Lösung zu Blatt 4

Aufgabe 2: Poissonmodelle, Tests, Devianz

(a) Sei $\mu = E(y|\underline{x})$ die erwartete Anzahl der Arztbesuche

Modellgleichung in Link-Form:

$log(\mu) = \beta_0 + x_g\beta_1 + x_{inc}\beta_2 + x_{ill}\beta_3 + x_{red}\beta_4 + x_h\beta_5 + x_{fp}\beta_6 + x_{inc}x_{fp}\beta_7$

Modellgleichung in Response-Form:

$$\mu = exp(\underline{x}^T\underline{\beta}) = exp(\beta_0) \cdot exp(x_g\beta_1) \cdot exp(x_{ill}\beta_3) \cdot ...$$

Vorteile gegenüber einem klassischen linearen Modell:

- Logarithmus nur im positiven Bereich definiert $\Rightarrow$ negative Erwartungswerte nicht möglich
- Erwartungswerte des klassischen linearen Modells im gesamten Bereich der reellen Zahlen möglich; für Zähldaten nicht geeignet

(b) Es gelte bzw. gilt: $x_{inc} = 1$ und $\beta_2 =$ -0.216862

$\Rightarrow exp(1 \cdot (-0.216862)) = 0.805041$

Interpretation: Mit der Erhöhung der Einkommensklasse um eins verringert sich die erwartete Anzahl der Arztbesuche einer Person um den Faktor 0.81 [bei Konstanthaltung aller anderen Variablen].

(c)
- p-Wert von `freepoor` $= 0.223307 > 0.1 \Rightarrow$ nicht-signifikanter Einfluss
- Aber: höchst signifikante Interaktion zw. `income` und `freepoor`
- Vorschlag: Nimm `freepoor` und die Interaktion aus dem Datensatz und teste (z.B. mittels eines LR-Tests), ob sich der Fit entscheidend verschlechtert
- in Teil (d) wird zunächst nur der Interaktionseffekt aus dem Modell genommen:

(d) **Likelihood-Ratio-Test**:

- H_0: $\beta_7 = 0$
  ("Modell M_0 ohne Interaktion genügt ")
- H_1: $\beta_7 \neq 0$
  ("Interaktion sollte im Modell behalten werden (M_1)")
- Teststatistik:
  $$\lambda = \frac{\texttt{Residual deviance}(M_0) - \texttt{Residual deviance}(M_1)}{\phi} = \frac{4398.2 - 4381.9}{1} = 16.3$$
- Vergleichswert: $\chi^2_{0.95}(1) = 3.84$
- $\chi^2_{0.95}(1) < \lambda \Rightarrow H_0$ verwerfen (zum Niveau $\alpha = 5\%$),
  d.h.: zu großer Informationsverlust bei Übergang von M_1 zu M_0

- H_0 sogar zu einem Niveau von 0.5% verwerfbar
- Anzahl der Freiheitsgrade der χ^2-Statistik bestimmbar über:
  Anzahl der unter H_0 restringierten Parameter bzw.
  df(`Residual deviance` (M_0)) - df(`Residual deviance`(M_1))

- Vergleich der AICs:
  `AIC`(M_0) $>$ `AIC`(M_1) $\Rightarrow$ Bestätigung des Ergebnisses des LR-Tests;
  d.h. deutlich besserer Fit bei Verwendung von M_1