Preferred citation style for this presentation

Schlich, R . (2001) Analysing intrapersonal variability of travel behaviour using the sequence alignment method, European Transport Conference, Cambridge, September 2001.

Analysing intrapersonal variability of travel behaviour using the sequence alignment method

Robert Schlich

IVT
ETH
Zürich

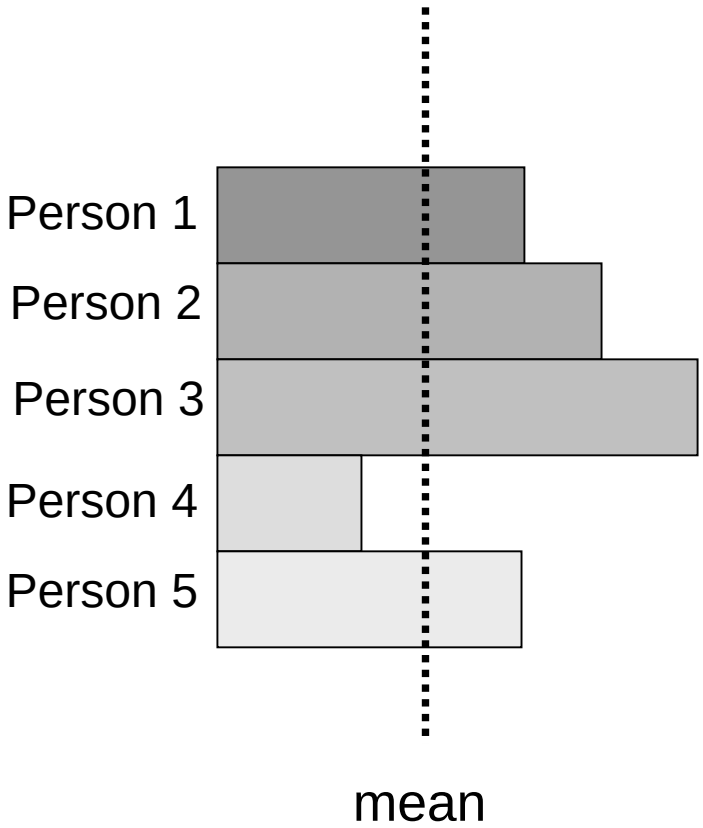
September 2001

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Intrapersonal Variability

