Lösung zu Aufgabe 1

(a) *Formulieren Sie das zugehörige multinomiale Logit-Modell.*

Responsevariable:

$$Y_i = \begin{cases} 0 & \text{custodial} \\ 1 & \text{admin} \\ 2 & \text{manage} \end{cases} \qquad i = 1, \ldots, 474$$

Verteilungsannahme:

$$\boldsymbol{y}_i | \boldsymbol{x}_i \sim \mathrm{M}(1, \boldsymbol{\pi}_i) \qquad i = 1, \ldots, n$$

$$\text{mit} \quad \boldsymbol{y}_i = (y_{i1}, y_{i2})^\top \quad \text{wobei} \quad y_{ir} = \begin{cases} 1 & Y_i = r \\ 0 & \text{sonst} \end{cases} \qquad r \in \{1, 2\} \quad Y_i \in \{0, 1, 2\}$$

$$\boldsymbol{x}_i = (1,\ \texttt{education}_i,\ \texttt{minority}_i)^\top$$

$$\boldsymbol{\pi}_i = (\pi_{i1}, \pi_{i2})^\top \quad \text{wobei} \quad \pi_{ir} = P(Y_i = r) \quad \text{für} \quad r \in \{1, 2\}$$

$$\text{und} \quad \pi_{i0} = P(Y_i = 0) = 1 - \pi_{i1} - \pi_{i2} \qquad \text{(Referenzkategorie)}$$

Strukturannahme multinomiales Logit-Modell:

$$P(Y_i = r) = \pi_{ir} = \frac{\exp(\beta_{r0} + \beta_{r1}\texttt{education}_i + \beta_{r3}\texttt{minority}_i)}{1 + \sum_{s=1}^{2} \exp(\beta_{s0} + \beta_{s1}\texttt{education}_i + \beta_{s3}\texttt{minority}_i)}$$

$$\text{für} \quad r \in \{1, 2\}$$

$$= \frac{\exp(\eta_{ir})}{1 + \exp(\eta_{i1}) + \exp(\eta_{i2})}$$

$$\text{mit} \quad \eta_{ir} = \beta_{r0} + \beta_{r1}\texttt{education}_i + \beta_{r3}\texttt{minority}_i$$

$$P(Y_i = 0) = \pi_{i0} = 1 - \pi_{i1} - \pi_{i2}$$

(b) *Interpretieren Sie die geschätzten Koeffizienten.*

- Interpretation von $\hat{\beta}_{10} = -2.232145$ und $\hat{\beta}_{20} = -30.268977$:

  Falls education = 0 und die Person keiner Minderheit angehört, dann beträgt die geschätzte Wahrscheinlichkeit für einen Job in der Verwaltung

  $$\hat{\pi}_{i1} = \frac{\exp(\hat{\beta}_{10})}{1 + \exp(\hat{\beta}_{10}) + \exp(\hat{\beta}_{20})} = \frac{\exp(-2.232145)}{1 + \exp(-2.232145) + \exp(-30.268977)} = 0.09690077 \ ,$$

  und für einen Job im Management

  $$\hat{\pi}_{i2} = \frac{\exp(\hat{\beta}_{20})}{1 + \exp(\hat{\beta}_{10}) + \exp(\hat{\beta}_{20})} = \frac{\exp(-32.809291)}{1 + \exp(-4.767341) + \exp(-32.809291)} = 6.457818e - 14 \ .$$

  Die Wahrscheinlichkeit für einen Job am Bankschalter ist also

  $$\hat{\pi}_{i0} = 1 - \hat{\pi}_{i1} - \hat{\pi}_{i2} = 0.9030992$$

- Interpretation von $\hat{\beta}_{11} = 0.4550169$:

  Wenn die Dauer der Ausbildung um ein Jahr steigt (und alle anderen Kovariablen gleich bleiben), dann

  – steigt die geschätzte logarithmierte Chance $\log\left(\frac{\hat{\pi}_{i1}}{\hat{\pi}_{i0}}\right)$ für einen Job in der Verwaltung im Vergleich zu einem Job am Schalter um 0.4550169

