

# Übungsaufgabe 1

Fabio Votta

2. November 2017

## Aufgabe 1

### Aufgabe 1a

Was ist unter Auspartialisierung zu verstehen und wieso ist es aufgrund der beteiligten Mechanismen wichtig immer mehrere Prädiktorvariablen zu berücksichtigen, auch wenn diese ggf. keinen Einfluss auf die abhängige Variable haben?

-> [ANTWORT HIER EINTRAGEN]

### Aufgabe 1b

Wieso können unabhängige Variablen ( $x_i$ ) im multiplen Regressionsmodell einen Einfluss auf  $Y$  haben, obwohl die bivariate Korrelation zwischen ihnen und  $Y$  nicht signifikant ist?

-> [ANTWORT HIER EINTRAGEN]

## Aufgabe 2

Bevor Sie die Analysen durchführen, suchen Sie im Codebuch (o. Variablenliste) Ihres Datensatzes (ALLBUS 2014) am besten Mittels STRG+F (aufrufen der "Suchenfunktion" in nahezu allen Programmen) die folgenden Variablen heraus: Alter, Geschlecht, Schulabschluss und individuelles Nettoeinkommen in der Fassung "Offene Angaben+Listeangaben".

Kodieren Sie dann diese Variablen wie folgt:

- Alter: Startwert auf 0 setzen; 18=0, 48=30
- Schulabschluss- bzw. Schulbildung: 5 Ausprägungen; 0=kein Schulabschluss, 1=HS, 2=RS, 3=FHR, 4=Abi; Rest=-1 bzw. Missing
- Geschlecht: 0=weiblich; 1=männlich

### 1. Schritt: Datensatz einladen

```
allbus <- read_spss("allbus2014.sav")  
  
head(allbus)
```

### 2. Schritt: relevante Variablen identifizieren

```
var_names(allbus, "alter") # V84 ALTER: BEFRAGTE<R>  
var_names(allbus, "schulabschluss") # V86 ALLGEMEINER SCHULABSCHLUSS  
var_names(allbus, "geschl") # V81 GESCHLECHT, BEFRAGTE<R>  
var_names(allbus, "eink") # V420 NETTOEINKOMMEN<OFFENE+LISTENANGABE>, KAT.
```

### 3. Schritt: Jetzt wählen wir die Variablen und erstellen ein Subset!

```
allb_sub <- select(allbus, V84, V86, V81, V420)
```

### 4. Schritt: Als nächstes benennen wir die Variablen um!

```
allb_sub <- rename(allb_sub,
  alter = V84,
  bildung = V86,
  geschl = V81,
  einkommen = V420)
```

## 5. Schritt: Als nächstes Rekodieren wir die Variablen

```
allb_sub <- mutate(allb_sub,
  alter0 = alter - 18,
  bildung_rec = ifelse(bildung == 6 | bildung == 7, NA, bildung - 1),
  geschl_rec = ifelse(geschl == 2, 0, 1))
```

```
allb_sub
```

## A tibble: 3,471 x 7

```
      alter  bildung  geschl einkommen  alter0 bildung_rec
1 33 5 2 14 15 4 2 50 3 2 9 32 2 3 56 3 1 17 38 2 4 61 3 1 8 43 2 5 59 3 2 9 41 2 6 56 3 1 21 38 2 7 66 2 1 12
48 1 8 25 5 2 NA 7 4 9 58 4 1 13 40 3 10 53 3 1 19 35 2 # ... with 3,461 more rows, and 1 more variables:
geschl_rec
```

### Bonus: Alles mit dem pipe operator %>%

```
allb_sub <- allbus %>%
  select(V84,V86, V81, V420) %>%
  rename(alter = V84,
    bildung = V86,
    geschl = V81,
    einkommen = V420) %>%
  mutate(alter0 = alter - 18,
    bildung_rec = ifelse(bildung == 6 | bildung == 7, NA, bildung - 1),
    geschl_rec = ifelse(geschl == 2, 0, 1)) %>%
  select(einkommen, alter0, geschl_rec, bildung_rec)
```

```
allb_sub
```

## A tibble: 3,471 x 4

```
einkommen alter0 geschl_rec bildung_rec 1 14 15 0 4 2 9 32 0 2 3 17 38 1 2 4 8 43 1 2 5 9 41 0 2 6 21 38 1 2
7 12 48 1 1 8 NA 7 0 4 9 13 40 1 3 10 19 35 1 2 # ... with 3,461 more rows
```

## Aufgabe 3

Berechnen Sie folgende (sequentielle) Regressionsmodelle:

- Modell a: Einkommen auf Alter;
- Modell b: Einkommen auf Bildung;
- Modell c: Einkommen auf Geschlecht;

- Modell ab: Einkommen auf Alter und Bildung;
- Modell abc: Einkommen auf Alter, Bildung und Geschlecht.

```
modell_a <- lm(einkommen ~ alter0, data = allb_sub)
modell_b <- lm(einkommen ~ bildung_rec, data = allb_sub)
modell_c <- lm(einkommen ~ geschl_rec, data = allb_sub)

modell_ab <- lm(einkommen ~ alter0 + bildung_rec, data = allb_sub)
modell_abc <- lm(einkommen ~ alter0 + bildung_rec + geschl_rec, data = allb_sub)

texreg(list(modell_a,
            modell_b,
            modell_c,
            modell_ab,
            modell_abc),
        float.pos = "ht!")
```

	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	10.53*** (0.19)	8.75*** (0.19)	9.33*** (0.12)	7.17*** (0.28)	5.15*** (0.28)
alter0	0.02*** (0.01)			0.04*** (0.01)	0.04*** (0.00)
bildung_rec		1.05*** (0.07)		1.20*** (0.07)	1.24*** (0.07)
geschl_rec			3.47*** (0.17)		3.56*** (0.16)
R <sup>2</sup>	0.00	0.07	0.12	0.08	0.21
Adj. R <sup>2</sup>	0.00	0.06	0.12	0.08	0.21
Num. obs.	3064	3040	3065	3039	3039
RMSE	4.95	4.78	4.65	4.74	4.40

\*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$

Table 1: Statistical models

### Aufgabe 3a

Vergleichen Sie die Regressionskoeffizienten über die Modelle und erläutern Sie was hier festzustellen ist!

-> [ANTWORT HIER EINTRAGEN]

### Aufgabe 3b

Vergleichen Sie  $R^2$  über die Modelle und erläutern Sie was hier festzustellen ist!

-> [ANTWORT HIER EINTRAGEN]