

Ludwig-Maximilians-Universität, München
Institut für Statistik



Empirischer Vergleich ordinaler Regressionsmodelle

Bachelorarbeit

Verfasser: Susanne Dandl
Betreuer: Prof. Dr. Gerhard Tutz, Institut für Statistik, LMU
Moritz Berger, Institut für Statistik, LMU
Eingereicht am: 30. Mai 2016

Zusammenfassung

In vielen Regressionsproblemen mit kategorialen, ordinalem Response ist dem Anwender zwar die Multinomialverteilung der Zielvariable bewusst, jedoch wird die Ordnung der Kategorien häufig in der Wahl der Linkfunktion missachtet. Somit wird nicht der volle Informationsgehalt der Variablen genutzt. Zwei Modelle, welche die Rangordnung einer Variablen beachten, sind das kumulative und Adjacent-Type Logit-Modell. Diese unterscheiden sich in der Art der verwendeten Logits. Während beim Adjacent-Type Modell die Logits benachbarter Kategorien genutzt werden, sind es beim kumulativen Modell kumulierte Logits. Dies führt zu unterschiedlichen Interpretationen. Welches Verfahren nun für die Berechnung von Regressionsmodellen zu bevorzugen ist, wird anhand mehrerer Datenbeispiele erläutert. Dazu werden zunächst, auf Grundlage des multinomialen Modells, die Komponenten des kumulativen Modells und des Adjacent-Type Modells genauer spezifiziert. Für beide Verfahren können sowohl Modelle unter Partial-Proportional-Odds Annahme als auch unter Annahme kategoriespezifischer Koeffizienten modelliert werden. Ein Likelihood-Ratio-Test kann darüber Aufschluss geben, welche Annahme getroffen werden sollte. Das kumulative Modell hat jedoch den Nachteil, dass es bei einer großen Anzahl geordneter Kategorien der Zielvariablen und mehreren kategorialen Einflussgrößen zu keinem Schätzwerten führen kann. Die Konzepte beider Methoden können auch auf die Modellierung longitudinaler Daten übertragen werden, indem zusätzlich ein Random Intercept in den Prädiktor aufgenommen wird.

Für die Modellierung der Vertrauensskalen des **Allbus**-Datensatzes wird das **R**-Package **vgam** mit der Funktion `vglm()` verwendet. Für die Modellierung der Schmerzkategorien des **Knee**-Datensatzes kann das **vcrrpart** Package mit der Funktion `olmm()` genutzt werden. Der Vergleich beider Verfahren anhand des **Allbus**-Datensatzes und **Knee**-Datensatzes gibt Aufschluss darüber, dass sowohl das kumulative Logit-Modell als auch das Adjacent-Category Logit-Modell in ähnlichen Situationen einen vergleichbaren Fit aufweisen. Dies bestätigen auch die Werte des AICs und der Devianz, welche als Vergleichskriterien für den Modellfit verwendet werden. Allgemein sollte das Verfahren genutzt werden, dessen Art der Logits bzw. dessen Interpretationsform präferiert wird.

Inhaltsverzeichnis

1	Einleitung	1
2	Grundlagen	2
2.1	Multinomialverteilung	2
2.1.1	Inferenz	4
2.1.2	Multinomiales Logit-Modell	5
2.2	Vorteile ordinaler Methoden	8
3	Ordinale Modelle	9
3.1	Kumulatives Modell	9
3.1.1	Motivation	9
3.1.2	Definition des kumulativen Logit-Modells	11
3.1.3	Kumulatives Logit-Modell für Kontingenztabellen	13
3.1.4	Inferenz	15
3.1.5	Methoden zur Modellüberprüfung	16
3.1.6	Erweiterung des kumulativen Logit-Modells	18
3.2	Adjacent-Type Modell	20
3.2.1	Definition des Adjacent-Type Logit-Modells	20
3.2.2	Verbindung zum multinomialen Logit-Modell	21
3.2.3	Adjacent-Type Logit-Modell für Kontingenztabellen	22
3.2.4	Inferenz	22
4	Vergleich der Methoden	24
4.1	VGAM Package	24
4.2	Allbus-Datenbeispiel	25
4.2.1	Variablenbeschreibung und Modellspezifikationen	25
4.2.2	Vergleich der Testergebnisse Zielvariable <i>government</i>	36
4.2.3	Vergleich der Testergebnisse Zielvariable <i>fellow</i>	44
4.3	Zusammenfassung des Vergleichs	49
5	Longitudinale Daten	51
5.1	Ordinale gemischte Modelle	51
5.2	Methodenvergleich anhand des Knee -Datensatzes	53
6	Fazit	62

Literaturverzeichnis	63
A Verfügbare Dateien	68
B R-Code und Ergebnistabellen	71
B.1 Likelihood-Ratio-Test kumulatives Logit-Modell Variable <i>government</i>	71
B.2 Auswertung Zielvariable <i>justice</i>	72

1 Einleitung

Die Theorie der ordinalen, kategorialen Daten und die Unterscheidung ihrer Methoden zu den ungeordneten Skalen erhalten seit 1980 durch Publikationen wie „Regression Models for Ordinal Data“ von McCullagh (siehe Peter McCullagh (1980)) Aufmerksamkeit in der Datenanalyse. Ordinale Skalen unterscheiden sich von nominalen Skalen darin, dass ihre Ausprägungen geordnet werden können. Zum Beispiel lassen sich Augenfarben in keine hierarchische Anordnung bringen. Somit gehören sie zu den nominalen Variablen. Ordinale Kategorien werden besonders häufig im Bereich medizinischer Gesundheitseinschätzungen und Messungen von Einstellungen und Meinungen zu öffentlichen Themen verwendet (vgl. Agresti (2010): S. 1 f.). Beispiele sind die in Kapitel 4 verwendeten Skalen zum Vertrauen in die Regierung und die Mitmenschen sowie die in Kapitel 5 verwendete Skala mit Stufen von 1 bis 5 für die Einschätzung von Knieschmerzen.

Es gibt zwei Arten geordneter kategorialer Variablen. Sie können zum einen auf festgesetzten, ordinalen Skalen beruhen. Ein Beispiel sind die Kategorien verschiedener Schulformen, welche anhand ihrer Abschlüsse in Hauptschule, Realschule und Gymnasium gestaffelt werden können. Zum anderen können sie durch das Zusammenfassen stetiger Größen entstehen (vgl. Tutz (2012): S. 241). Diese werden Intervallskalen genannt (vgl. Agresti (2010): S. 2). Ein Beispiel ist die Einteilung des Body Mass Indexes (BMI) in verschiedene Gruppen. Der BMI wird anhand des Gewichts in Kilogramm dividiert durch die quadrierte Größe in Metern berechnet. Wird ein Index von kleiner als 18.5 errechnet, so wird die Person der Gruppe der Untergewichtigen zugeordnet. Bei einem BMI von 18.5-24.99 erfolgt die Zuteilung zu den Normalgewichtigen, bei Werten zwischen 25 und 29.99 zu den Übergewichtigen und ab 30 zur Personengruppe der Fettleibigen. Diese Gruppengrenzwerte wurden von der World Health Organization (kurz: WHO) festgelegt (vgl. WHO - World Health Organization (2000)). Ein Nachteil dieses Vorgehens ist der Verlust von Information durch die Bildung von Intervallen. Im Gegensatz zu festgesetzten, ordinalen Skalen ist bei den Intervallskalen die Distanz zwischen den Kategorien genau festgelegt.

In vielen Regressionsproblemen mit kategorialem, ordinalem Response ist dem Anwender zwar die Multinomialverteilung der Zielvariable bewusst, jedoch wird die Ordnung der Kategorien in der Wahl seiner Methoden häufig nicht genutzt. Somit wird nicht der volle Informationsgehalt der Variablen verwendet (vgl. Agresti (2010): S. 1 ff.).