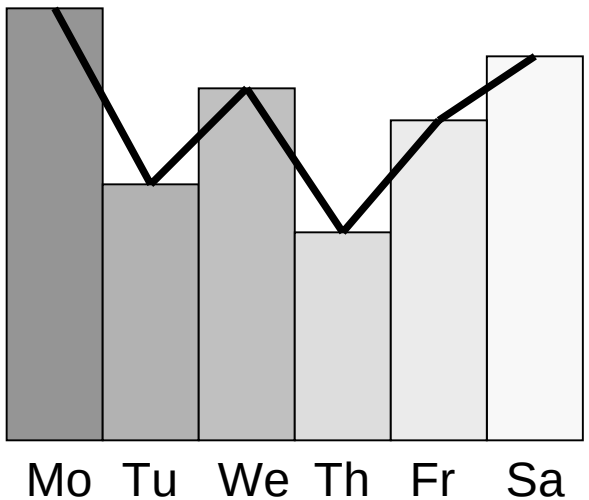
Inter-personal Variability

One single day

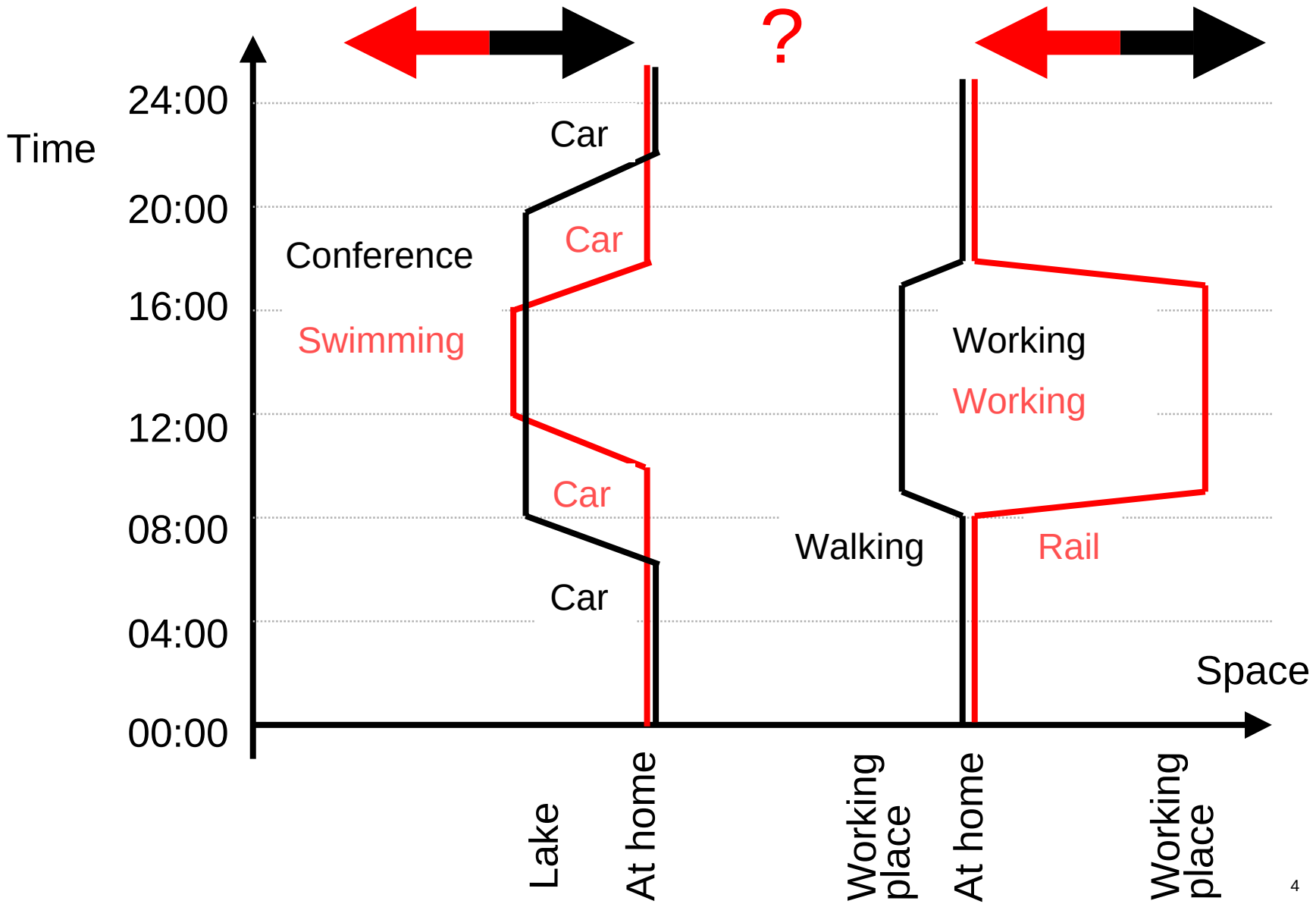


Intra-personal Variability

Person 1



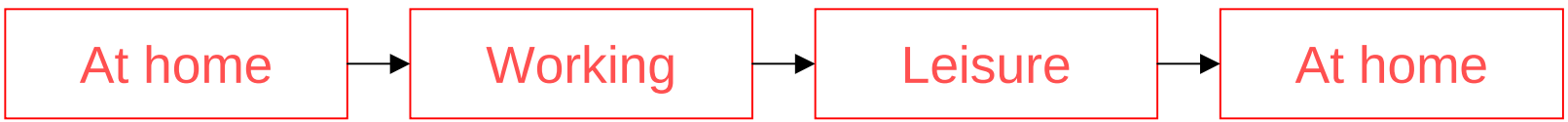
Measuring similarity



Trips and activities: different characteristics

Activity attributes:

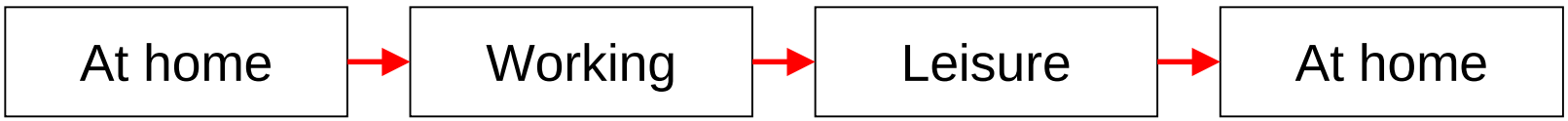
- Duration
- Location
- Activity type



Trips and activities: different characteristics

Activity attributes:

- Duration
- Location
- Activity type



Trip attributes:

- Mode choice
- Trip distance
- Trip duration
- Route choice

Theory of sequence alignment I

Measuring differences between two strings *s* [s_1, s_2,.....s_n] and *g* [g_1, g_2,.....g_n]

$d(s,g) = \sum_{i=1}^{n} f(x)$ and

$f(x) = 1$ if $s_i \neq g_i$

$f(x) = 0$ if $s_i = g_i$

Example:

s=ABCDE

g=AFBCDE

d(*s*,*g*)=4

→ Problem of recognising sequential order or duration

Theory of sequence alignment II: Levenshtein

Similarity as total amount of effort to equalise $s[s_1, s_2,.....s_n]$ and $g[g_1, g_2,.....g_n]$

Four basic operation:

- Identity: $w_e(s_i,g_i)=0$
- Insertion: $w_i(\varnothing,g_i)=1$
- Deletion: $w_d(s_i, \varnothing)=1$
- Substitution: $w_s(s_i,g_i)=w_d(s_i,g_i)+w_i(s_i,g_i)=2$

Definition Levenshtein Distance:

Smallest sum of operation weighting values required to change $s[s_1, s_2,.....s_n]$ into g $[g_1, g_2,.....g_n]$:

Theory of sequence alignment III: Trajectories

- Different possibilities to equal two strings
- Combination of operations are called trajectories

Example

s=CAMBRIDGE
g=CAMPING

1) substitute s_4(B:P), s_5(R:I), s_6(I:N), s_7(D:G) delete s_8(G), s_9(E)
   d=10
2) substitute s_4(B:P), delete s_5(R), substitute s_6(D:N), delete s_8 (E)
   d=6

Theory of sequence alignment IV: Problems

- Qualitative and quantitative data
- Cost of operation weights
- Including duration of activities as attribute or using equal time slices?
- Different attributes
  - Sum of „unidimensional“ sequence alignments across all attribute
  - Optimum trajectory based sequence alignment (Joh et al. 1999)

Software

Dana (C.H. Joh, Universiteit Eindhoven)

- multidimensional
- restricted number of allowed elements per string
- restricted possibilities to change operation weights

ClustalG (C. Wilson, A. Harvey, and J. Thompson)

- Unidimensional
- large strings allowed
- Better possibilities to change operation weights

