

Übungsaufgabe 9

Fabio Votta

10. Januar 2017

Aufgabe 1

Erstellen Sie zunächst dichotome Variablen für SPD-Wahl bei der letzten Bundestagswahl (aus V810), die Gewerkschaftsmitgliedschaft (z. B. V606) und die ev. Konfessionszugehörigkeit (V601).

V810 ZWEITSTIMME, LETZTE BUNDESTAGSWAHL

- 1 CDU-CSU
- 2 SPD
- 3 FDP
- 4 DIE GRUENEN
- 6 DIE LINKE
- 20 NPD
- 41 PIRATEN
- 42 AFD
- 43 FREIE WAEHLER
- 90 ANDERE PARTEI

V606 MITGLIED IN EINER GEWERKSCHAFT?

- 1 Ja
- 2 Nein

V601 KONFESSION, BEFRAGTE

- 1 EVANG.OHNE FREIKIRCH
- 2 EVANG.FREIKIRCHE
- 3 ROEMISCH-KATHOLISCH
- 4 AND.CHRISTL.RELIGION
- 5 AND.NICHT-CHRISTLICH
- 6 KEINER RELIGIONSGEM.

```
allbus %<>%  
  select(V810, V606, V601) %>%  
  mutate(spd = ifelse(V810 == 2, 1, 0)) %>%  
  mutate(gewerkschaft = ifelse(V606 == 1, 1, 0)) %>%  
  mutate(evangelisch = ifelse(V601 == 1 | V601 == 2, 1, 0))
```

Aufgabe 1a

Berechnen Sie eine log. Regression (Modell 1) AV: SPD-Wahl und UV: die Gewerkschaftsmitgliedschaft und die ev. Konfessionszugehörigkeit.

```
mod1 <- glm(spd ~ gewerkschaft + evangelisch,  
            data = allbus, family = "binomial")  
  
texreg(mod1)  
  
mod1_wide <- tidy_wide(mod1)
```

```
mod1_wide %>%
  mutate(z = intercept + 1 * gewerkschaft + 1 * evangelisch) %>%
  mutate(p = logit2prob(z)) %>%
  kable()
```

Die Wahrscheinlichkeit für ein evangelisches, aktuelles Gewerkschaftsmitglied, die SPD zu wählen liegt bei knapp 40 Prozent.

Aufgabe 1b

Berechnen Sie ein weiteres (nested) log. Regression mit zwei weiteren (frei wählbaren) unabhängigen Variablen (Modell 2)

```
mod1_wide %>%
  mutate(z = intercept + 0 * gewerkschaft + 0 * evangelisch) %>%
  mutate(p = logit2prob(z)) %>%
  kable()
```

Die Wahrscheinlichkeit für ein nicht-evangelisches Nicht-Gewerkschaftsmitglied die SPD zu wählen liegt bei 21.96 Prozent.

Aufgabe 1c

Erläutern Sie das Prinzip von McFaddens- R^2 und des Nagelkerke- R^2

```
mod1_wide %>%
  mutate(z = intercept + 1 * gewerkschaft + 1 * evangelisch) %>%
  mutate(odds = 1 / exp(z)) %>%
  kable()
```

```
mod1_wide %>%
  mutate(z = intercept + 0 * gewerkschaft + 0 * evangelisch) %>%
  mutate(odds = 1 / exp(z)) %>%
  kable()
```

Die Odds für eine Person (A), die sowohl evangelisch, als auch aktuell Gewerkschaftsmitglied ist, beträgt: 1,502 Das heißt: Damit ist die Chance, dass diese Person (A) SPD statt einer anderen Partei wählt 1,5 mal höher.

Die Odds für eine Person (B), die nicht-evangelisch und aktuell kein Gewerkschaftsmitglied ist, beträgt: 3,553 Das heißt: Damit ist die Chance, dass diese Person (B) SPD statt einer anderen Partei wählt 3,5 mal höher.

Aufgabe 1d

Interpretieren Sie diese für beide Modelle.

```
tidy(mod1) %>%
  mutate(odds_ratio = exp(estimate)) %>%
  kable()
```

Im Vergleich zu Personen, die keine evangelische Konfessionsangehörigkeit haben, haben Personen mit evangelischer Zugehörigkeit eine 1,64 mal höhere Chance die SPD zu wählen.

Im Vergleich zu Personen, die nicht aktuell Gewerkschaftsmitglied sind, haben Gewerkschaftsmitglieder eine 1,44 mal höhere Chance die SPD zu wählen.

Visualisierung

```
plot_model(mod1, show.values = T, show.p = T)

plot_model(mod1, terms = c("evangelisch", "gewerkschaft"), type = "pred")
```

Aufgabe 2

Welches der beiden Modelle weist die bessere Passung an die empirischen Daten auf? Erläutern Sie, wie Sie zur Feststellung dieses Sachverhaltes vorgehen müssen und beziehen Sie sich dabei auf das Prinzip der Devianz (vgl. die Berechnung von Pseudo- R^2). Berechnen Sie einen Likelihood Ratio- χ^2 -Test mit Modell 2 und Modell 1.

Das Modell kann nur Aussagen über Personen mit evangelischer Konfessionszugehörigkeit und Personen ohne evangelische Konfessionszugehörigkeit machen. Das heißt, es kann —keine— Aussagen über Katholiken, Muslime etc. machen. Genauso kann das Modell nur Aussagen über Personen machen, die aktuell Gewerkschaftsmitglieder sind oder keine sind. Also nicht über Personen, die beispielsweise mal welche waren oder zukünftig gerne einer Gewerkschaft angehören möchten.

Aufgabe 3

Regressionsdiagnostik zu Modell 2

Aufgabe 3a

Was können Sie über die Verteilung der Fälle aussagen? Erläutern Sie das Vorgehen und Ihr Ergebnis.

Aufgabe 3b

Können Sie Verzerrungen mit Blick auf die Y-Werte Verteilung, die Vollständigkeit der Information und die (vollständige) Separation identifizieren?