Nach einer Einführung in das Multinomialmodell, seine Inferenz und das multinomiale Logit-Modell, werden am Ende von Kapitel 2 die Vorteile ordinaler Methoden erläutert. In Kapitel 3 wird näher auf das kumulative Modell und das Adjacent-Type Modell eingegangen, welche die

ordinale Struktur der Daten ausnutzen. Diese beiden Methoden werden in Kapitel 4 anhand des **Allbus**-Datensatzes verglichen. Anschließend folgt in Kapitel 5 ein kurzer Überblick über die Verwendung gemischter ordinaler Modelle im kumulativen und im Adjacent-Category Modell für Daten mit Messwiederholungen, den sogenannten longitudinalen Daten. Beide Methoden werden dann anhand des **Knee**-Datensatzes kontrastiert. Zum Abschluss folgt eine Zusammenfassung der Erkenntnisse in einem Fazit.

2 Grundlagen

Ist die Zielvariable in einem Regressionsproblem kategorial, so kann angenommen werden, dass diese einer Multinomialverteilung folgt (Fahrmeir et al. (2009): S. 236). Bei der Verteilung wird nicht zwischen ordinalen und nominalen Skalen unterschieden. Die Ordnung fließt in die Wahl der Linkfunktion ein. In diesem Kapitel werden Annahmen, Wahrscheinlichkeitsfunktionen und Inferenz des Multinomialmodells sowie das multinomiale Logit-Modell näher erläutert.

2.1 Multinomialverteilung

Die folgenden Spezifikationen der Multinomialverteilung beziehen sich auf Fahrmeir et al. (2009) S. 236 ff. sowie Tutz (2012) S.209. Ausgehend von einer Variable Y mit Kategorien $1, 2, \dots, k$ ist die Wahrscheinlichkeit auf Kategorie $r \in 1, 2, \dots, k$ als

$$P(Y = r) = \pi_r \text{ mit } \pi_r \in [0, 1] \text{ und } \sum_{r=1}^k \pi_r = 1 \quad (1)$$

definiert. Jede Kategorie hat also eine eigene Wahrscheinlichkeit aufzutreten, sodass Y ein multivariater Response ist mit $\mathbf{y}^T = (y_1, y_2, \dots, y_k)$.

Wird nur eine Person, eine Befragung oder ein Messpunkt i betrachtet, so kann eine binäre Variable y_{ir} definiert werden mit

$$y_{ir} = \begin{cases} 1, & Y_i = r \\ 0, & \text{sonst} \end{cases} \text{ mit } r \in 1, \dots, k$$

Gilt $Y_i = r$, so entspricht $\mathbf{y}_i^T = (y_{i1}, y_{i2}, \dots, y_{ik})$ dem Vektor $(0, \dots, 0, 1, 0, \dots, 0)$, wobei die 1 an der r -ten Stelle des Vektors für die Wahl der r -ten Kategorie steht. \mathbf{y}_i folgt einer Multinomialverteilung und hat mit der Wahrscheinlichkeit $\boldsymbol{\pi}_i^T = (\pi_{i1}, \dots, \pi_{ik})$ und $\pi_{ir} = P(Y_i = r) =$

$P(y_{ir} = 1)$ die Wahrscheinlichkeitsfunktion

$$f(\mathbf{y}_i | \boldsymbol{\pi}_i) = \pi_{i1}^{y_{i1}} \cdot \dots \cdot \pi_{ik}^{y_{ik}}$$

Werden nun n Befragungen durchgeführt, so entspricht y_r mit $r \in 1, \dots, k$ nicht mehr binären Variablen, sondern der Anzahl an Wiederholungen in denen Kategorie r bei den n Befragungen aufgetreten ist. Der Vektor $\mathbf{y}^T = (y_1, \dots, y_k)$ ist dann multinomialverteilt mit $\mathbf{y} \sim M(n, \boldsymbol{\pi})$ und $\boldsymbol{\pi}^T = (\pi_1, \dots, \pi_k)$. Die Wahrscheinlichkeitsfunktion wird folgendermaßen definiert:

$$f(\mathbf{y} | \boldsymbol{\pi}) = \frac{n!}{y_1! \cdot \dots \cdot y_k!} \pi_1^{y_1} \cdot \dots \cdot \pi_k^{y_k}$$

Da y_k durch y_1, \dots, y_{k-1} ausgedrückt werden kann, wie die Umformung von $\sum_{i=1}^n y_i = n$ zu

$$y_k = n - \sum_{i=1}^{k-1} y_i = n - y_1 - \dots - y_{k-1},$$

zeigt, kann eine Reduzierung des Vektors \mathbf{y}^T zu $\mathbf{y}^T = (y_1, \dots, y_q)$ mit $q = k - 1$ vorgenommen werden. Zudem kann π_k nach Umformung von (1) zu

$$\pi_k = 1 - \sum_{r=1}^{k-1} \pi_r = 1 - \pi_1 - \dots - \pi_{k-1}$$

durch π_1, \dots, π_q mit $q = k - 1$ berechnet werden. Somit ergibt sich die Wahrscheinlichkeitsfunktion unter Reduzierung der verwendeten Anzahl an Parametern mit $\boldsymbol{\pi}^T = (\pi_1, \dots, \pi_q)$ als:

$$f(\mathbf{y} | \boldsymbol{\pi}) = \frac{n!}{y_1! \cdot \dots \cdot y_q! \cdot (n - y_1 - \dots - y_q)!} \pi_1^{y_1} \cdot \dots \cdot \pi_q^{y_q} \cdot (1 - \pi_1 - \dots - \pi_q)^{n - y_1 - \dots - y_q}$$

Für den Erwartungswert von \mathbf{y} gilt

$$E(\mathbf{y}) = n\boldsymbol{\pi} = \begin{pmatrix} n\pi_1 \\ \vdots \\ n\pi_q \end{pmatrix}$$

und für die Kovarianz von \mathbf{y}

$$\text{Cov}(\mathbf{y}) = \begin{pmatrix} n\pi_1(1 - \pi_1) & \dots & -n\pi_1\pi_q \\ \vdots & \ddots & \vdots \\ -n\pi_q\pi_1 & \dots & n\pi_q(1 - \pi_q) \end{pmatrix}$$

Anstelle der multinomialen Verteilung $\mathbf{y} \sim M(n, \boldsymbol{\pi})$ kann auch die skalierte Multinomialverteilung verwendet werden, sodass nicht die absolute Anzahl betrachtet wird, sondern die relativen Häufigkeiten der Ausprägungen $1, \dots, q$

$$\bar{\mathbf{y}} = \frac{\mathbf{y}}{n} = \left(\frac{y_1}{n}, \dots, \frac{y_q}{n} \right)^T = (\bar{y}_1, \dots, \bar{y}_q)^T$$

In diesem Fall gilt $\bar{\mathbf{y}} \sim M(n, \boldsymbol{\pi})/n$. Der Erwartungswert ist gegeben als

$$E(\bar{\mathbf{y}}) = \boldsymbol{\pi} = \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_q \end{pmatrix}$$

und die Kovarianz als

$$\text{Cov}(\bar{\mathbf{y}}) = \frac{1}{n} \begin{pmatrix} \pi_1(1 - \pi_1) & \dots & -\pi_1\pi_q \\ \vdots & \ddots & \vdots \\ -\pi_q\pi_1 & \dots & \pi_q(1 - \pi_q) \end{pmatrix}$$

2.1.1 Inferenz

Im vorherigen Kapitel wurde die Wahrscheinlichkeitsfunktion für eine Person $i \in 1, \dots, n$ als multinomialverteilt mit

$$f(\mathbf{y}_i | \boldsymbol{\pi}_i) = \pi_{i1}^{y_{i1}} \cdot \dots \cdot \pi_{iq}^{y_{iq}}$$

definiert (vgl. Tutz (2012): S. 218). Alle folgenden Spezifikationen des Abschnitts beziehen sich nun auf Fahrmeir et al. (2009) S. 247 ff.. Werden n Subjekte betrachtet, so folgt unter Annahme der Unabhängigkeit der Beobachtungen die Likelihoodfunktion als:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n f(\mathbf{y}_i | \boldsymbol{\pi}_i) = \prod_{i=1}^n \pi_{i1}^{y_{i1}} \cdot \dots \cdot \pi_{iq}^{y_{iq}} \cdot (1 - \pi_{i1} - \dots - \pi_{iq})^{(n - y_{i1} - \dots - y_{iq})} \quad (2)$$

Dabei gilt $q = k - 1$.

