Übungsaufgabe 5

Fabio Votta

29.November 2018

Aufgabe 1

Benutzen Sie den Datensatz „SOEP_wide_lehrdaten.sav“

Aufgabe 1a

Berechnen Sie ein Cross-lagged-Design mit zwei Regressionen bzgl. der Variablen Subjektive Gesundheit und Lebenszufriedenheit (Jahre 2003 und 2004)

Subjektive Gesundheit - `gesund_org.2003` - `gesund_org.2004`

Lebenszufriedenheit - `lebensz_org.2003` - `lebensz_org.2004`

```
lebens_mod <- lm(lebensz_org.2004 ~ gesund_org.2003 + lebensz_org.2003, data = soep)
gesund_mod <- lm(gesund_org.2004 ~ gesund_org.2003 + lebensz_org.2003, data = soep)

screenreg(list(lebens_mod, gesund_mod))
```

```
##
## ==========================================
##                   Model 1      Model 2
## ------------------------------------------
## (Intercept)          2.25 ***     0.98 ***
##                     (0.15)       (0.07)
## gesund_org.2003      0.27 ***     0.66 ***
##                     (0.04)       (0.02)
## lebensz_org.2003     0.52 ***     0.02 *
##                     (0.02)       (0.01)
## ------------------------------------------
## R^2                  0.32         0.45
## Adj. R^2             0.32         0.45
## Num. obs.         2111         2111
## RMSE                 1.49         0.71
## ==========================================
## *** p < 0.001, ** p < 0.01, * p < 0.05
```

Aufgabe 1b

Welche Aussage lässt sich bzgl. der Stabilität der Variablen Lebenszufriedenheit und Subj. Gesundheit machen?

Beide Stabilitätskoeffizienten sind über 0,5 und können somit als stabil gelten. Die Stabilität der Variable Gesundheit ist ein wenig höher als die der Lebenszufriedenheit. Es lässt sich jedoch keine Aussage über die Richtung bzw. Kausalität treffen.

Aufgabe 1c

Wie bewerten Sie die kausale Reihenfolge bzgl. der Effekte Subj. Gesundheit <-> Lebenszufriedenheit anhand der „Kreuzkoeffizienten“. Vergleichen Sie hierzu die Kreuzladungen.

T-Test:

$$t_{df} = \frac{b_1 - b_2}{\sqrt{(SE_{b_1})^2 + (SE_{b_2})^2}}$$

```r
t_test_coefs <- function(mod1, mod2, coef1 = NULL, coef2 = NULL) {

   tidy_mod1 <- tidy(mod1) %>%
      filter(term == coef1)

   tidy_mod2 <- tidy(mod2) %>%
      filter(term == coef2)

   b1 <- tidy_mod1$estimate
   b2 <- tidy_mod2$estimate

   se1 <- tidy_mod1$std.error
   se2 <- tidy_mod2$std.error

   oben <- b1 - b2

   unten <- sqrt((se1)^2 + (se2)^2)

   t_value = oben/unten

   return(list(t_value = t_value))
}


t_test_coefs(lebens_mod, gesund_mod,
             coef1 = "gesund_org.2003",
             coef2 = "lebensz_org.2003")
```

```
## $t_value
## [1] 6.489016
```

Der Kreuzkoeffizient von Gesundheit 2003 auf Lebenszufriedenheit 2004 ist mit 0,27 größer als der von Lebenszufriedenheit 2003 auf Gesundheit 2004 (0,02). Das ist ein Hinweis auf Kausalität zwischen Gesundheit und Lebenszufriedenheit: der Gesundheitszustand hat demnach Einfluss auf die Lebenszufriedenheit. Der Unterschied zwischen den Koeffizienten ist auf dem 99%igem Signifikanz-Niveau signifikant (t = 6.49)

Aufgabe 2

Erstellen Sie eine Interaktionsvariable zwischen Geschlecht und Alter und reduzieren Sie vor den folgenden Regressionsanalysen den Datensatz (Alter unter 46 Jahren, alter0 < 28). Modell 1 enthält dann Alter und Geschlecht, in Modell 2 kommt die Interaktionsvariable hinzu.

