

GESIS-Workshop
"Datenanalyse mit R"
DLR Berlin

Jan-Philipp Kolb

Freitag, 07. November, 2014



Inhaltsverzeichnis

Aufgabe in SPSS

Gliederung

Aufgabe in SPSS

Ausgangspunkt

Bsp.: mittlere Tagesstrecke mit dem Fahrrad zur Arbeit

Wegedatensatz: mittlere Wegelänge nach Verkehrsmittel und Zweck

MiD 2008

Abbildung in Deutschland

5

mittlere Wegelänge nach Hauptverkehrsmittel und Hauptwegezweck

		hvm Hauptverkehrsmittel				
		1 zu Fuß	2 Fahrrad	3 MIV (Mitfahrer)	4 MIV (Fahrer)	5 ÖPV
		wegkm_k Wegelänge [km] korrigiert	wegkm_k Wegelänge [km] korrigiert	wegkm_k Wegelänge [km] korrigiert	wegkm_k Wegelänge [km] korrigiert	wegkm_k Wegelänge [km] korrigiert
		Mittelwert	Mittelwert	Mittelwert	Mittelwert	Mittelwert
hvwzweck Hauptzweck des Weges	1 Arbeit	,9	3,5	27,1	19,7	25,9
	2 dienstlich	2,0	2,1	43,7	20,0	50,3
	3 Ausbildung	,9	2,4	8,9	23,6	11,9
	4 Einkauf	,8	2,0	9,8	6,7	8,3
	5 Erledigung	1,0	2,2	14,5	10,6	12,0
	6 Freizeit	2,0	4,7	25,0	19,1	32,1
	7 Begleitung	1,0	1,6	12,2	8,0	18,2

Quelle: MID2008_PUF_Wege (gewichtet)

Die durchschnittliche Länge eines Weges mit dem Fahrrad zur Arbeit beträgt 3,5 km.

Diese Info erhält man direkt aus dem Wegedatensatz.

Wie viele Kilometer fährt eine Person pro Tag im Mittel mit dem Fahrrad zur Arbeit?
(meist Hin- und Rückweg → die Strecke wird also fast doppelt so lang sein)



infas



Ausgangspunkt

Bsp. aus dem Wegedatensatz – Arbeitswege mit dem Fahrrad

MiD 2008

Mobilität in Deutschland

6

	hhid	pid	wsid	hwzweck	hvm	wegkm_k	
1	200816	3	1	Arbeit	Fahrrad	0,98	→ 1,96 km
2	200816	3	2	Arbeit	Fahrrad	0,98	
3	200816	3	3	Arbeit	MIV (Fahrer)	0,95	
4	200816	3	4	Arbeit	MIV (Fahrer)	0,95	
5	201435	1	1	Arbeit	Fahrrad	1,47	→ 2,94 km
6	201435	1	2	Arbeit	Fahrrad	1,47	
7	201435	1	3	Freizeit	MIV (Fahrer)	19,00	
8	201435	1	4	Freizeit	MIV (Fahrer)	19,00	
9	237431	2	1	Arbeit	Fahrrad	1,96	→ 3,92 km
10	237431	2	2	Arbeit	Fahrrad	1,96	
11	237431	2	3	Freizeit	Fahrrad	1,96	
12	237431	2	4	Freizeit	Fahrrad	1,96	
13	201874	1	1	Arbeit	Fahrrad	8,33	→ 16,66 km
14	201874	1	2	Arbeit	Fahrrad	8,33	

- Für jede Person müssen die Wegelängen für Arbeitswege mit dem Fahrrad aufsummiert werden und dann dem Personendatensatz zugespielt werden.



infas



Sozialwissenschaftliches
Institut für
Quantitative Sozialforschung

Ausgangspunkt

Durchführung mit SPSS – Filter setzen

MID 2008
Mobilität in Deutschland

7

- Im Wagedatensatz muss zunächst ein **Filter** gesetzt werden, der die Arbeitswege mit dem **Fahrrad** auswählt.
 - Arbeit: hzwzweck = 1
 - Fahrrad: hvm = 2

SPSS-Menü: Daten → Fälle auswählen → Falls Bedingung zutrifft
→ Falls: hzwzweck = 1 & hvm = 2

SPSS-Syntax:

```
USE ALL.  
COMPUTE filter_$=(hzwzweck = 1 & hvm = 2).  
VARIABLE LABEL filter_$ 'hzwzweck = 1 & hvm = 2 (FILTER)'.  
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.  
FORMAT filter_$ (f1.0).  
FILTER BY filter_$.  
EXECUTE.
```

➤ **Achtung:**

- Die **Gewichtung** muss **ausgeschaltet** sein: **WEIGHT OFF**.
- **Missings** bei der **Wegelänge** müssen als **fehlend** deklariert sein