Die Funktion wird verwendet, um die Parameter nach dem Maximum-Likelihood-Prinzip zu schätzen. Zur Vereinfachung wird sie zunächst zur Log-Likelihood-Funktion

$$l(\boldsymbol{\beta}) = \sum_{i=1}^n (y_{i1} \log(\pi_{i1}) + \dots + y_{iq} \log(\pi_{iq}) + (n - y_{i1} - \dots - y_{iq}) \log(1 - \pi_{i1} - \dots - \pi_{iq}))$$

logarithmiert. Nach Ableitung der Log-Likelihood-Funktion nach $\boldsymbol{\beta}$ ergibt sich $s(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$. Wird diese Score-Funktion nun 0 gesetzt und nach $\boldsymbol{\beta}$ aufgelöst, so erhält man $\hat{\boldsymbol{\beta}}$, für das

$$s(\hat{\boldsymbol{\beta}}) \stackrel{!}{=} 0$$

gilt. Durch die zweite Ableitung wird überprüft, ob es sich um ein Maximum handelt. Im Folgenden wird von p Einflussgrößen und damit von einer Designmatrix der Form $\mathbf{x}_i^T = (1, x_{i1}, \dots, x_{ip})$ ausgegangen. Der Zusammenhang zwischen $\boldsymbol{\pi}_i$ und dem Prädiktor $\mathbf{x}_i^T \boldsymbol{\beta}$ mit $\boldsymbol{\beta}^T = (\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_p)$ wird über die Responsefunktion h festgelegt. Es gilt:

$$\pi_{ir} = h(\eta_{ir}) = h(\mathbf{x}_i^T \boldsymbol{\beta})$$

Die Schätzung von $\hat{\boldsymbol{\beta}}$ erfolgt in iterativer Weise nach dem Fisher-Scoring-Algorithmus, welcher in Kapitel 3.1 zum kumulativen Modell näher erläutert wird. Dazu wird die erwartete Fisher-Informationsmatrix benötigt. Diese Matrix $F(\boldsymbol{\beta})$ wird als

$$F(\boldsymbol{\beta}) = E \left(- \frac{\partial^2 l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right)$$

definiert. Der ML-Schätzer $\hat{\boldsymbol{\beta}}$ ist dann asymptotisch normalverteilt mit

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

und $F^{-1}(\hat{\boldsymbol{\beta}})$ als Inverses der Fisher-Informationsmatrix.

2.1.2 Multinomiales Logit-Modell

Das multinomiale oder mehrkategoriale Logit-Modell ist der natürliche Link der Multinomialverteilung. Sie verbindet eine Responsevariable Y mit geordneten oder ungeordneten Kategorien $1, \dots, k$ durch eine Linkfunktion mit dem Designvektor $\mathbf{x}^T = (1, x_1, \dots, x_p)$. Die folgenden Definitionen zum Logit-Modell sind in Fahrmeir et al. (2009) S. 239 zu finden. Zur Verein-

fachung wird angenommen, dass für jede Kategorie der Zielvariable der Kovariablenvektor \mathbf{x} identisch ist. Für die Auftrittswahrscheinlichkeit π_r von Kategorie r gilt:

$$\pi_r = P(Y = r|\mathbf{x}) = \frac{\exp(\mathbf{x}^T \boldsymbol{\beta}_r)}{1 + \sum_{s=1}^q \exp(\mathbf{x}^T \boldsymbol{\beta}_s)} \text{ mit } q = k - 1 \text{ und } r = 1, \dots, q \quad (3)$$

Zudem gilt $\boldsymbol{\beta}_r^T = (\beta_{r0}, \beta_{r1}, \dots, \beta_{rp})$ mit $r = 1, \dots, q$, wenn die Wirkung der Einflussvariablen über die Kategorien variiert. Wird dies nicht angenommen, so entspricht $\boldsymbol{\beta}_r^T$ dem Vektor $(\beta_{r0}, \beta_1, \dots, \beta_p)$. Bei dem mehrkategorialen Logit-Modell wird die (logarithmierte) Chance zwischen der betrachteten Kategorie r und einer Referenzkategorie betrachtet. In Softwareprogrammen wie **R** wird als Referenzkategorie häufig die letzte Kategorie k verwendet. Die Auftrittswahrscheinlichkeit für Kategorie k ist gegeben durch:

$$\pi_k = 1 - \pi_1 - \dots - \pi_q = \frac{1}{1 + \sum_{s=1}^q \exp(\mathbf{x}^T \boldsymbol{\beta}_s)}$$

Um nun die Chance oder das relative Risiko von Kategorie r zu der letzten Kategorie k zu betrachten, wird die Auftrittswahrscheinlichkeit von Kategorie r mit der von Kategorie k verglichen:

$$\frac{P(Y = r|\mathbf{x})}{P(Y = k|\mathbf{x})} = \frac{\pi_r}{\pi_k} = \exp(\mathbf{x}^T \boldsymbol{\beta}_r) \text{ mit } r = 1, \dots, q \quad (4)$$

Nach dem Logarithmieren von (4) wird folgende äquivalente Log-Odds-Darstellung erhalten, wenn categoriespezifische Koeffizienten angenommen werden:

$$\log \frac{\pi_r}{\pi_k} = \log (\exp(\mathbf{x}^T \boldsymbol{\beta}_r)) = \mathbf{x}^T \boldsymbol{\beta}_r = \beta_{r0} + x_1 \beta_{r1} + \dots + x_p \beta_{rp} \quad (5)$$

Der Koeffizientenvektor $\boldsymbol{\beta}_r$ hängt von der betrachteten Kategorie r und der gewählten Referenzgruppe ab. Zur Interpretation des Einflusses einer Variablen x_j mit $j \in 1, \dots, p$ auf die Wahl der Kategorie r , wird die Veränderung der Chance zwischen Kategorie r und k bei einer Steigerung von x_j um c Einheiten betrachtet:

$$\begin{aligned} & \frac{P(Y = r|\mathbf{x} \text{ mit } x_j + c)/P(Y = k|\mathbf{x} \text{ mit } x_j + c)}{P(Y = r|\mathbf{x})/P(Y = k|\mathbf{x})} \\ & \stackrel{(4)}{=} \frac{\exp(\beta_{r0} + x_1 \beta_{r1} + \dots + (x_j + c) \beta_{rj} + \dots + x_p \beta_{rp})}{\exp(\beta_{r0} + x_1 \beta_{r1} + \dots + x_j \beta_{rj} + \dots + x_p \beta_{rp})} \\ & \stackrel{(5)}{=} \frac{\exp((x_j + c) \beta_{rj})}{\exp(x_j \beta_{rj})} = \frac{\exp(x_j \beta_{rj} + c \beta_{rj})}{\exp(x_j \beta_{rj})} = \exp(c \beta_{rj}) \end{aligned}$$

