SM II Abgabe 2

Ilirjana Ajazaj

12.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!

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

Der standardisierte Regressionskoeffizient informiert über Veränderungen auf der Standardskala von Y, die durch Veränderungen von X um jeweils eine Standardeinheit ausgelöst werden. Dazu wird der Mittelwert der empirischen Variablenwerte berechnet, um anschlißend die Differenz einzelner Variablenwerte von diesem Mittelwert zu bilden. Diese Werte werden wiederum jeweils durch die Standardabweichung geteilt, woraus die Standardisierung folgt (Um die Variablenwerte von X und Y zu standardisieren, müssen wir also die von uns berechneten Differenzen zwischen Rohwerten und Mittelwert zusätzlich quad- rieren, dann addieren, diese Summe durch die Anzahl der Personen (N) dividieren und aus dem derart gefundenen Wert die Wurzel ziehen.). Der standardisierte Regressionskoeffizient b* wird genutzt um darüber zu informieren, wieviel Standardeinheit(en) Y zu- oder abnimmt, wenn X um eine Standardeinheit steigt.b* macht somit eine Aussage über den Teil der Standardabweichung von Y, der durch X hervorgerufen wird bzw. darüber, welcher Anteil der Varianz von Y durch X gebunden wird. Standardisierte Regressionskoeffizienten werden zum Vergleich von Einflussstärken unterschiedlicher unab- hängiger Variablen innerhalb eines Modells (bei Schätzung mit einem bestimmten Datensatz) ein- gesetzt. Sie ermöglichen keine inhaltlichen Schätzungen auf einer empirischen Skala (Urban/Mayer 2011:80) BSP.: aus (Urban/Mayerl 2011: 73): Nehmen wir an, die Angaben zum Alter von Personen seien nicht in Jahren, sondern in Monatsangaben in die Analyse eingegangen. So würde aus dem Wert 17 (Jahre) der Wert 204 (Monate) und aus 52 der Wert 624. Wird mit diesen neuen Werten eine unstandardisierte Regressionsanalyse durchgeführt, die ansonsten mit der nach Tabelle 2.2 gerechneten Analyse iden- tisch ist, so ergibt sich ein Regressionskoeffizient von 0,78 (statt 9,37). Mit jedem zusätzlichen Lebensmonat steigt dann also das zu erwartende Nettoein- kommen um ca. 78 Cent.

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) und das Einkommen. [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".

Durch die Z-Transformation wird der Mittelwert der Verteilung jeweils ohnehin nach '0' verschoben, von welchem die Abweichungen beider Variablen, mit gleichen Fällen, entsprechend gleich ausfallen. Die z-

Transformation erzeugt grundsätzlich eine Verteilung mit den Eigenschaften MW=0 und s=1. Schiefe und Kurtosis bleiben jedoch unberührt.

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

	n	mean	sd	median	min	max	range	skew	kurtosis
alter_z	3468	0	1	0.0319708	-1.79595	2.373994	4.169944	0.0622101	-0.8939557
alter0_z	3468	0	1	0.0319708	-1.79595	2.373994	4.169944	0.0622101	-0.8939557

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.

Die Steigungskoeffizienten/Regressionskoeffizienten "b" für "Modell 1" betragen 0,039 (alter0) und 1,199 (Bildungsabschluss) und für "Model_z" betragen sie ("b*")0,678 (alter0_Z) und 1,453 (bildung_z).

```
library(lm.beta)

library(stargazer)
```

```
##
## Please cite as:
##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
##  R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
model_1 <-lm (einkommen ~ alter0 + bildung_rec, data = allb_sub_z)

model_z <-lm (einkommen_z ~ alter0_z + bildung_z, data =allb_sub_z)

screenreg (list(model_1, model_z), float.pos ="ht!")
```

```
##
## =====================================
##              Model 1      Model 2
## -------------------------------------
## (Intercept)     7.17 ***     0.00
##                (0.28)       (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
```

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.*

In Modell_1 sind die UV's unstandardisiert und in Modell_Z standardisiert.

Aufgabe 3

Erstellen Sie ein multivariates Regressionsmodell mit Y=Einkommen. Versuchen Sie dabei den R2-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.*

Tipp: Mit `binoculaR` könnt ihr den Datensatz nach relevanten Variablen durchsuchen :)

```r
#Regressionsmodell einfach um eine möglichst große Menge von theoretisch begründeten oder unbegründet
# Urban/Mayerl 2011: 109 Giftliste

allb_r <-allbus %>%
select (V84, V86, V81, V420, V269, V103, V7, V13, V14,
V16, V20, V25, V30,V31, V494, V9, V209, V279, V216, V215, V495, V513) %>%
rename (alter =V84,
        bildung =V86,
          geschl =V81,
          einkommen =V420,
          beruf =V103,
          westost =V7,
          internet =V14,
          computer =V16,
          essen =V20,
          kunst =V25,
          theater =V30,
          museum =V31,
          haushaltseinkommen =V494,
          wirtschaftslage =V9)%>%
          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),
       west = ifelse(westost ==1,1,0))

giftliste <- lm (einkommen~geschl_rec +alter0 +alter0quad +bildung_rec + internet +computer +essen +kuns

screenreg(giftliste, float.pos ="ht!")
```

```
##
## ========================
##              Model 1
```

```
## ------------------------
## (Intercept)   -20.97 ***
##               (0.45)
## geschl_rec      1.14 ***
##               (0.07)
## alter0          0.47 ***
##               (0.01)
## alter0quad     -0.00 ***
##               (0.00)
## bildung_rec     2.16 ***
##               (0.04)
## internet       -1.29 ***
##               (0.03)
## computer        1.17 ***
##               (0.03)
## essen          -3.07 ***
##               (0.05)
## kunst           3.45 ***
##               (0.05)
## theater        -0.91 ***
##               (0.08)
## museum          3.21 ***
##               (0.06)
## ------------------------
## R^2             0.93
## Adj. R^2        0.93
## Num. obs.    2717
## RMSE            1.59
## ========================
## *** p < 0.001, ** p < 0.01, * p < 0.05
```