Dataset Mobidrive

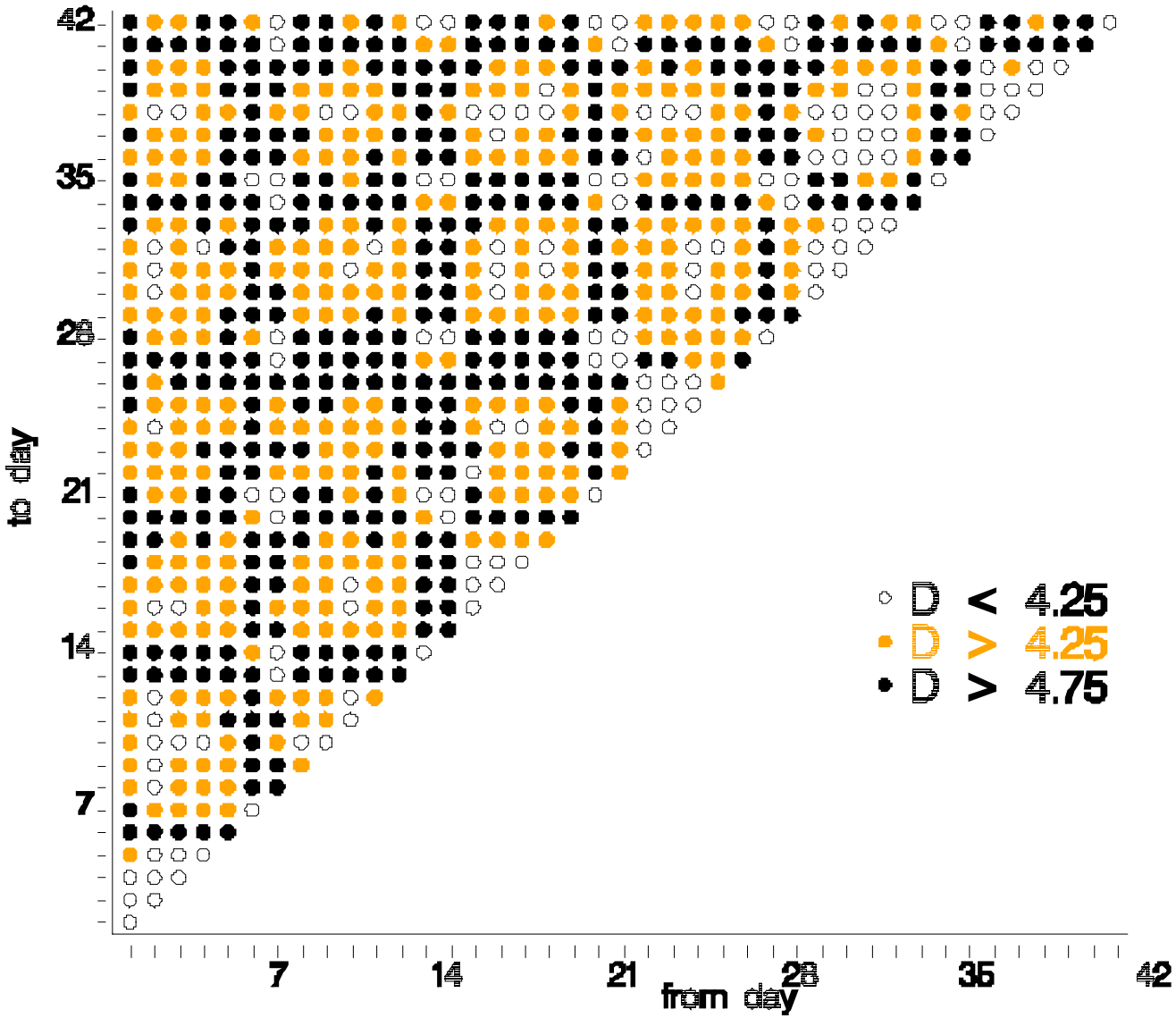
- Reporting period: six weeks
- Travel diary, weekly send out, mailed back and checked via phone
- Cities of Karlsruhe und Halle/Germany
- 162 households, 361 persons
- ca. 52.000 trips and 15.000 days reported September - November 1999 (Pretest: May-July 1999)
- used in analysis: City of Halle (159 persons, 21.000 trips)

Results: Intrapersonal variability over 42 days

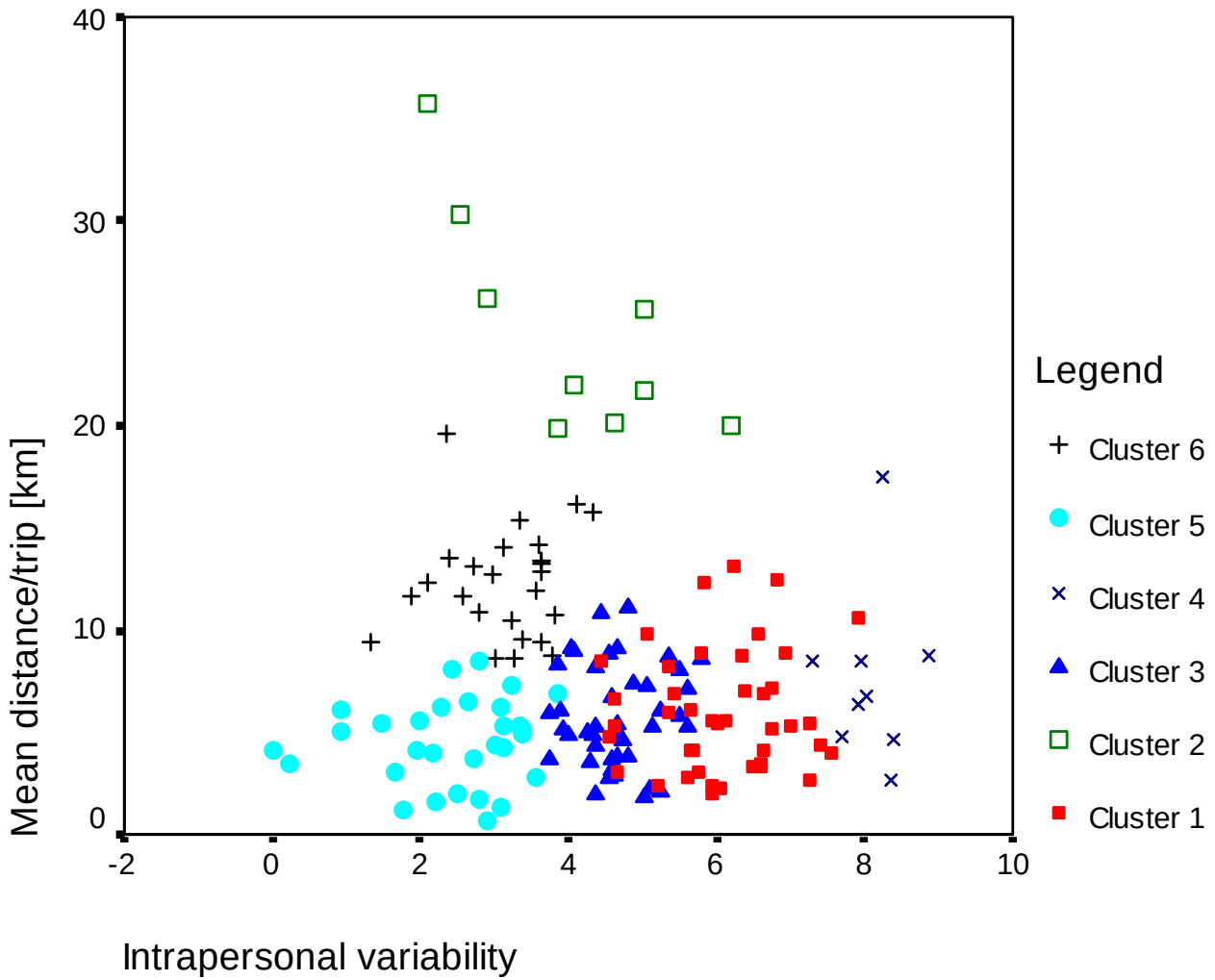
Mean daily Levensthein distance by licence ownership and age