Je nach Vorzeichen kann die Chance für Kategorie r im Gegensatz zu Kategorie k um den Faktor $\exp(c\beta_{rj})$ steigen (bei einem positiven Vorzeichen) oder sinken (bei einem negativen Vorzeichen), wenn sich x_j um c Einheiten verändert und alle anderen Variablen festgehalten werden. Gilt $c = 1$, so wird die Veränderung der Chance auf Kategorie r zu k um den Faktor $\exp(\beta_{rj})$ bei Steigerung der Einflussgröße x_j um eine Einheit betrachtet, unter Konstanthalten aller anderen Kovariablen. Zur Interpretation der Veränderung der Chance auf Stufe r zu $l \in 1, \dots, (k - 1)$, also einer anderen Kategorie als k , wird folgender Bruch betrachtet:

$$\begin{aligned} \frac{P(Y = r | \mathbf{x} \text{ mit } x_j + c)}{P(Y = l | \mathbf{x} \text{ mit } x_j + c)} &\stackrel{(3)}{=} \frac{\frac{\exp(\beta_{ro} + \dots + (x_j + c)\beta_{rj} + \dots + x_p\beta_{rp})}{1 + \sum_{s=1}^q \exp(\beta_{so} + \dots + (x_j + c)\beta_{sj} + \dots + x_p\beta_{sp})}}{\frac{\exp(\beta_{lo} + \dots + (x_j + c)\beta_{lj} + \dots + x_p\beta_{lp})}{1 + \sum_{s=1}^q \exp(\beta_{so} + \dots + (x_j + c)\beta_{sj} + \dots + x_p\beta_{sp})}} \\ &= \frac{\exp(\beta_{ro} + \dots + (x_j + c)\beta_{rj} + \dots + x_p\beta_{rp})}{\exp(\beta_{lo} + \dots + (x_j + c)\beta_{lj} + \dots + x_p\beta_{lp})} \end{aligned} \quad (6)$$

Für den Vergleich der Chance der Kategorie r zu der Kategorie l folgt:

$$\begin{aligned} \frac{\frac{P(Y=r|\mathbf{x} \text{ mit } x_j+c)}{P(Y=l|\mathbf{x} \text{ mit } x_j+c)}}{\frac{P(Y=r|\mathbf{x})}{P(Y=l|\mathbf{x})}} &\stackrel{6}{=} \frac{\frac{\exp(\beta_{ro} + \dots + (x_j + c)\beta_{rj} + \dots + x_p\beta_{rp})}{\exp(\beta_{lo} + \dots + (x_j + c)\beta_{lj} + \dots + x_p\beta_{lp})}}{\frac{\exp(\beta_{ro} + \dots + x_j\beta_{rj} + \dots + x_p\beta_{rp})}{\exp(\beta_{lo} + \dots + x_j\beta_{lj} + \dots + x_p\beta_{lp})}} \\ &= \frac{\exp((x_j + c)\beta_{rj})\exp(x_j\beta_{lj})}{\exp((x_j + c)\beta_{lj})\exp(x_j\beta_{rj})} = \frac{\exp((x_j + c)\beta_{rj} - x_j\beta_{rj})}{\exp((x_j + c)\beta_{lj} - x_j\beta_{lj})} \\ &= \frac{\exp(c\beta_{rj})}{\exp(c\beta_{lj})} = \exp(c(\beta_{rj} - \beta_{lj})) \end{aligned}$$

Alternativen zum Logit-Modell sind das Probit-Modell und das c-Log-Log-Modell. Beim Probit-Modell wird für die Auftrittswahrscheinlichkeit π_r nicht (3) angenommen, sondern die Standardnormalverteilung:

$$\pi_r = \phi(x^T \beta_r)$$

Beim c-Log-Log Modell wird die Auftrittswahrscheinlichkeit

$$\pi_r = 1 - \exp(-\exp(x^T \beta_r))$$

verwendet.

2.2 Vorteile ordinaler Methoden

Das mehrkategoriale Logit-Modell missachtet bei Anwendung auf geordnete, kategoriale Daten die ordinale Struktur der Variablen und nutzt damit nicht die volle Information aus. Deshalb sind ordinale Methoden für Daten mit geordneten Kategorien zu bevorzugen. Agresti gibt in seinem Buch „Analysis of Ordinal Categorical Data“ (siehe Agresti (2010): S. 3 f.) den Hinweis, dass beim Aufdecken relevanter Trends und alternativer Nachweise bei Tests des Effekts von Kovariablen auf eine ordinale Zielvariable ordinale Methoden eine größere Power aufweisen als nominale Methoden.

Die Power ist ein Maß für die Qualität eines statistischen Tests. Sie gibt die Wahrscheinlichkeit an eine Nullhypothese zu verwerfen, wenn die Alternative richtig ist (vgl. Fahrmeir et al. (2011): S. 420-424). Zudem kann bei der Analyse ordinaler Daten auf ein größeres Verfahrensspektrum zurückgegriffen werden, da Methoden stetiger Variablen wie Korrelation oder Mittelwerte in ähnlicher Form angewendet werden können (vgl. Agresti (2010): S. 3). Der Grund dafür ist, dass sowohl bei den stetigen als auch ordinalen Skalen eine Ordnung der Daten vorliegt. Außerdem können aus stetigen Variablen durch Intervallbildung ordinale Größen gebildet werden (siehe Einleitung). Das folgende Kapitel, welches das kumulative und das Adjacent-Type Logit-Modell vorstellt, beweist zudem, dass die ordinalen Modelle parametersparsamer als das multinomiale Logit-Modell sind (vgl. Tutz (2012): S. 241).

3 Ordinale Modelle

In Kapitel 2 wurde das multinomiale Modell vorgestellt, aus dessen kanonischem Link das multinomiale Logit-Modell resultiert. Es missachtet jedoch bei Anwendung auf Regressionsprobleme mit ordinaler Zielvariable deren zugrundeliegende Ordnung. Deshalb werden in diesem Kapitel zwei Modelle vorgestellt, welche die Ordnung der Kategorien ausnutzen. Zunächst wird das kumulative Modell erläutert.

3.1 Kumulatives Modell

Eine Möglichkeit zur Konstruktion von Modellen, welche die Ordnung beachten, ist benachbarte Kategorien in zwei Gruppen zusammenzufassen, sodass eine binäre Variable entsteht (vgl. Agresti (2010): S. 44). Deren kumulierte Wahrscheinlichkeiten werden dann miteinander verglichen.

3.1.1 Motivation

Die folgende Motivation des kumulativen Modells basiert auf Fahrmeir et al. (2009) S. 242-245. Es wird angenommen, dass es hinter der ordinalen, beobachtbaren Responsevariable Y_i mit Kategorien $r = 1, \dots, k$ eine latente (nicht beobachtbare) Variable Y_i^* gibt. Dieses Y_i^* ist definiert als:

$$Y_i^* = -\mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i \quad (7)$$

mit \mathbf{x}_i als Designvektor, $\boldsymbol{\beta}$ als Parametervektor und ϵ_i als Störvariable mit Verteilungsfunktion F . Das sogenannte Schwellenwertkonzept verknüpft die latente Variable und die Beobachtungen durch folgende Annahme:

$$Y_i = r \Leftrightarrow \theta_{r-1} < Y_i^* \leq \theta_r \text{ mit } r = 1, \dots, k-1 \quad (8)$$

