

GESIS-Workshop
"Datenanalyse mit R"
DLR Berlin

Jan-Philipp Kolb

Freitag, 07. November, 2014



Inhaltsverzeichnis

Aufgabe in SPSS

Ausgangstabelle erzeugen

Variable erzeugen

Gliederung

Aufgabe in SPSS

Ausgangstabelle erzeugen

Variable erzeugen

Ausgangspunkt

Führerscheinanteile nach Alter und Geschlecht an einen (anderen) Datensatz anlesen

- ▶ **Ziel:** bei generierten Bevölkerung mit Eigenschaften Alter und Geschlecht)
- ▶ die Eigenschaft Führerschein vorhanden ja/nein anzulesen nach den Prozents, die aus dem MiD-Datensatz vorliegen (siehe Tabelle)

Ausgangstabelle

MiD2008-Personendatensatz, gewichtet mit <u>p_gew</u>				<u>hp_pkwfs</u> Pkw-Führerschein (aus p061_3, ergänzt um Angabe aus HH-Int.)		
				1 ja	2 nein	Gesamt
				Anzahl der Zeilen (%)	Anzahl der Zeilen (%)	Anzahl der Zeilen (%)
<u>hp_sex</u> Geschlecht (fehlende Angaben ergänzt aus <u>HH-Interview</u>)	1 männlich	<u>hp_altg3</u> Altersgruppen (Variante 3)	5 18 - 29	86%	14%	100%
			6 30 - 39	95%	5%	100%
			7 40 - 49	97%	3%	100%
			8 50 - 59	96%	4%	100%
			9 60 - 64	95%	5%	100%
			10 65 - 74	95%	5%	100%
			11 75 und älter	88%	12%	100%
	2 weiblich	<u>hp_altg3</u> Altersgruppen (Variante 3)	5 18 - 29	87%	13%	100%
			6 30 - 39	94%	6%	100%
			7 40 - 49	95%	5%	100%
			8 50 - 59	89%	11%	100%
			9 60 - 64	86%	14%	100%
			10 65 - 74	71%	29%	100%
			11 75 und älter	44%	56%	100%

Daten einlesen

Das `.RData`-File wurde zuvor erzeugt:

```
setwd(SPSS.path)  
(load("MIDPERS.RData"))
```

Gliederung

Aufgabe in SPSS

Ausgangstabelle erzeugen

Variable erzeugen

Werte vorbereiten

```
# Variablen  
MIDPERS$fs <- rep(0,nrow(MIDPERS))  
MIDPERS$fs[MIDPERS$hp_pkwfs=="ja"] <- 1  
MIDPERS$hp_sex <- as.character(MIDPERS$hp_sex)  
  
MIDPERS$p_gew <- as.numeric(MIDPERS$p_gew)
```


Tabelle erzeugen

```
Tab0 <- with(MIDPERS ,  
             tapply(p_gew ,  
                   list (hp_altg3 , hp_pkwfs , hp_sex) ,  
                   sum , na.rm=T)  
)
```

Tabelle erzeugen

```
Tab0 <- Tab0[-(1:4), -c(2,4,5), -1]  
  
Tab0[, ,1] <- Tab0[, ,1]/rowSums(Tab0[, ,1])  
Tab0[, ,2] <- Tab0[, ,2]/rowSums(Tab0[, ,2])
```

Führerschein vorhanden - Männer

	ja	nein
18 - 29	0.86	0.14
30 - 39	0.95	0.05
40 - 49	0.97	0.03
50 - 59	0.96	0.04
60 - 64	0.95	0.05
65 - 74	0.95	0.05
75 und älter	0.88	0.12

Führerschein vorhanden - Frauen

	ja	nein
18 - 29	0.87	0.13
30 - 39	0.94	0.06
40 - 49	0.95	0.05
50 - 59	0.89	0.11
60 - 64	0.86	0.14
65 - 74	0.71	0.29
75 und älter	0.44	0.56

Gliederung

Aufgabe in SPSS

Ausgangstabelle erzeugen

Variable erzeugen

Variable erzeugen

Die Variable Führerschein (fs) erzeugen:

```
N <- nrow(MIDPERS)
fs <- rep(0,N)
```

Variable erzeugen

Nach Häufigkeit im Personendatensatz:

```
ind1 <- MIDPERS$hp_sex==1 & MIDPERS$hp_altg3=="18 - 29"  
fs[ind1] <- sample(0:1,N,prob=c(.86,.14),replace=T)
```

Variable erzeugen

Etwas mehr automatisiert:

```
ind1 <- MIDPERS$hp_sex==1 & MIDPERS$hp_altg3=="18 - 29"  
fs[ind1] <- sample(0:1,N,prob=c(Tab0[1,1,1],Tab0[1,2,1]),  
                  replace=T)
```


Variable erzeugen

Für die zweite Altersgruppe:

```
ind2 <- MIDPERS$hp_sex==1 & MIDPERS$hp_altg3=="30 - 39"  
fs[ind2] <- sample(0:1,N,prob=c(.95,.05),replace=T)
```

Variable erzeugen - Schleife

Für die Männer `sex==1`

```
Tab <- data.frame(Tab0)

fs <- vector()

for(i in 1:nrow(Tab)){
  indi <- MIDPERS$hp_sex==1 &
    MIDPERS$hp_altg3==rownames(Tab)[i]
  fs[indi] <- sample(0:1,N,prob=c(Tab[i,1],Tab[i,2]),
replace=T)
}
```

Variable erzeugen - Schleife

Für die Frauen `sex==2`

```
for(i in 1:nrow(Tab)){  
  indi <- MIDPERS$hp_sex==2 &  
    MIDPERS$hp_altg3==rownames(Tab)[i]  
  fs[indi] <- sample(0:1,N,prob=c(Tab[i,3],Tab[i,4]),  
replace=T)  
}
```