  – steigt die geschätzte Chance $\frac{\hat{\pi}_{i1}}{\hat{\pi}_{i0}}$ für einen Job in der Verwaltung im Vergleich zu einem Job am Schalter um den Faktor $\exp(0.4550169) \approx 1.58$

- Interpretation von $\hat{\beta}_{21} = 2.2002726$:

  Wenn die Dauer der Ausbildung um ein Jahr steigt (und alle anderen Kovariablen gleich bleiben), dann

  – steigt die geschätzte logarithmierte Chance $\log\left(\frac{\hat{\pi}_{i1}}{\hat{\pi}_{i0}}\right)$ für einen Job im Management im Vergleich zu einem Job am Schalter um 2.2002726

  – steigt die geschätzte Chance $\frac{\hat{\pi}_{i1}}{\hat{\pi}_{i0}}$ für einen Job im Management im Vergleich zu einem Job am Schalter um den Faktor $\exp(2.2002726) \approx 9.03$

- Interpretation von $\hat{\beta}_{12} = -1.174635$:

  Wenn der/die Angestellte einer Minderheit angehört (und alle anderen Kovariablen gleich bleiben), dann

  – verändert sich die geschätzte logarithmierte Chance für einen Job in der Verwaltung im Vergleich zu einem Job am Schalter um -1.174635

  – verändert sich die geschätzte Chance für einen Job in der Verwaltung im Vergleich zu einem Job am Schalter um den Faktor $\exp(-1.174635) \approx 0.31$

- Interpretation von $\hat{\beta}_{22} = -3.293551$:

  Wenn der/die Angestellte einer Minderheit angehört (und alle anderen Kovariablen gleich bleiben), dann

- verändert sich die geschätzte logarithmierte Chance für einen Job im Management im Vergleich zu einem Job am Schalter um -3.293551
- verändert sich die geschätzte Chance für einen Job im Management im Vergleich zu einem Job am Schalter um den Faktor $\exp(-3.293551) \approx 0.04$

(c) *Erklären Sie, wie im multinomialen Logit-Modell die Signifikanz einzelner Parameter getestet werden kann.*

Die Parameter werden im multinomialen Logit-Modell per Likelihood-Prinzip geschätzt. Für den ML-Schätzer gilt asymptotisch

$$\hat{\boldsymbol{\beta}} \overset{a}{\sim} \mathrm{N}(\boldsymbol{\beta}, F^{-1}(\hat{\boldsymbol{\beta}}))$$

und es kann zum Beispiel der Wald-Test zum Niveau α angewendet werden.

Erinnerung: Wald-Test allgemein

Allgemeine lineare Hypothese:

$$H_0 : \boldsymbol{C\beta} = \boldsymbol{d} \qquad \text{vs.} \qquad H_1 : \boldsymbol{C\beta} \neq \boldsymbol{d}$$

Wald-Teststatistik:

$$w = (\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{d})^\top [\boldsymbol{C}\widehat{\mathbf{Cov}}(\hat{\boldsymbol{\beta}})\boldsymbol{C}^\top]^{-1}(\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{d})$$

Ablehnbereich:

$$w > \chi^2_{1-\alpha}(s) \qquad \text{mit} \quad s = \dim(\boldsymbol{C})$$

Spezialfall für einen einzelnen Parameter

Die Nullhypothese reduziert sich beim Test auf einzelne Parameter im multinomialen Logit-Modell zu:

$$H_0 : \beta_{rj} = 0 \qquad \text{vs.} \qquad H_1 : \beta_{rj} \neq 0$$

mit $r = 1, \ldots, k-1$ und $j = 1, \ldots, p$, wobei k die Anzahl der Kategorien der Response-Variable und p die Anzahl der Kovariablen (ohne Intercept) bezeichnen.