Für die Schwellenwerte $\theta_0, \dots, \theta_k$ wird angenommen, dass diese geordnet sind mit $-\infty = \theta_0 < \theta_1 < \dots < \theta_k = \infty$. Liegt also eine latente Variable zwischen den Schwellen θ_{r-1} und θ_r , so liegt eine Beobachtung der Kategorie r vor. Da beim kumulativen Modell durch Zusammenfassung aneinandergrenzender Kategorien eine binäre Zielvariable mit Ausprägungen $Y_i \leq r$ und $Y_i > r$ entstehen soll, wird folgende Auftrittswahrscheinlichkeit betrachtet:

$$P(Y_i \leq r) = P(Y_i = 1) + \dots + P(Y_i = r) \text{ mit } r = 1, \dots, q \quad (9)$$

Diese werden als kumulierte Wahrscheinlichkeiten bezeichnet. Bei Anwendung des Schwellenwertkonzeptes ergibt sich als Zielvariable $Y_i \leq r$, falls $Y_i^* \leq \theta_r$ und $Y_i > r$, wenn $Y_i^* > \theta_r$ ist. Abbildung 1 zeigt, wie sich die Chance auf eine Kategorie bei einer Veränderung der Einflussgrößen ändert. Es sind die Dichten der latenten Variable Y_i^* für zwei unterschiedliche Ausprägungen von \mathbf{x}_i , also \mathbf{x}_{i1} und \mathbf{x}_{i2} , dargestellt. Beide Dichten haben die gleiche Form, da jeweils $\epsilon_i \sim F$ gilt. Auf der x-Achse sind die Schwellenwerte markiert. Die Fläche zwischen zwei Schwellenwerten unter der Dichtekurve gibt an mit welcher Wahrscheinlichkeit eine bestimmte Kategorie auftritt. Dadurch, dass die Dichtekurven aufgrund der unterschiedlichen Ausprägungen der Einflussgrößen auf den latenten Kontinuum verschoben sind, verändert sich die Wahrscheinlichkeit für eine bestimmte Responsekategorie bei Änderung von \mathbf{x}_{i1} zu \mathbf{x}_{i2} . Zum Beispiel ist die Wahrscheinlichkeit für ein Subjekt mit Ausprägungen \mathbf{x}_{i2} auf Kategorie 2 viel kleiner als die Wahrscheinlichkeit für ein Subjekt mit den Ausprägungen \mathbf{x}_{i1} auf Kategorie 2, wie die grauen Flächen unter den Dichtekurven zwischen θ_1 und θ_2 zeigen.

Aus (7) und (8) folgt die Definition des kumulativen Modells:

$$P(Y_i \leq r | \mathbf{x}_i) = F(\theta_r + \mathbf{x}_i^T \boldsymbol{\beta}), \quad r = 1, \dots, k$$

Die Auftrittswahrscheinlichkeit einer Kategorie r wird definiert als:

$$P(Y_i = r | \mathbf{x}_i) = F(\theta_r + \mathbf{x}_i^T \boldsymbol{\beta}) - F(\theta_{r-1} + \mathbf{x}_i^T \boldsymbol{\beta}), \quad r = 2, \dots, k - 1$$

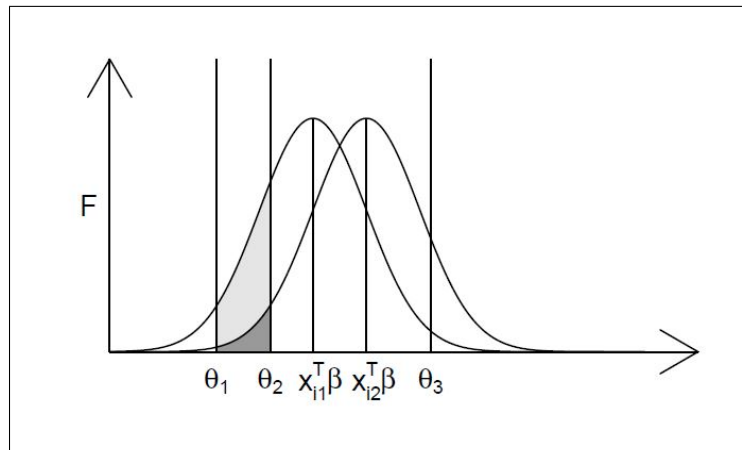


Abbildung 1: Schwellenwerte und Dichten der latenten Variable für die unterschiedlichen Designvektoren \mathbf{x}_{i1} und \mathbf{x}_{i2} (eigene Darstellung nach Fahrmeir et al. (2009): S. 243)

Für die Verteilungsfunktion F von ϵ_i können verschiedene Varianten verwendet werden. Für das kumulative Proportional-Odds Modell, welches im nächsten Abschnitt vorgestellt wird, gilt:

$$P(Y_i \leq r | \mathbf{x}_i) = \frac{\exp(\theta_r + \mathbf{x}_i^T \boldsymbol{\beta})}{1 + \exp(\theta_r + \mathbf{x}_i^T \boldsymbol{\beta})}$$

Die äquivalente Form dazu ist:

$$\log \frac{P(Y_i \leq r | \mathbf{x}_i)}{P(Y_i > r | \mathbf{x}_i)} = \theta_r + \mathbf{x}_i^T \boldsymbol{\beta}$$

Wird für F jedoch die Minimum-Extremwertverteilung bzw. Gompertzverteilung $F(x) = 1 - \exp(-\exp(x))$ gewählt, so ergibt sich das kumulative Extremwertmodell bzw. Complementary Log-Log-Modell:

$$P(Y_i \leq r | \mathbf{x}_i) = 1 - \exp(-\exp(\theta_r + \mathbf{x}_i^T \boldsymbol{\beta}))$$

Dieses ist äquivalent zu

$$\log(-\log(P(Y_i > r | \mathbf{x}_i))) = \theta_r + \mathbf{x}_i^T \boldsymbol{\beta}$$

Es wird auch als „gruppiertes Cox“- bzw. „Proportional Hazards“-Modell bezeichnet, da es in der Verweildaueranalyse Anwendung findet.

3.1.2 Definition des kumulativen Logit-Modells unter Proportionalitätsannahme

Beim kumulativen Modell werden die Kategorien in zwei Gruppen aufgeteilt. Einmal in die Gruppe $Y \leq j$ und in die Gruppe $Y > j$. Die folgenden Definitionen basieren auf Agresti (2010) S. 44-53 sowie Agresti (2013) S.275. Die kumulativen Logits werden unter Verwendung der kumulativen Wahrscheinlichkeiten (9) wie folgt definiert:

$$\text{logit}(P(Y \leq r)) = \log \frac{P(Y \leq r)}{P(Y > r)} = \log \frac{P(Y \leq r)}{1 - P(Y \leq r)} = \log \frac{\pi_1 + \dots + \pi_r}{\pi_{r+1} + \dots + \pi_k}$$

$$\text{mit } r = 1, \dots, (k-1) = 1, \dots, q$$

Mit Variation von r über alle möglichen Responsekategorien von Y ergeben sich $k - 1$ verschiedene Logits. Der Vorteil dieser Logits zu den ursprünglichen mehrkategorialen Logits aus Kapitel 2.1.2 ist, dass bei der Berechnung alle k Responsekategorien einfließen. Im Folgenden wird das Proportional-Odds Modell vorgestellt, welches alle $k - 1$ kumulativen Logits in einem

einzigem, sparsamen Modell verwendet und die erklärenden Variablen für die betrachtete Responsevariable mit einbezieht.