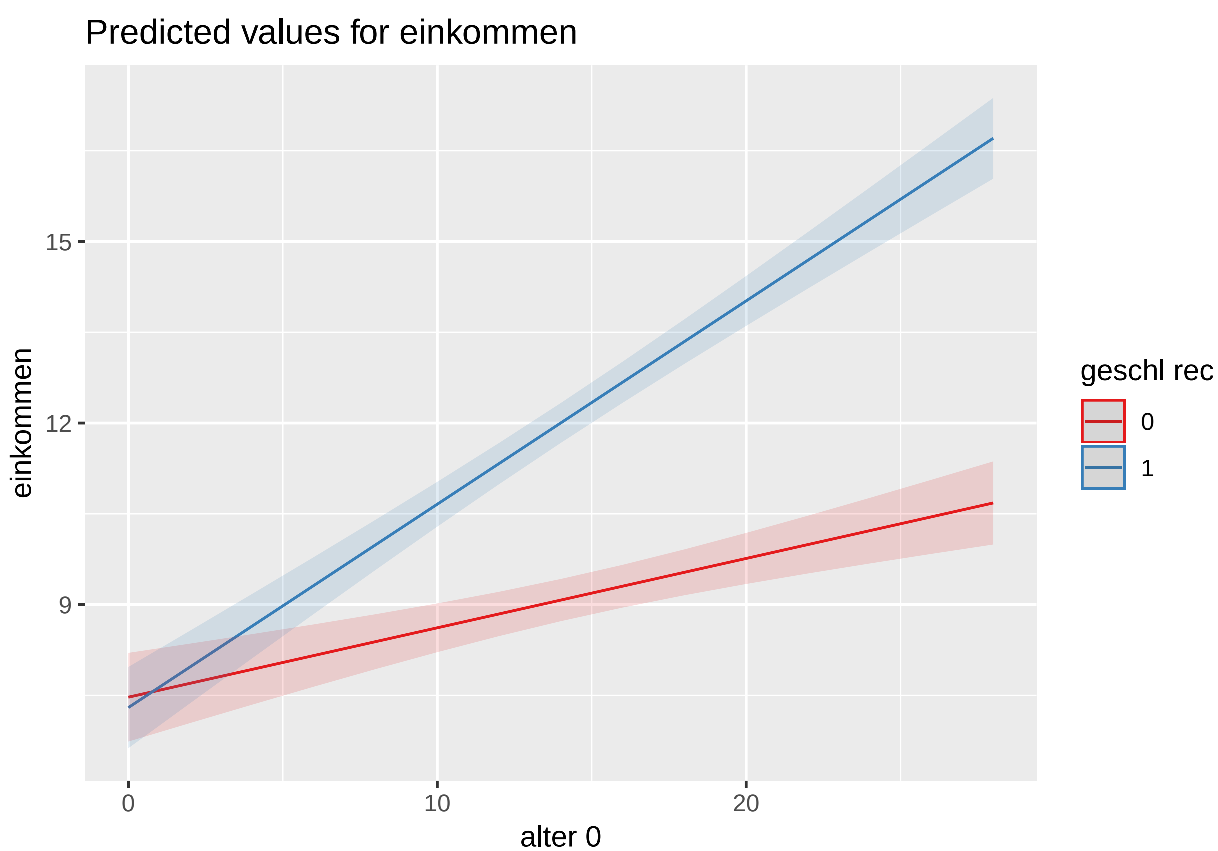
```r
allb_sub_u46 <- filter(allb_sub, alter0 < 28)
```

```r
mod1 <- lm(einkommen ~ alter0 + geschl_rec, data = allb_sub_u46)
mod2 <- lm(einkommen ~ alter0 + geschl_rec + alter0*geschl_rec, data = allb_sub_u46)

screenreg(list(mod1, mod2))
```

```
##
## ============================================
##                      Model 1      Model 2
## --------------------------------------------
## (Intercept)             5.76 ***     7.47 ***
##                        (0.29)       (0.37)
## alter0                  0.23 ***     0.11 ***
##                        (0.02)       (0.02)
## geschl_rec              3.00 ***    -0.17
##                        (0.25)       (0.51)
## alter0:geschl_rec                    0.22 ***
##                                     (0.03)
## --------------------------------------------
## R^2                     0.22         0.25
## Adj. R^2                0.22         0.25
## Num. obs.            1224         1224
## RMSE                    4.36         4.27
## ============================================
## *** p < 0.001, ** p < 0.01, * p < 0.05
```

```r
plot_model(mod2, type = "int")
```



Aufgabe 2a

Berechnen Sie anhand von Modell 1 und Modell 2 jeweils das prognostizierte Einkommen eines Mannes und einer Frau im Alter von jeweils 30 Jahren.

```
## Modell 1
intercept_mod1 <- mod1$coefficients[1]
alter_mod1 <- mod1$coefficients[2]
geschl_mod1 <- mod1$coefficients[3]

## Modell 1 Mann
intercept_mod1 + 12*alter_mod1 + 1*geschl_mod1
```

```
## (Intercept)
##    11.54269
```

```
## Modell 1 Frau
intercept_mod1 + 12*alter_mod1 + 0*geschl_mod1
```

```
## (Intercept)
##    8.540775
```

```
## Modell 2
intercept_mod2 <- mod2$coefficients[1]
alter_mod2 <- mod2$coefficients[2]
geschl_mod2 <- mod2$coefficients[3]
altgeschl_mod2 <- mod2$coefficients[4]

## Modell 2 Mann
intercept_mod2 + 12*alter_mod2 + 1*geschl_mod2 + 12*1*altgeschl_mod2
```

```
## (Intercept)
##    11.32997
```

```
## Modell 2 Frau
intercept_mod2 + 12*alter_mod2 + 0*geschl_mod2 + 12*0*altgeschl_mod2
```

```
## (Intercept)
##    8.845454
```

Aufgabe 2b

Was ist dabei der Interaktionseffekt und wie lässt er sich inhaltlich begründen?

Der Interaktionsffekt besteht darin, dass der Zusammenhang zwischen Alter und Einkommen bei Frauen weniger stark positiv ist als bei Männern. In anderen Worten, zunehmendes Alter wirk sich für Frauen weniger stark auf das Einkommen auf als für Männer. Das Geschlecht moderiert also den Effekt von Alter auf Einkommen. Interpretiert werden könnte dies so, dass Frauen häufiger einen weniger kontinuierlichen Berufsweg haben, vor allem durch Schwangerschaft und Kindererziehung.