 .	4
77 .	6
79 .	14
89 .	2

Range of Valid Data Values: 2 to 89

Total Responses: Summation of listed categories: 9028

Summary Statistics:

Minimum : 2

Maximum : 89

Mean : 14.894

Standard deviation : 14.587

Variable Format: numeric

Variable: **Travel cost with car (in CHF)**

Location:

Width: 11

Value	Label	Frequency
0 .		35
1 .		802
2 .		905
3 .		609
4 .		678
5 .		541
6 .		335
7 .		344
8 .		322
9 .		292
10 .		242
11 .		311
12 .		213
13 .		220
14 .		166
15 .		169
16 .		154
17 .		169
18 .		128
19 .		137
20 .		133
21 .		168
22 .		147
23 .		137
24 .		117
25 .		114
26 .		89
27 .		113
28 .		82

29 .	86
30 .	87
31 .	64
32 .	93
33 .	72
34 .	60
35 .	56
36 .	61
37 .	57
38 .	49
39 .	52
40 .	37
41 .	26
42 .	31
43 .	34
44 .	34
45 .	29
46 .	49
47 .	24
48 .	28
49 .	39
50 .	7
51 .	7
52 .	5
53 .	6
54 .	9
55 .	3
57 .	5
58 .	7
60 .	10
61 .	6

63 .	2
64 .	6
65 .	3
66 .	9
73 .	3

Range of Valid Data Values: 0 to 73

Total Responses: Summation of listed categories: 9028

Summary Statistics:

Minimum : 0

Maximum : 73

Mean : 12.874

Standard deviation : 12.442

Variable Format: numeric

Variable: **Reliability with train (delay in minutes)**

Location:

Width: 11

Value	Label	Frequency
0 .		2247
5 .		2279
15 .		4502

Range of Valid Data Values: 0 to 15

Total Responses: Summation of listed categories: 9028

Summary Statistics:

Minimum : 0

Maximum : 15

Mean : 8.742

Standard deviation : 6.488

Variable Format: numeric

Variable: **Reliability with cars (dealy in minutes)**

Location:

Width: 11

Value	Label	Frequency
5 .		2261
10 .		2266
20 .		4501

Range of Valid Data Values: 5 to 20

Total Responses: Summation of listed categories: 9028

Summary Statistics:

Minimum : 5

Maximum : 20

Mean : 13.733

Standard deviation : 6.495

Variable Format: numeric

Variable: **Employed**

Location:

Width: 4

Value	Label	Frequency
1 .	fultime	4566
2 .	parttime	1444
3 .	unemployed	3018

Total Responses: Summation of listed categories: 9028

Summary Statistics:

Variable Format: character

Variable: **Car availability**

Location:

Width: 4

Value	Label	Frequency
1 .	always	5519
2 .	after arrangement	1910
3 .	no	1599

Total Responses: Summation of listed categories: 9028

Summary Statistics:

Variable Format: character

Variable: **Age (in years)**

Location:

Width: 11

Value	Label	Frequency
15 .		32
16 .		143
17 .		105
18 .		119
19 .		168
20 .		136
21 .		148
22 .		96
23 .		160
24 .		72
25 .		102
26 .		136
27 .		135
28 .		152
29 .		72
30 .		223
31 .		96
32 .		165
33 .		208
34 .		190
35 .		219
36 .		182
37 .		184
38 .		279
39 .		186
40 .		277
41 .		159
42 .		295
43 .		132

44 .	200
45 .	183
46 .	121
47 .	167
48 .	118
49 .	163
50 .	194
51 .	160
52 .	119
53 .	172
54 .	143
55 .	144
56 .	207
57 .	117
58 .	156
59 .	120
60 .	188
61 .	135
62 .	92
63 .	103
64 .	80
65 .	212
66 .	135
67 .	119
68 .	125
69 .	103
70 .	86
71 .	48
72 .	88
73 .	81
74 .	38

75 .	102
77 .	32
78 .	32
79 .	24
80 .	79
81 .	16
82 .	15
83 .	8
84 .	32

Total Responses: Summation of listed categories: 9028

Summary Statistics:

Variable Format: character

Variable: **Gender**

Location:

Width: 4

Value	Label	Frequency
1 .	male	4739
2 .	female	4289

Total Responses: Summation of listed categories: 9028

Summary Statistics:

Variable Format: character

Variable: **Number of km in '000 travelled by car in the last year**

Location:

Width: 11

Value	Label	Frequency
0 .		1816
1 .		48
10 .		1113
11 .		79
12 .		463
13 .		56
14 .		63
15 .		965
16 .		16
17 .		64
18 .		112
2 .		40
20 .		670
22 .		24
23 .		7
24 .		15
25 .		327
27 .		16
3 .		102
30 .		342
35 .		112
38 .		8
4 .		56
40 .		64
45 .		8
5 .		260
50 .		88
6 .		159
60 .		32

7 .	222
70 .	8
72 .	8
74 .	8
75 .	8
8 .	301
80 .	8
9 .	158
90 .	8
99 .	96

Total Responses: Summation of listed categories: 7950

Summary Statistics:

Variable Format: character

Variable: **Season ticket availability**

Location:

Width: 4

Value	Label	Frequency
0 .	Without PT-card	3801
1 .	annual season tickets	951
2 .	Half-price-discount card	3737
3 .	Other PT-card	539

Range of Valid Data Values: 0 to 3

Total Responses: Summation of listed categories: 9028

Summary Statistics:

Variable Format: numeric

Variable: **Trip purpose**

Location:

Width: 4

Value	Label	Frequency
1 .	Commuters	1080
2 .	Busines	650
3 .	Shopping	1181
4 .	Leisure/ vacation	6117

Range of Valid Data Values: 1 to 4

Total Responses: Summation of listed categories: 9028

Summary Statistics:

Variable Format: numeric

Variable: **Language**

Location:

Width: 11

Value	Label	Frequency
0 .	others	2391
1 .	german	6637

Range of Valid Data Values: 0 to 1

Total Responses: Summation of listed categories: 9028

Summary Statistics:

Variable Format: numeric