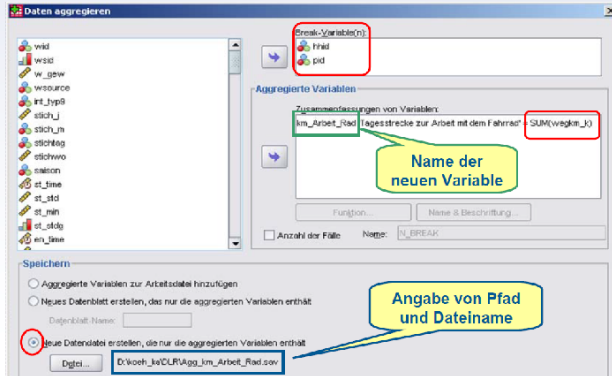
Ausgangspunkt

Durchführung mit SPSS – Daten aggregieren

MID 2008
Mikroskop in Deutschland

8

- Für jede Person werden die Wegelängen aufsummiert.
 - SPSS-Menü: Daten → Aggregieren



Ausgangspunkt

Durchführung mit SPSS – Daten aggregieren

MiD 2008

Modeller in Deutschland

9

- Für jede Person werden die Wegelängen aufsummiert.

- SPSS-Syntax: **AGGREGATE**

```

/OUTFILE='Pfad\Agg_km_Arbeit_Rad.sav'
/BREAK=hhid pid
/km_Arbeit_Rad 'Tagesstrecke zur Arbeit mit dem Fahrrad'=SUM(wegkm_k).

```

- Die neu erstellte Variable km_Arbeit_Rad (Datensatz: Agg_km_Arbeit_Rad.sav) enthält 4 Sysmis. Ursache: keine Kilometerangabe vorhanden

	hhid	pid	wsid	hwzweck	hvm	wegkm_k		hhid	pid	km_Arbeit_Rad
	262101	3	1	Arbeit	Fahrrad	weiß nicht		262101	3	
	262101	3	2	Arbeit	Fahrrad	weiß nicht				

- Je nach weiterem Auswertungsziel können die Sysmis umkodiert werden:

```

RECODE km_Arbeit_Rad (SYSMIS=9995).
EXECUTE.
MISSING VALUES km_Arbeit_Rad (9995).
VALUE LABELS km_Arbeit_Rad
9995 'Wert nicht zu berechnen'.

```


Ausgangspunkt

Durchführung mit SPSS – Dateien zusammenfügen

Modellbau in Deutschland

10

- Die Variable km_Arbeit_Rad (Datensatz: Agg_km_Arbeit_Rad.sav) muss nun dem Personendatensatz zugespielt werden.

- Beide Dateien müssen in der gleichen Reihenfolge sortiert sein!

→ SORT CASES BY hhid(A) pid(A).

- SPSS-Menü:

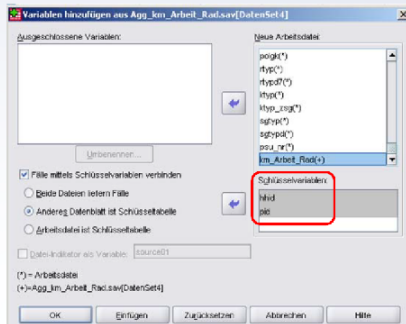
→ Daten

→ Dateien zusammenfügen

→ Variablen hinzufügen

- SPSS-Syntax:

```
MATCH FILES /FILE=*  
/TABLE='DatenSet4'  
/BY hhid pid.  
EXECUTE.
```



Ausgangspunkt

Auswertungen im Personendatensatz

 MID 2000
 Multiple in Deutschland

11

- Die Variable `km_Arbeit_Rad` enthält im Personendatensatz nur bei den Personen einen Wert, die auch am Stichtag einen Weg mit dem Fahrrad zur Arbeit angegeben haben. Bei allen anderen Personen steht kein Wert (Sysmis).
 ➤ Sysmis-Kodierung analog der Variablen `anzkm` (Tagesstrecke) möglich.

- Ausgabe der mittleren Tagesstrecke pro Person mit dem Fahrrad zur Arbeit

- SPSS-Menü:

- Analysieren
- Deskriptive Statistiken
- Häufigkeiten
- Statistiken: Mittelwert

- SPSS-Syntax:

```
WEIGHT BY p_gew.  
FREQUENCIES VARIABLES=km_Arbeit_Rad  
/FORMAT=NOTABLE  
/STATISTICS=MEAN  
/ORDER=ANALYSIS.  
WEIGHT OFF.
```

Statistiken

km_Arbeit_Rad Tagesstrecke zur Arbeit mit dem Fahrrad		
N	Gültig	1427
	Fehlend	59286
	Mittelwert	6,8

	hhid	pid	km_Arbeit_Rad
	200816	3	1,96
	201435	1	2,94
	237431	2	3,92
	201874	1	16,66