

SM II Abgabe 1

Robin Gerl

31. Oktober 2018

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?

Auspartialisierung ist ein Term in der Empirischen Statistik. Bei der Auspartialisierung werden Varianzen von zwei verschiedenen Unabhängigen Variablen rausgerechnet. Diese rausgerechneten Varianzen sind von den Variablen Faktoren, die gemeinsam einen Einfluss auf die abhängige Variable besitzen.

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?

Weil die Beziehungen zwischen Variablen komplex sind. Befunde aus den bivariaten Korrelation können durch Kontrollvariablen beeinflusst werden. In diesem Fall die wahre Einflusswirken überdeckt sein von Suppressoreffekten. Durch Kontrollvariablen könnte dieser wahre Effekt ermittelt werden.

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

```
library(haven)
allbus <- read_sav("allbus.sav")
View(allbus)
```

2. Schritt: relevante Variablen identifizieren

- V84 ALTER: BEFRAGTE
- V86 ALLGEMEINER SCHULABSCHLUSS
- V81 GESCHLECHT, BEFRAGTE
- V420 NETTOEINKOMMEN,KAT.

```
# binocular(allbus) # zum Datensatz inspizieren :)
#binocular(allbus)
```

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

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

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

```
allbussub <- rename(allbus, Alter = V84,
                    Bildung = V86,
                    Geschl= V81,
                    einkommen = V420)
```

5. Schritt: Als nächstes Rekodieren wir die Variablen

```
allbussub <- mutate(allbussub, alt <- Alter-18,
                    bild = ifelse(Bildung == 6 | Bildung == 7, NA, Bildung-1),
                    ges = ifelse(Geschl == 2, 0, 1))
```

Bonus: Alles mit dem pipe operator %>%

```
allbussub <- allbus %>%
  select(V84,V86,V81,V420) %>%
  rename(Alter=V84, Bildung=V86, geschl=V81, einkomm=V420) %>%
  mutate(alt=Alter-18,
         bild = ifelse(Bildung==6|Bildung==7,NA,Bildung-1),
         ges = ifelse(geschl == 2,0,1))
```

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.

```

ma <-lm (einkomm~alt,data = allbussub)
mb <-lm (einkomm~bild, data=allbussub)
mc <-lm (einkomm~ges,data=allbussub)
mab <-lm(einkomm~alt+bild, data=allbussub)
mabc <-lm(einkomm~alt+bild+ges, data=allbussub)

screenreg(list(ma,mb,mc,mab,mabc))

```

```

=====
Model 1 Model 2 Model 3 Model 4 Model 5
----- (Intercept) 10.53 *** 8.75 *** 9.33 *** 7.17 ***
5.15 (0.19) (0.19) (0.12) (0.28) (0.28)
alt 0.02 0.04 *** 0.04 (0.01) (0.01) (0.00)
bild 1.05 1.20 *** 1.24 (0.07) (0.07) (0.07)
ges 3.47 3.56 (0.17) (0.16)
----- R^2 0.00 0.07 0.12 0.08 0.21
Adj. R^2 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

```

Aufgabe 3a

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

In allen Modellen ist der Regressionskoeffizient der Variable signifikant, aber besitzt mit 0,02 bis 0,04 keine Aussagekraft. Das bedeutet dass, das Alter nicht mit dem Einkommen zusammenhängt. Die Bildung wiederum besitzt einen signifikanten positiven Effekt auf die Variable Einkommen. Dieser Effekt wird sogar höher wenn man auf die Variablen Alter und Geschlecht kontrolliert. Das Geschlecht hat den stärksten Effekt auf Einkommen, der ebenfalls signifikant ist.

Aufgabe 3b

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

Das R^2 der Modelle steigt mit jeder hinzugefügten Variable an. Den größten R^2 anstieg besitzt das Modell 3 welches alleine 12% der Varianz erläutert. Eine höhere Erklärungskraft besitzt nur das Modell 5, welches alle Variablen beinhaltet.

Bonus:

Für die Visualisierung des Modell können wir `plot_model` aus dem `sjPlot` package benutzen.