Das Proportional-Odds Modell ist unter Berücksichtigung des Vektors der Kovariablen $\mathbf{x} = (x_1, \dots, x_p)$ definiert als:

$$\text{logit}(P(Y \leq r|\mathbf{x})) = \log \left(\frac{P(Y \leq r|\mathbf{x})}{P(Y > r|\mathbf{x})} \right) = \alpha_r + \boldsymbol{\beta}^T \mathbf{x} = \alpha_r + \beta_1 x_{i_1} + \beta_2 x_{i_2} + \dots + \beta_p x_{i_p}$$

mit $r = 1, \dots, (k - 1)$

Dabei hat jeder kumulative Logit einen eigenen Intercept α_r , wobei sich $P(Y \leq r|\mathbf{x})$ und somit die Intercepts bei steigendem r erhöhen. Zudem gilt $\alpha_0 = -\infty$ und $\alpha_k = \infty$. Äquivalent dazu kann die Wahrscheinlichkeit für $Y \leq r$ definiert werden als:

$$P(Y \leq r|\mathbf{x}) = \frac{\exp(\alpha_r + \boldsymbol{\beta}^T \mathbf{x})}{1 + \exp(\alpha_r + \boldsymbol{\beta}^T \mathbf{x})}$$

Für die Wahrscheinlichkeit von $Y = r$ gilt:

$$\begin{aligned} P(Y = r|\mathbf{x}) &= P(Y \leq r|\mathbf{x}) - P(Y \leq r - 1|\mathbf{x}) \\ &= \frac{\exp(\alpha_r + \boldsymbol{\beta}^T \mathbf{x})}{1 + \exp(\alpha_r + \boldsymbol{\beta}^T \mathbf{x})} - \frac{\exp(\alpha_{r-1} + \boldsymbol{\beta}^T \mathbf{x})}{1 + \exp(\alpha_{r-1} + \boldsymbol{\beta}^T \mathbf{x})} \end{aligned} \quad (10)$$

Beim Proportional-Odds Modell wird angenommen, dass für jeden Logit der Koeffizientenvektor $\boldsymbol{\beta}$ identisch ist. Dies führt zu einem sparsameren Modell als das mehrkategoriale Logit-Modell in Kapitel 2, welches für jeden Logit einen eigenen separaten Parameter annimmt. Zur Motivation des kumulativen Logit-Modells und zum besseren Verständnis seiner Interpretation wird zunächst angenommen, dass die einzige Einflussgröße eine stetige Variable x ist. Das Proportional-Odds Modell ist dann definiert als:

$$\text{logit}(P(Y \leq r|\mathbf{x})) = \alpha_r + \beta x$$

Abbildung 2 zeigt für eine ordinale Variable Y mit vier Kategorien die zugehörigen kumulativen Wahrscheinlichkeiten. Diese haben durch die Annahme der Proportionalität einen gleichen Kurvenverlauf, da der Parameter β als identisch angenommen wird. Die Größe von $|\beta|$ gibt an wie stark die Kurve steigt bzw. fällt. Bei fixierten x sind die Werte der kumulierten Wahrscheinlichkeiten geordnet, wobei der Wert für $P(Y \leq 1)$ immer am kleinsten und für $P(Y \leq k - 1)$ am größten ist.

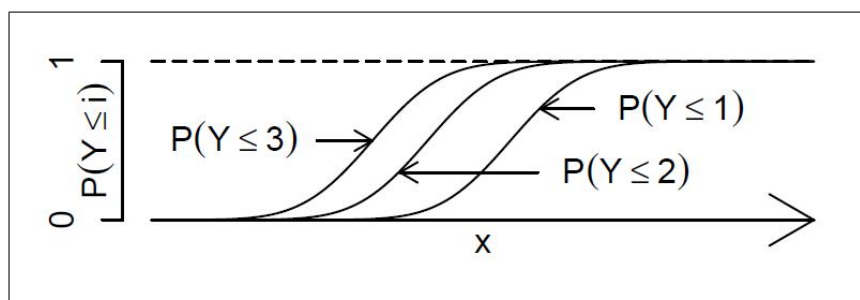


Abbildung 2: Darstellung der Wahrscheinlichkeiten des kumulativen Proportional-Odds Logit-Modell bei vier Kategorien der Zielvariablen Y (eigene Darstellung nach (Agresti (2010): S. 47))

Für $s > r$ und $s \in 1, \dots, k-1$ entspricht der Kurvenverlauf von $P(Y \leq s)$ dem von $P(Y \leq r)$, wobei dieser um $\alpha_s - \alpha_r / \beta$ Einheiten verschoben ist. Je größer die Differenz zwischen α_s und α_r ist, desto größer ist der horizontale Abstand zwischen den Kurven. Zur Interpretation der Koeffizienten wird die Veränderung der Chance auf $Y \leq r$ zu $Y > r$ bei Änderung des Wertes der einzigen Einflussvariable x von x_1 zu x_2 , mit $x_1 \leq x_2$, betrachtet:

$$\begin{aligned} & \text{logit}(P(Y \leq r|x_2)) - \text{logit}(P(Y \leq r|x_1)) \\ &= \log \frac{P(Y \leq r|x_2)/P(Y > r|x_2)}{P(Y \leq r|x_1)/P(Y > r|x_1)} = \frac{\alpha_r + \beta x_1}{\alpha_r + \beta x_2} = \beta(x_2 - x_1) \end{aligned} \quad (11)$$

Der kumulative Log-Odds-Ratio ist also proportional zu der Distanz zwischen x_2 und x_1 . Zur Interpretation wird $\exp(\beta(x_2 - x_1))$ betrachtet. Dieses besagt, dass die Chance für die Responsekategorie r oder niedriger im Verhältnis zu höheren Kategorien um den Faktor $\exp(\beta(x_2 - x_1))$ bei einem positiven Vorzeichen steigt bzw. bei einem negativen Vorzeichen sinkt, wenn sich x_1 zu x_2 ändert. Beträgt die Distanz zwischen x_2 und x_1 gleich 1, so ändert sich die Chance um den Faktor $\exp(\beta)$. Formel (11) zeigt, dass die Odds unabhängig von den betrachteten Kategorien sind. Durch diese Eigenschaft erhält das Modell den Namen Proportional-Odds Modell.

3.1.3 Kumulatives Logit-Modell für Kontingenztabelle

Die folgenden Ausführungen beruhen auf Agresti (2010) S. 49-51. Zur Vereinfachung wird nur eine einzige Einflussgröße betrachtet, welche kategorial ist. Somit wird der Prädiktor als Faktor in das Modell aufgenommen. Dargestellt in einer Kreuztabelle entsprechen die Zeilen den Ausprägungen der erklärenden Variablen X und die Spalten den Kategorien der Responsevariable Y . Im Fall einer nominal skalierten Prädiktorvariable mit Kategorien $1, 2, \dots, l$ erhalten die

Logits die Form

$$\text{logit}(P(Y \leq r)) = \alpha_r + \tau_1 z_1 + \tau_2 z_2 + \dots + \tau_{l-1} z_{l-1}$$

mit $r = 1, \dots, q$ und $z_j = 1$, wenn es sich um eine Beobachtung aus Zeile j handelt, wobei gilt $j \in 1, \dots, l$. Ansonsten gilt $z_j = 0$. Für die Identifizierbarkeit des Modells wird zudem $\tau_l = 0$ oder $\sum_{i=1}^l \tau_i = 0$ angenommen. Dabei entspricht τ_j dem Effektparameter für Zeile j .

Die Differenz zwischen den Effektparametern von Kategorie a zu b der ordinalen Variable X entsprechen dem kumulativen Log-Odds-Ratio für ordinale erklärende Variablen:

$$\text{logit}(P(Y \leq r|X = b)) - \text{logit}(P(Y \leq r|X = a)) = \tau_b - \tau_a$$

Der kumulative Log-Odds-Ratio für die Zeilen a und b ist für alle Kategorien $r = 1, \dots, q$ identisch. Gilt $\tau_1 = \tau_2 = \dots = \tau_l$, so sind Y und X unabhängig.

