

```

---
title: "SM II Abgabe 1"
author: "Fabio Votta"
date: "25.Oktober 2018"
output: pdf_document
---

```{r packages, include=FALSE, echo=F}
install.packages("pacman") #muss einmal installiert werden
install.packages("devtools") #muss einmal installiert werden
devtools::install_github("favstats/binocularR") # zum Datensatz inspizieren :)

pacman::p_load(tidyverse, haven, texreg, binocularR, rmarkdown, knitr)

```

## 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 wird in einer multivariaten Regression angewendet, um sicher zu sein, dass die unabhängige (X) Variablen keine Wirkungseffekte mit einander haben. Bivariate Regressionen zwischen alle X-Variablen wurden durchgeführt, und nur der unerklärte Effekt wird in der multivariate Regression berücksichtigt.



### 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?*



Es kann sein, dass eine bestimmte X-Variable einen Einfluss auf Y hat, aber nur wahrgenommen sein kann, wenn den Einfluss einer dritten Variable berücksichtigt ist. Es liegt daran, dass die erste unabhängige Variable X1 eine schwache Wirkung auf Y aber eine starke Wirkung auf die zweite unabhängige Variable X2 hat. Die Variable X2 spielt die Rolle einer Kontrollvariable.



## 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**
```

```
```{r results='hide'}
allbus <- read_sav("data/allbus2014.sav")

allbus %>% select(V1:V15) %>% head() %>% kable("html")

head(allbus)
```
```

```
**2. Schritt: relevante Variablen identifizieren**
```

```
+ V84    ALTER: BEFRAGTE<R>
+ V86    ALLGEMEINER SCHULABSCHLUSS
+ V81    GESCHLECHT, BEFRAGTE<R>
+ V420   NETTOEINKOMMEN<OFFENE+LISTENANGABE>,KAT.

```{r}
#binocularR(allbus) # zum Datensatz inspizieren :)

sollte auskommentiert sein wenn man das Dokument "knitted"
Fehler kann ignoriert werden: object 'datatables_html' not found
```
```

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

```
Tipp: Nutze `select`
```

```
```{r}

select(allbus, V84, V86, V81)

```
```

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

```
Tipp: Nutze `rename`
```

```
```{r}

rename(allbus, V84_new = V84,
V86_new = V86,
V81_new = V81)

```
```

```
**5. Schritt: Als nächstes Rekodieren wir die Variablen**
```

```
Tipp: `mutate` und `ifelse` machen die Aufgabe einfacher :)
```

```
```{r}

mutate(allbus, V84_new = V84_new - 18,
V86_new = ifelse(V86_new == 1 | V86_new == 2 | V86_new == 3 | V86_new == 4 | V86_new
== 5 , V86_new == V86_new - 1 , V86_new == Missing),
V81_new = ifelse(V81_new == 1, V81_new == 1, V81_new == 2))
```

```
```
```

```
**Bonus: Alles mit dem pipe operator `%>%`**
```

```
```{r}
```

```
```
```

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.*

`lm` ist die Funktion für lineare Regression

```
```{r, results='asis'}
```

```
```
```

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

```
GET
FILE='C:\Users\Abby\Documents\SM Übungen\02_r\smII_aufgaben\data\allbus2014.sav'.
```

```
>Avertissementn° 5281. Nom de la commande : GET FILE
>SPSS Statistics s'exécute en mode de codage Unicode. Le fichier utilise u
ne
>page de codes spécifique à l'environnementlocal. La largeur définie pour
les
>variables de chaîne est automatiquementmultipliéepar trois afin d'éviter
>toute perte de données. Vous pouvez faire appel à ALTER TYPE pour définir
la
>largeur des variables de chaîne sur la plus grande valeur observée parmi l
es
>chaînes de variable.
ALTER TYPE ALL(A=AMIN).
```

Modifier le type

C:\Users\Abby\Documents\SM Übungen\02_r\smII_aufgaben\data\allbus2014.sav

Types modifiés

| | | |
|---------------------------|-----|------|
| DIGITAL OBJECT IDENTIFIER | A75 | AMIN |
| RELEASE | A75 | AMIN |

```
DATASET NAME Jeu_de_données1WINDOW=FRONT.
* Encoding: UTF-8.
***** Aufgabe 1 *****
```

*1. Schritt: Variablen auswählen.

*Alter.

```
COMPUTE alterneu=V84 - 18.
EXECUTE.
```

*Schulabschluss

```
Recode v86 (1=0) (2=1) (3=2) (4=3) (5=4) (ELSE=SYSMIS).
EXECUTE.
```

*Geschlecht.

Recode v81 (2=0).
EXECUTE.

*Einkommen.

*2. Schritt: Rekodieren.

Aufgabe 3

*Model a.

REGRESSION
 /MISSING LISTWISE
 /STATISTICS COEFF OUTS R ANOVA
 /CRITERIA=PIN(.05) POUT(.10)
 /NOORIGIN
 /DEPENDENT V417
 /METHOD=ENTER alterneu
 /METHOD=ENTER V86
 /METHOD=ENTER V81
 /METHOD=ENTER alterneu v86
 /METHOD=ENTER alterneu v86 v81.