Spezielle Teststatistik:

$$w = \left(\frac{\hat{\beta}_{rj}}{\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_{rj})}} \right)^2$$

Die geschätzte Varianz im Nenner von w entspricht den Diagonalelementen der inversen Fisher-Matrix. Sie kann im R-Output bei den Standardfehlern abgelesen werden, wobei die geschätzte Varianz den quadrierten Standardfehlern entspricht.

Ablehnbereich:

$$w > \chi^2_{1-\alpha}(1)$$

(d) *Testen Sie, ob der Zusammenhang zwischen der Dauer der Ausbildung und einem Job im Management zum Signifikanzniveau von 5% signifikant ist.*

$$H_0 : \beta_{21} = 0 \qquad \text{vs.} \qquad H_1 : \beta_{21} \neq 0$$

Teststatistik:

$$w = \left(\frac{\hat{\beta}_{21}}{\sqrt{\widehat{\mathrm{Var}}(\hat{\beta}_{21})}} \right)^2 = \left(\frac{2.2002726}{0.27296417} \right)^2 = 8.060665^2 = 64.97431$$

Ablehnbereich:

$$w > \chi^2_{0.95}(1) = 3.841$$

Testentscheidung:

$$w = 64.97431 > 3.841 \qquad \Rightarrow \qquad H_0 \text{ wird zum Niveau von 5\% abgelehnt.}$$

(e) *Geben Sie die geschätzte Wahrscheinlichkeitsverteilung für* **`admin`** *an, wenn die betrachtete Person keiner Minderheit angehört und eine 12-jährige Ausbildung hat.*

$$\begin{aligned}
\hat{\eta}_{i1} &= \hat{\beta}_{10} + \hat{\beta}_{11} \cdot 12 + \hat{\beta}_{13} \cdot 0 \\
&= -2.232145 + 0.4550169 \cdot 12 - 1.174635 \cdot 0 \qquad\qquad = 3.228058
\end{aligned}$$

$$\begin{aligned}
\hat{\eta}_{i2} &= \hat{\beta}_{20} + \hat{\beta}_{21} \cdot 12 + \hat{\beta}_{22} \cdot 0 \\
&= -3.865706
\end{aligned}$$

$$\begin{aligned}
\hat{\pi}_{i1} &= \hat{P}(Y_i = 1 | \texttt{education}_i = 12, \texttt{minority}_i = 0) \\
&= \frac{\exp(\hat{\eta}_{i1})}{1 + \exp(\hat{\eta}_{i1}) + \exp(\hat{\eta}_{i2})} \\
&= \frac{\exp(3.228058)}{1 + \exp(3.228058) + \exp(-3.865706)} \\
&= 0.961109
\end{aligned}$$

$$\begin{aligned}
\hat{\pi}_{i2} &= \hat{P}(Y_i = 2 | \texttt{education}_i = 12, \texttt{minority}_i = 0) \\
&= \frac{\exp(\hat{\eta}_{i2})}{1 + \exp(\hat{\eta}_{i1}) + \exp(\hat{\eta}_{i2})} \\
&= \frac{\exp(-3.865706)}{1 + \exp(3.228058) + \exp(-3.865706)} \\
&= 0.0007979765
\end{aligned}$$

$$\begin{aligned}
\hat{\pi}_{i0} &= \hat{P}(Y_i = 0 | \texttt{education}_i = 12, \texttt{minority}_i = 0) \\
&= 1 - \hat{\pi}_{i1} - \hat{\pi}_{i2} = 0.03809297
\end{aligned}$$

(f) *Geben Sie das zugrundeliegende (kumulative) Modell an.*

Modell:

$$\mathbb{P}(Y_i \leq r) = \frac{\exp(\gamma_{0r} + \mathbf{x}_i^T \gamma)}{1 + \exp(\gamma_{0r} + \mathbf{x}_i^T \gamma)}$$

mit

- $r \in \{0, 1\}$ Kategorien `custodial` und `admin` der Responsevariable
- γ_{0r}, $r \in \{0, 1\}$, "Schwellen" zwischen `custodial/admin`, `admin/manage`
  Unterschiedliche Intercepts für die verschiedenen Kategorien.
- $\gamma = (\gamma_1, \gamma_2)^\mathbf{T}$ Vektor mit Koeffizienten zu `education` und `minority`
  Ein Vektor für alle Kategorien von Y!

