

# SM II Abgabe 2

Fabio Votta

9. November 2018

## Aufgabe 1

Wozu werden Standardisierungen durchgeführt und wie wird dabei vorgegangen? Erläutern Sie zudem exemplarisch wozu  $b^*$  benutzt wird und wie man diesen interpretiert!

$$b^* = b * \frac{s_x}{s_y}$$

## Aufgabe 2

Führen Sie eine z-Standardisierung für die Originalaltersvariable ( $alter\_z$ ) und die auf Null gesetzte Altersvariable ( $alter\_0z$ ) sowie für “unsere” Bildungsvariable (0 bis 4). [Daten: ALLBUS 2014]

```
allb_sub_z <- allb_sub %>%
  select(einkommen, alter, alter0, geschl_rec, bildung_rec) %>%
  mutate(alter_z = scale(alter),
         alter0_z = scale(alter0),
         bildung_z = scale(bildung_rec),
         einkommen_z = scale(einkommen))
```

## Aufgabe 2a

Vergleichen Sie die Zahlenwerte, Mean und die Standardabweichung von  $alter_z$  und  $alter\_0z$  und erklären Sie Ihre “Beobachtung”.

```
allb_sub_z %>%
  select(alter_z, alter0_z) %>%
  describe() %>%
  select(-vars, -range, -trimmed, -mad, -skew, -kurtosis, -se) %>%
  kable()
```

	n	mean	sd	median	min	max
alter_z	3468	0	1	0.0319708	-1.79595	2.373994
alter0_z	3468	0	1	0.0319708	-1.79595	2.373994

## Aufgabe 2b

Führen Sie eine Regression von Einkommen auf  $alter\_0$  und  $bildung$  (Modell 1) und eine Regression von  $einkommen\_z$  auf  $alter\_0z$  und  $bildung\_z$  (Modell 2) durch und vergleichen Sie die b-Koeffizienten.

```
mod1 <- lm(einkommen ~ alter0 + bildung_rec, data = allb_sub_z)
mod2 <- lm(einkommen_z ~ alter0_z + bildung_z, data = allb_sub_z)

texreg(list(mod1,
            mod2))
```

	Model 1	Model 2
(Intercept)	7.17*** (0.28)	0.00 (0.02)
alter0	0.04*** (0.01)	
bildung_rec	1.20*** (0.07)	
alter0_z		0.14*** (0.02)
bildung_z		0.29*** (0.02)
R <sup>2</sup>	0.08	0.08
Adj. R <sup>2</sup>	0.08	0.08
Num. obs.	3039	3039
RMSE	4.74	0.96

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

Table 2: Statistical models

#ODER

tbl\_std(mod1)

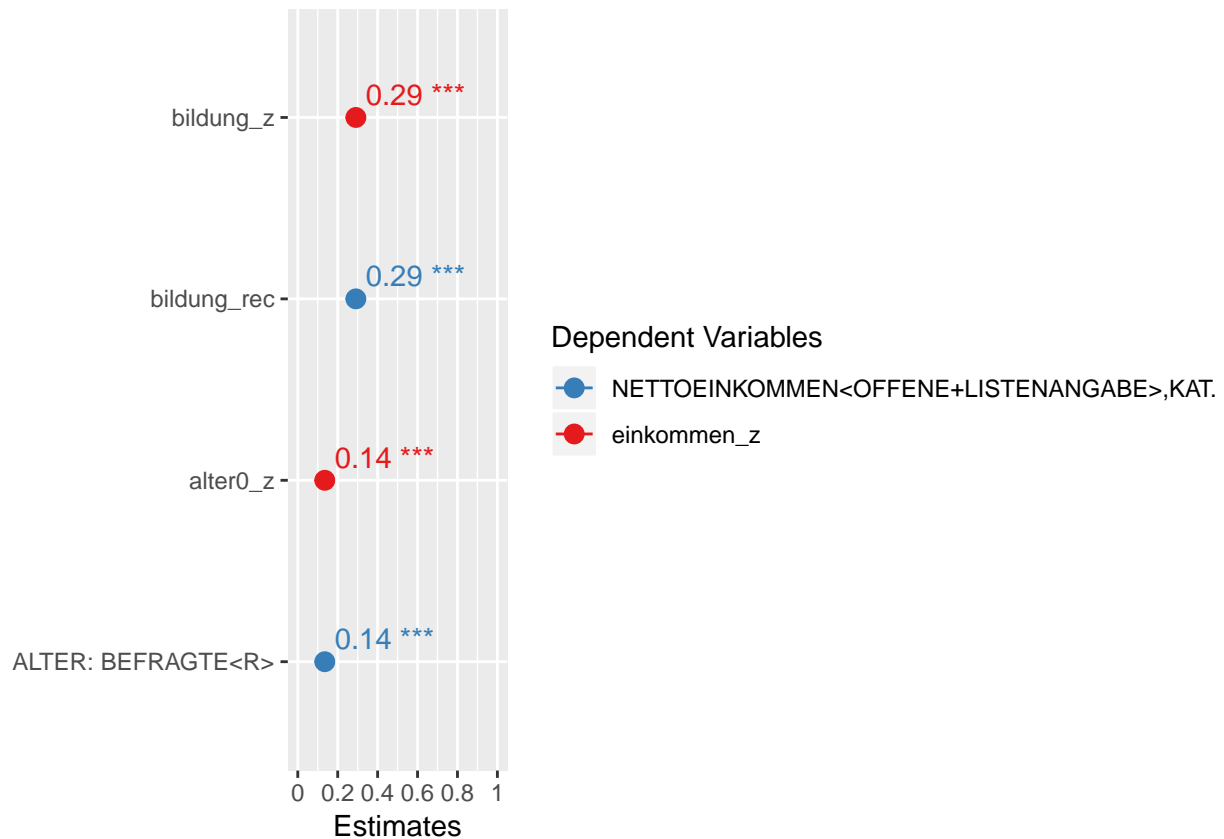
Table 3:

	<i>Dependent variable:</i>	
	einkommen	
	b (1)	std.b (2)
alter0	0.039*** (0.005)	0.135*** (0.005)
bildung_rec	1.199*** (0.074)	0.291*** (0.074)
Constant	7.165*** (0.282)	0.000 (0.282)
Observations	3,039	3,039
R <sup>2</sup>	0.082	0.082
Adjusted R <sup>2</sup>	0.081	0.081
Residual Std. Error (df = 3036)	4.741	4.741
F Statistic (df = 2; 3036)	135.439***	135.439***

Note:

\* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$

sjPlot::plot\_models(mod1, mod2, std.est = "std", show.values = T, show.p = T)



### Aufgabe 2c

Wie erklären Sie die Werte  $b$  und  $b^*$  in Modell 2? TIPP: Verwenden Sie bei Modell 2 das z-transformierte Einkommen als abhängige Variable.

### Aufgabe 3

Erstellen Sie ein multivariates Regressionsmodell mit  $Y = \text{Einkommen}$ . Versuchen Sie dabei den  $R^2$ -Wert so groß wie nur irgendwie möglich zu bekommen. Jeder schmutzige Trick der Sozialforschung ist erlaubt (und in diesem Fall erwünscht).

- Einzige Einschränkung: Keine Regression von  $Y$  auf  $Y$ .

```
allb_r <- allbus %>%
  select(V84, V86, V81, V420, V98, V118, V269, V103, V7, V13, V14,
         V16, V20, V21, V22, V25, V30, V31, V494, V9, V209, V279,
         V71, V711, V216, V215, V495, V513, V514, V377) %>%
  rename(alter = V84,
         bildung = V86,
         geschl = V81,
         einkommen = V420,
         arbeitsstd = V118,
         keineberufsausbildung = V98,
         beruf = V103,
         westost = V7,
```

```

internet = V14,
computer = V16,
essen = V20,
besuchfreunde = V21,
besuchfamilie = V22,
kunst = V25,
theater = V30,
museum = V31,
haushaltseinkommen = V494,
wirtschaftslage = V9,
fernsehenmin = V71,
dauerbildung = V711,
demzufu = V216,
linksrechts = V215,
prokopfeink = V495 ,
krankengeldhh = V513,
elterngeldhh = V514,
gebd = V377) %>%
na.omit() %>%
mutate(alter0 =alter -18,
       alter0quad =alter0*alter0,
       bildung_rec = ifelse(bildung ==6|bildung ==7,0, bildung -1),
       geschl_rec = ifelse(geschl ==2,0,1),
       ganztags = ifelse(beruf ==1,1,0),
       halbtags = ifelse(beruf ==2,1,0),
       west = ifelse(westost ==1,1,0),
       immigrant = ifelse(gebd ==2,1,0))

highr2 <- lm(einkommen~geschl_rec +alter0 +alter0quad +bildung_rec +
keineberufsausbildung +arbeitsstd +halbtags +west +
internet +computer +essen +besuchfreunde +besuchfamilie +kunst +
theater +museum +fernsehenmin +
haushaltseinkommen +wirtschaftslage +
dauerbildung +demzufu +linksrechts +
prokopfeink + krankengeldhh +
elterngeldhh +immigrant,data =allb_r)

texreg(highr2, float.pos ="ht!")

```

	Model 1
(Intercept)	-0.84 (1.49)
geschl_rec	1.92*** (0.22)
alter0	0.30*** (0.03)
alter0quad	-0.00*** (0.00)
bildung_rec	0.25* (0.12)
keineberufsausbildung	-2.71*** (0.41)
arbeitsstd	0.06*** (0.01)
halbtags	-2.74*** (0.33)
west	1.53*** (0.23)
internet	-0.27** (0.09)
computer	0.13 (0.08)
essen	-0.32* (0.12)
besuchfreunde	-0.02 (0.12)
besuchfamilie	-0.22* (0.10)
kunst	0.14 (0.10)
theater	0.12 (0.18)
museum	0.10 (0.19)
fernsehenmin	-0.00 (0.00)
haushaltseinkommen	0.18*** (0.03)
wirtschaftslage	-0.41** (0.15)
dauerbildung	0.14*** (0.04)
demzufr	0.05 (0.09)
linksrechts	-0.02 (0.06)
prokopfeink	0.00*** (0.00)
krankengeldhh	-1.99 (1.82)
elterngeldhh	0.88 (0.62)
immigrant	-0.24 (0.33)
R <sup>2</sup>	0.68
Adj. R <sup>2</sup>	5 0.67
Num. obs.	764
RMSE	2.54

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