Ist die erklärende Variable ordinal, so ist es besser jeder Kategorie bzw. Zeile numerische Scores zuzuordnen, da damit die Ordnung der Kategorien beachtet wird. Diese Scores werden anschließend wie eine quantitative Variable behandelt. Sie werden im Folgenden als u_j mit $j = 1, \dots, l$ bezeichnet und sind geordnet in $u_1 \leq u_2 \leq \dots \leq u_l$. Werden diese in das Logit-Modell aufgenommen, so wird das Modell

$$\text{logit}(P(Y \leq r|X = u_j)) = \alpha_r + \beta u_j$$

mit $r = 1, \dots, q$ und $j = 1, \dots, (l - 1)$ erhalten. Für die Interpretation von β werden zwei Zeilen j und $j + 1$ mit $j = 1, \dots, l - 1$ und die Differenz der Log-Odds betrachtet:

$$\text{logit}(P(Y \leq r|X = u_{j+1})) - \text{logit}(P(Y \leq r|X = u_j)) = \beta(u_{j+1} - u_j) \quad (12)$$

Dieser sogenannte kumulative Log-Odds-Ratio ist proportional zur Distanz zwischen den Zeilen und für fixiertes j sind die Werte der Odds-Ratios über alle Kategorien $1, \dots, q$ identisch. Beträgt die Distanz zwischen u_j und $u_j + 1$ gleich 1, so vereinfacht sich (12) zu:

$$\text{logit}(P(Y \leq r|X = u_{j+1})) - \text{logit}(P(Y \leq r|X = u_j)) = \beta$$

Bei gleicher Distanz zwischen den Zeilenscores wird dieses Modell mit nur einer ordinalen Einflussgröße und Zielgröße auch als kumulatives Logit Uniform Association Modell bezeichnet.

3.1.4 Inferenz

Die Erläuterungen zu der Inferenz über Modellparameter und Wahrscheinlichkeiten in diesem Abschnitt basieren auf Agresti (2010) S. 58-61. Wenn \mathbf{x}_i der Vektor mehrerer erklärender Variablen für Person i ist, so ist die Wahrscheinlichkeit von Kategorie r bei gegebenen \mathbf{x}_i gleich $\pi_r(\mathbf{x}_i) = P(Y_i = r | \mathbf{X} = \mathbf{x}_i) = P(Y_i \leq r | \mathbf{x}_i) - P(Y_i \leq r - 1 | \mathbf{x}_i)$. Die Likelihood-Funktion des multinomialen Modells (2) erweitert sich dann im Fall einer ordinaler Responsevariable zu:

$$\begin{aligned} L(\beta) &= \prod_{i=1}^n \left(\prod_{r=1}^k \pi_r(\mathbf{x}_i)^{y_{ir}} \right) = \prod_{i=1}^n \left(\prod_{r=1}^k [P(Y_i \leq r | \mathbf{x}_i) - P(Y_i \leq r - 1 | \mathbf{x}_i)]^{y_{ir}} \right) \\ &\stackrel{(10)}{=} \prod_{i=1}^n \left(\prod_{r=1}^k \left[\frac{\exp(\alpha_r + \beta^T \mathbf{x}_i)}{1 + \exp(\alpha_r + \beta^T \mathbf{x}_i)} - \frac{\exp(\alpha_{r-1} + \beta^T \mathbf{x}_i)}{1 + \exp(\alpha_{r-1} + \beta^T \mathbf{x}_i)} \right]^{y_{ir}} \right) \end{aligned}$$

Zur Berechnung der Log-Likelihood-Funktion wird

$$G(z) = \frac{\exp(z)}{1 + \exp(z)} \quad g(z) = \frac{\exp(z)}{(1 + \exp(z))^2}$$

verwendet, woraus folgt:

$$l(\beta) = \sum_{i=1}^n \left(\sum_{r=1}^k y_{ir} x_{ik} \frac{g(\alpha_r + \beta^T \mathbf{x}_i) - g(\alpha_{r-1} + \beta^T \mathbf{x}_i)}{G(\alpha_r + \beta^T \mathbf{x}_i) - G(\alpha_{r-1} + \beta^T \mathbf{x}_i)} \right)$$

Zur Lösung von $s(\beta) = \frac{\partial l(\beta)}{\partial \beta} \stackrel{!}{=} 0$ wird der Fisher-Scoring-Algorithmus verwendet, ein iteratives Verfahren bei dem in jedem Schritt neu gewichtete kleinste Quadrate verwendet werden. Bei jeder Iteration wird also die nicht konstante Varianz der Beobachtungen beachtet, indem die Gewichte variieren bis die Schätzung konvergiert.

Bereits 1980 hat McCullagh in seinem Artikel „Regression Models for Ordinal Data“ (siehe Peter McCullagh (1980)) gezeigt, dass bei einem suffizient großen Stichprobenumfang n ein einziges Maximum der Likelihood-Funktion erreicht wird. Zu beachten ist, dass für endliches n die Einzigartigkeit der Lösung nicht gewährleistet sein kann oder unendliche Koeffizienten geschätzt werden können.

Um zu testen, ob eine Variable x_j mit $j \in 1, \dots, p$ Einfluss auf die Wahl der Responsekategorie hat, werden Nullhypothesen der Form $H_0 : \beta_j = 0$ betrachtet. Wird als Testverfahren der Wald-Test verwendet, so wird für seine Testgröße $z = \hat{\beta}_j / \text{SE}$ der Standardfehler SE benötigt. Zur Schätzung des Standardfehlers wird die Fisher-Informationsmatrix verwendet. Die Informationsmatrix hat zwei mögliche Versionen. Zum einen gibt es die beobachtete Fisher-

Informationsmatrix (vgl. Fahrmeir et al. (2009) S. 469)

$$F_{\text{obs}}(\boldsymbol{\beta}) = -\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}$$

und zum anderen die erwartete Fisher-Informationsmatrix (vgl. Fahrmeir et al. (2009): S. 470)

$$F(\boldsymbol{\beta}) = E(F_{\text{obs}}(\boldsymbol{\beta})) = E\left(-\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}\right)$$

Das Inverse der erwarteten Informationsmatrix fließt beim Fisher-Scoring-Algorithmus ein. Wird stattdessen die beobachtete Fisher-Informationsmatrix benutzt, so wird der Newton-Raphson-Algorithmus verwendet. Die Standardfehler ergeben sich aus der Wurzel der Diagonalelemente beider Versionen von Informationsmatrizen. Somit kann der Wald-Test zum Testen der Nullhypothese $H_0 : \beta_j = 0$ angewendet werden. Unter H_0 wird z als approximativ $\chi^2(df = 1)$ -verteilt angenommen. Allerdings weist Agresti darauf hin, dass bei der Anwendung des Tests die Stichprobengröße nicht zu klein sein und kein großer Anteil der Beobachtungen in die höchste oder niedrigste Kategorie der Responsevariable fallen sollte (siehe Agresti (2010): S. 60). Dann kann die Annahme, dass $(\hat{\beta}_j - \beta_j)/SE$ standardnormalverteilt ist, nicht mehr erfüllt sein und es sollte auf einen Likelihood-Ratio-Test zurückgegriffen werden. Er spezifiziert allerdings nicht, ab welcher Anzahl der Beobachtungen und welchem Anteil dies der Fall ist. Für die Teststatistik des Likelihood-Ratio-Tests gilt:

$$LR = -2(L_0 - L_1) \tag{13}$$

L_0 ist die maximierte Log-Likelihood-Funktion unter Annahme der Nullhypothese und L_1 die maximierte Log-Likelihood-Funktion ohne Einschränkung. Die Teststatistik ist unter Annahme der Nullhypothese wie beim Wald-Test approximativ $\chi^2(df = 1)$ verteilt.