Verteilungsannahme: binäre Entscheidung ($y_i \leq$ oder $>$ Kategorie r)
Strukturannahme: in Term $\exp(\cdot)/(1 + \exp(\cdot))$

(g) *Interpretieren Sie die geschätzten Koeffizienten.*

Über logarithmierte, kumulierte Chancen

$$\log \frac{\mathbb{P}(Y_i \leq \texttt{custodial})}{\mathbb{P}(Y_i \geq \texttt{admin})} = \gamma_{00} - \mathbf{x}_i^T \gamma$$

$$\log \frac{\mathbb{P}(Y_i \leq \texttt{admin})}{\mathbb{P}(Y_i \geq \texttt{manage})} = \gamma_{01} - \mathbf{x}_i^T \gamma$$

- Interpretation von $\hat{\gamma}_{00}$:
  Die Chance auf eine Stelle am Schalter (oder niedriger) im Verhältnis zu einer Stelle im Verwaltungswesen (oder höher) liegt bei $\exp(7.9514) = 2839.548$, falls man keiner Minderheit angehört und ein Ausbildungsniveau von 0 Jahren aufweist.
- Interpretation von $\hat{\gamma}_{01}$:
  Die Chance auf eine Stelle im Verwaltungwesen (oder niedriger) im Verhältnis zum Managerposten (oder höher) liegt bei $\exp(14.1721) = 1428449$, falls man keiner Minderheit angehört und ein Ausbildungsniveau von 0 Jahren aufweist.
- Interpration von $\hat{\gamma}_1$ (`education`): Beachte das Minus im linearen Prädiktor!
  Die Chance auf einen Job der Kategorie r oder niedriger im Verhältnis zu einer höheren Jobkategorie fällt - mit dem Zuwachs des Ausbildungsniveaus um eine Einheit - um den Faktor $\exp(-0.8699976) = 0.4189526$; dieser Sachverhalt wird als unabhängig von den betrachteten Kategorien angenommen.
- Interpration von $\hat{\gamma}_2$ (`minority`): Beachte das Minus im linearen Prädiktor!
  Die Chance auf einen Job der Kategorie r oder niedriger im Verhältnis zu einer höheren Jobkategorie ist für ein Mitglied einer Minderheit um den Faktor $\exp(+1.0564379) = 2.876108$ höher als für jemanden, der keiner Minderheit angehört; dieser Sachverhalt wird als unabhängig davon betrachtet, um welche Jobkategorien es sich handelt.

(h) *Wie unterscheidet sich das sequentielle vom kumulativen Logit-Modell?*

Sequentielles Modell:

$$\mathbb{P}(Y_i = r | Y_i \geq r, \mathbf{x_i}) = \frac{\exp(\gamma_{\mathbf{0r}} + \mathbf{x_i^T}\gamma)}{\mathbf{1} + \exp(\gamma_{\mathbf{0r}} + \mathbf{x_i^T}\gamma)}$$

mit

- $r \in \{0, 1\}$ Kategorien `custodial` und `admin` der Responsevariable
- γ_{0r}, $r \in \{0, 1\}$, Parameter für Übergang zwischen `custodial`/`admin` und `admin`/`manage` Unterschiedliche Intercepts für die verschiedenen Kategorien.
- $\gamma = (\gamma_{\mathbf{1}}, \gamma_{\mathbf{2}})^{\mathbf{T}}$ Vektor mit Koeffizienten zu `education` und `minority` Ein Vektor für alle Kategorien von Y!

Verteilungsannahme: binäre Entscheidung
$y_i = r$ oder $y_i \neq r$, Übergang zwischen den Kategorien oder nicht.

Strukturannahme: in Term $\exp(\cdot)/(1 + \exp(\cdot))$

Unterschied: Annahme, dass Kategorien sukzsessive erreicht werden,
Modellierung als Prozess!