

# Übungsaufgabe 4

Fabio Votta

22.November 2018

## Aufgabe 1

Erstellen Sie eine Regression von Einkommen auf Bildung, Geschlecht und Alter sowie der Dummyvariablen Zugang zu tertiärer Bildung (*bild\_tert*), die null kodiert ist, wenn der betreffende Befragte einen niedrigeren Schulabschluss als Fachhochschulreife hat und eins, wenn Umgekehrtes der Fall ist. Hinzu kommen die Interaktionsvariablen zwischen Geschlecht und Alter (*gesch\_alter*) sowie zwischen Alter und Zugang zu tertiärer Bildung (*alt\_tert*).

- 0 OHNE ABSCHLUSS
- 1 VOLKS-,HAUPTSCHULE
- 2 MITTLERE REIFE
- 3 FACHHOCHSCHULREIFE
- 4 HOCHSCHULREIFE
- NA ANDERER ABSCHLUSS
- NA NOCH SCHUELER
- NA KEINE ANGABE

## Aufgabe 1a

Berechnen Sie das Konfidenzintervall für die Variablen *bild\_tert* und Alter mittels der Koeffizienten und interpretieren Sie diese.

$$KI_{95} = b \pm t_n \times SE_b$$

Für  $n > 120$  und 95% Signifikanzniveau ist der kritische Wert  $t_{krit} = 1.96$

Für  $n > 120$  und 99% Signifikanzniveau ist der kritische Wert  $t_{krit} = 2.58$

## Aufgabe 1b

Testen Sie das Gesamtmodell auf Linearität.

## Aufgabe 2

Was ist unter Multikollinearität zu verstehen, warum ist es ein Problem, wenn diese in einer Modellschätzung vorliegt und wie kann das Vorliegen derselben diagnostiziert werden?

## Aufgabe 3

Wie ausgeprägt ist die Multikollinearität im Regressionsmodell von Aufgabe 1? Welche Gründe (inhaltliche) lassen sich für die Multikollinearität identifizieren?

#### Aufgabe 4

Bestimmen Sie den minimalen Stichprobenumfang für eine Variablenbeziehung in der Höhe von ca.  $f^2=0.1$ . Die Variablenbeziehung soll in einem Regressionsmodell mit 20 weiteren Kontrollvariablen mit einer Power von 0.8 und einem Signifikanzniveau von 95% (bzw. Irrtumswahrscheinlichkeit 0.05) getestet werden. Stellen Sie Ihren Denk- /Rechenvorgang dar.\_\_

Tipp: siehe Urban/Mayerl 2011: 159f.

$$N = \frac{\lambda}{f^2}$$

#### Aufgabe 5

*Welche Form von Fehlschluss wird durch ein niedriges Signifikanzniveau “begünstigt”?*

#### Aufgabe 6

*In welchen Fällen ist es sinnvoll das Signifikanzniveau höher anzusetzen als 95%?*