Régression

[Jeu_de_données1] C:\Users\Abby\Documents\SM Übungen\02_r\smII_aufgaben\data\allbus2014.sav

Variables introduites/éliminées^a

| Modèle | Variables introduites | Variables éliminées | Méthode |
|--------|--|---------------------|------------|
| 1 | alterneu ^b | . | Introduire |
| 2 | ALLGEMEINE R SCHULABSC HLUSS ^b | . | Introduire |
| 3 | GESCHLECH T, BEFRAGTE< R> ^b | . | Introduire |

a. Variable dépendante : BEFR.: NETTOEINKOMMEN, OFFENE ABFRAGE

b. Toutes les variables demandées ont été introduites.

Récapitulatif des modèles

| Modèle | R | R-deux | R-deux ajusté | Erreur standard de l'estimation |
|--------|-------------------|--------|---------------|---------------------------------|
| 1 | ,033 ^a | ,001 | ,001 | 1564,605 |
| 2 | ,189 ^b | ,036 | ,035 | 1537,658 |
| 3 | ,302 ^c | ,091 | ,090 | 1492,892 |

a. Prédicteurs : (Constante), alterneu

b. Prédicteurs : (Constante), alterneu, ALLGEMEINER SCHULABSCHLUSS

c. Prédicteurs : (Constante), alterneu, ALLGEMEINER SCHULABSCHLUSS , GESCHLECHT, BEFRAGTE<R>

ANOVA^a

| Modèle | | Somme des carrés | ddl | Carré moyen | F | Sig. |
|--------|------------|------------------|------|-------------|--------|-------------------|
| 1 | Régression | 7269146,274 | 1 | 7269146,274 | 2,969 | ,085 ^b |
| | de Student | 6602225252 | 2697 | 2447988,599 | | |
| | Total | 6609494399 | 2698 | | | |
| 2 | Régression | 235096522,9 | 2 | 117548261,5 | 49,716 | ,000 ^c |
| | de Student | 6374397876 | 2696 | 2364390,903 | | |
| | Total | 6609494399 | 2698 | | | |
| 3 | Régression | 603074089,3 | 3 | 201024696,4 | 90,197 | ,000 ^d |
| | de Student | 6006420309 | 2695 | 2228727,388 | | |
| | Total | 6609494399 | 2698 | | | |

a. Variable dépendante : BEFR.: NETTOEINKOMMEN, OFFENE ABFRAGE

b. Prédicteurs : (Constante), alterneu

c. Prédicteurs : (Constante), alterneu, ALLGEMEINER SCHULABSCHLUSS

d. Prédicteurs : (Constante), alterneu, ALLGEMEINER SCHULABSCHLUSS, GESCHLECHT, BEFRAGTE<R>

Coefficients^a

| Modèle | | Coefficients non standardisés | | Coefficients standardisés | t |
|--------|----------------------------|-------------------------------|-----------------|---------------------------|--------|
| | | B | Erreur standard | Bêta | |
| 1 | (Constante) | 1456,796 | 62,379 | | 23,354 |
| | alterneu | 2,974 | 1,726 | ,033 | 1,723 |
| 2 | (Constante) | 721,574 | 96,789 | | 7,455 |
| | alterneu | 7,862 | 1,768 | ,088 | 4,448 |
| | ALLGEMEINER SCHULABSCHLUSS | 251,837 | 25,655 | ,193 | 9,816 |
| 3 | (Constante) | 311,797 | 99,235 | | 3,142 |
| | alterneu | 8,074 | 1,716 | ,090 | 4,705 |
| | ALLGEMEINER SCHULABSCHLUSS | 259,003 | 24,915 | ,199 | 10,396 |
| | GESCHLECHT, BEFRAGTE<R> | 739,438 | 57,547 | ,236 | 12,849 |

Coefficients^a

| Modèle | | Sig. |
|--------|----------------------------|------|
| 1 | (Constante) | ,000 |
| | alterneu | ,085 |
| 2 | (Constante) | ,000 |
| | alterneu | ,000 |
| | ALLGEMEINER SCHULABSCHLUSS | ,000 |
| 3 | (Constante) | ,002 |
| | alterneu | ,000 |
| | ALLGEMEINER SCHULABSCHLUSS | ,000 |
| | GESCHLECHT, BEFRAGTE<R> | ,000 |

a. Variable dépendante : BEFR.: NETTOEINKOMMEN, OFFENE ABFRAGE

Variables exclues^a

| Modèle | | Bêta In | t | Sig. | Corrélation partielle |
|--------|----------------------------|-------------------|--------|------|-----------------------|
| 1 | ALLGEMEINER SCHULABSCHLUSS | ,193 ^b | 9,816 | ,000 | ,186 |
| | GESCHLECHT, BEFRAGTE<R> | ,232 ^b | 12,376 | ,000 | ,232 |
| 2 | GESCHLECHT, BEFRAGTE<R> | ,236 ^c | 12,849 | ,000 | ,240 |

Variables exclues^a

| Modèle | | Statistiques de colinéarité Tolérance |
|--------|----------------------------|--|
| 1 | ALLGEMEINER SCHULABSCHLUSS | ,921 |
| | GESCHLECHT, BEFRAGTE<R> | 1,000 |
| 2 | GESCHLECHT, BEFRAGTE<R> | ,999 |

a. Variable dépendante : BEFR.: NETTOEINKOMMEN, OFFENE ABFRAGE

b. Prédicteurs dans le modèle : (Constante), alterneu

c. Prédicteurs dans le modèle : (Constante), alterneu, ALLGEMEINER SCHULABSCHLUSS