	Not Licenced		Licenced		All	
	Mean	Std	Mean	Std	Mean	Std
18-24	4.5	2.6	5.6	1.7	5.4	1.8
25-34	4.8	1.2	5.3	1.8	5.2	1.7
35-44	5.9	1.9	5.0	1.9	5.1	1.9
45-54	4.0	2.0	4.5	2.0	4.4	2.0
55-64	2.9	1.3	4.6	1.5	4.0	1.7
65 and more	2.5	1.2	4.7	0.8	3.7	1.5
All	4.0	1.7	4.9	1.8	4.5	1.8

Results: Mean distance from different type of day



Cluster analysis



Behavioural characteristics of the clusters

	[Levensthein distance]	Mean trip distance [km]	Mean number of trip per day	Cluster size [n]
Cluster 1	6.1	6.0	4.6	42
Cluster 2	4.0	24.6	3.3	9
Cluster 3	4.7	5.9	3.4	43
Cluster 4	8.1	7.6	6.0	9
Cluster 5	2.4	4.5	2.5	30
Cluster 6	3.1	12.3	2.6	25
Overall	4.5	7.8	3.6	158

Sociodemographic characteristics of the clusters

	Cluster						
	1	2	3	4	5	6	Sum
Mean Age [years]	38.5	41.9	32.7	38.4	43.0	44.4	38.9
Proportion of females [%]	52.4	22.2	51.2	44.4	53.3	56.0	50.6
Proportion of parents [%]	47.6	55.6	9.3	55.6	23.3	48.0	33.5
Proportion of people without licences [%]	26.2	11.1	60.5	11.1	60.0	28.0	40.5
Proportion of fulltime employed people [%]	52.4	88.9	23.3	33.3	26.7	60.0	41.8

Summary

Day-to-day variability measured with multidimensional sequence alignment:

- Sociodemograpic characteristics as expected: highest variability for persons between 25 and 45 years with driving licence
- High variability between weekend days and week days

Clusters based on trip distance, number of trips per day and day to day variability

- 6 cluster solution
- Good differentiation in terms of travel characterisits
- Reasonable differences for the sociodemographic characterisitcs

Outlook

Further research: Sequence alignment

- Check for other operation costs (deletion, substitution, insertion)
- Check for other weights (Consideration of meaning of different attributes, e.g. mode choice)
- Consideration of duration of activities („long form“)

Further research: Travel behaviour

- Classification based on systematic and comprehensive description of travel behaviour
- Relevance for transport policy

Dataset: fatigue