

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))

allb_sub_z %>%
  select(alter_z, alter0_z) %>%
  descr() %>%
  #select(-vars, -trimmed, -mad, -se) %>%
  kable()
```

var	type	label	n	NA.prc	mean	sd	se	md	trimmed	range
alter_z	numeric	alter_z	3468	0.0864304	0	1	0.0169809	0.0319708	-0.0110151	4.17 (-1.8-2.37) 0.
alter0_z	numeric	alter0_z	3468	0.0864304	0	1	0.0169809	0.0319708	-0.0110151	4.17 (-1.8-2.37) 0.

Aufgabe 2a

Vergleichen Sie die Zahlenwerte, Mean und die Standardabweichung von `alter_z` und `alter_0z` und erklären Sie Ihre “Beobachtung”.

Aufgabe 2b

Führen Sie eine Regression von Einkommen auf `Alter_0` und Bildung (Modell 1) und eine Regression von Einkommen 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)
```

	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 ²	0.08	0.08
Adj. R ²	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

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

#ODER

```
mod1 <- lm(einkommen ~ alter0 + bildung_rec, data = allb_sub_z)
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 ²	0.082	0.082
Adjusted R ²	0.081	0.081
Residual Std. Error (df = 3036)	4.741	4.741
F Statistic (df = 2; 3036)	135.439***	135.439***

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

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 =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 .