3.1.5 Methoden zur Modellüberprüfung

Liegen die Daten als Kreuztabelle vor, so ist ein Test der Anpassungsgüte durch eine Pearson- oder Likelihood-Ratio-Statistik möglich. Diese testen das vorhandene Modell gegen ein Modell, welches die Daten perfekt fittet. Für die Tests wird die beobachtete Anzahl in einer Zelle der Tabelle mit der erwarteten Anzahl im perfekt gefitteten Modell verglichen. Die folgenden Spezifikationen beruhen auf Agresti (2010) S. 67-69. Wenn n_i die Anzahl an Beobachtungen für einen Vektor der Einflussgrößen \boldsymbol{x}_i und n_{ir} mit $r = 1, \dots, k$ die beobachtete Anzahl für die

r -te Responsekategorie sind, dann wird die erwartete Anzahl folgendermaßen geschätzt:

$$\hat{\mu}_{ir} = n_i \hat{P}(Y = r | \mathbf{x}_i) \text{ mit } r = 1, \dots, k$$

Die Pearson-Statistik zum Test der Anpassungsgüte unter Verwendung der geschätzten erwarteten Anzahl wird definiert als:

$$\chi^2 = \sum_{i=1}^n \sum_{r=1}^k \frac{(n_{ir} - \hat{\mu}_{ir})^2}{\hat{\mu}_{ir}}$$

Daraus folgt die Likelihood-Ratio-Statistik:

$$G^2 = 2 \sum_{i=1}^n \sum_{r=1}^k n_{ir} \log \frac{n_{ir}}{\hat{\mu}_{ir}}$$

Unter der Nullhypothese sind sowohl die Pearson-Statistik als auch die Likelihood-Ratio-Statistik χ^2 -verteilt. Die Anzahl der Freiheitsgrade ergibt sich aus der Differenz der Anzahl der modellierten kumulativen Logits zu der Zahl der Modellparameter. Die Anzahl der modellierten kumulativen Logits erhält man aus der Anzahl der multinomialen Parameter im perfekt gefitteten Modell, also $(k - 1)$ -mal die Anzahl der Kategorien der erklärenden Variable. In einer $(l \times k)$ -Kontingenztabelle wären dies also $l(k - 1)$ Parameter. Im Fall einer ordinalen Einflussgröße mit dem Logits $\text{logit}(P(Y \leq r | X = j)) = \alpha_r + \beta u_j$ mit $r = 1, \dots, q$ ist die Anzahl der Modellparameter gleich $q + 1 = k$. Somit ergeben sich die Freiheitsgrade in diesem Fall als $df = l(k - 1) - k = lk - l - k$. Ist die kategoriale Einflussgröße nicht geordnet, dann werden die Logits $\text{logit}(P(Y \leq r | X = j)) = \alpha_r + \tau_j$ betrachtet. In diesem Fall sind es $(k - 1) + (l - 1)$ Modellparameter. Somit werden für die χ^2 -Verteilung $df = l(k - 1) - ((k - 1) + (l - 1)) = (l - 1)(k - 2)$ Freiheitsgrade betrachtet. In Abschnitt 3.1.4 wurde bereits erwähnt, dass bei Daten mit wenigen Beobachtungen oder einer Verteilung der Beobachtungen auf wenige Kategorien unendliche Koeffizienten geschätzt werden könnten. In diesem Fall ist die Anwendung eines globalen Goodness of Fit-Tests nicht möglich. Es kann aber auf eine Pearson-Statistik zurückgegriffen werden, welche die ursprünglichen, ungruppierten Daten verwendet. Genauere Informationen dazu befinden sich in Lipsitz et al. (1996). Sind die Einflussgrößen nicht qualitativ, sondern stetig, so kann der oben beschriebenen Goodness of Fit-Test nicht angewendet werden. Um trotzdem zu testen, wie gut das Modell die Daten fittet, kann auf den Vergleich zu alternativen, komplexeren Modellen zurückgegriffen werden. Wird nur eine Einflussgröße betrachtet, so kann dessen Fit zum Beispiel mit einem Modell

unter zusätzlicher quadratischer Aufnahme der Einflussgröße verglichen werden. Falls mehrere Einflussgrößen in das Modell aufgenommen werden, kann zum Vergleich ein Modell mit zusätzlicher Aufnahme von Interaktionstermen verwendet werden. Ob das größere, komplexere Modell mit zusätzlichen Term(en) einen besseren Fit bewirkt, kann anhand eines Likelihood-Ratio-Tests, wie er bereits in Abschnitt 3.1.4 vorgestellt wurde, beurteilt werden. Dessen Teststatistik $LR = -2(L_0 - L_1)$ ist approximativ χ^2 -verteilt mit der Anzahl an zusätzlichen Parametern des komplexeren Modells als Freiheitsgrade. Ein kleiner p-Wert besagt, dass das einfachere Modell verworfen werden sollte.

Alternativ kann auch das Akaike Information Criterion, kurz AIC, betrachtet werden (vgl. Agresti (2010): S. 75)). Dieses ist definiert als (vgl. Fahrmeir et al. (2009): S. 206)

$$AIC = -2(l(\hat{\beta}) - p)$$

mit p als Anzahl an Parametern im Modell. Der Vorteil des AICs ist, dass die Zahl der Parameter bei der Berechnung einfließt und mit steigender Anzahl ein höheres AIC zur Folge hat. Sparsame Modelle können ebenfalls eine den wahren Werten nahe Schätzung liefern und sind dabei einfacher in ihrer Berechnung. Ist die Anzahl an Parametern identisch, so kann auch die Devianz (vgl. Tutz (2012): S. 67)

$$D = -2\phi \sum_{i=1}^n (l(\hat{\beta}) - l(y))$$

als Vergleichskriterium verwendet werden, mit ϕ als zugehöriger Dispersionsparameter.

3.1.6 Erweiterung des kumulativen Logit-Modells

Häufig ist die Annahme der Proportionalität nicht angemessen und statt eines einzigen Parametervektors β für alle Kategorien der Zielgröße, sollte für jede Responsekategorie ein eigenes β_r mit $r = 1, \dots, q$ geschätzt werden. Somit ergibt sich ein komplexeres Logit-Modell nach Agresti (2010) S. 70f.. Dieses wird definiert als:

$$\text{logit}(P(Y \leq r)) = \alpha_r + \beta_r^T \mathbf{x}$$

Das um $(k - 2)$ Parameter erweiterte Modell hat jedoch das strukturelle Problem, dass es vor allem bei einer großen Anzahl an Prädiktoren nicht verwendet werden kann, da die Maximierung der Likelihood-Funktion und damit die Schätzung der Parameter zu komplex ist. Um zu testen, ob ein Modell ohne Annahme der Proportionalität verwendet werden sollte, kann ein Score-

Test durchgeführt werden. Dieser liefert asymptotisch ähnliche Ergebnisse wie der Wald- und Likelihood-Ratio-Test. Er hat aber den Vorteil, dass er nur die Log-Likelihood-Funktion unter der Nullhypothese $H_0 : \beta_1 = \beta_2 = \dots = \beta_q$ betrachtet und nicht die Funktion unter der komplexen Alternative. Ist der p-Wert größer als 0.05, so kann angenommen werden, dass das komplexere Modell keinen signifikant besseren Fit liefert. Dieser Wert gibt jedoch nur Auskunft über die statistische Signifikanz, reflektiert aber nicht die Praktikabilität. Deshalb wird trotz Tests häufig das einfachere Modell wegen seiner Sparsamkeit bevorzugt. Das Ergebnis des Score-Tests ist zudem fragwürdig bei Daten mit wenigen Gesamtbeobachtungen oder wenn ein großer Teil der Beobachtungen in wenige Responsekategorien fallen.

Einen besonderen Fall bildet das kumulative Partial-Proportional-Odds Logit-Modell. Dieses beinhaltet sowohl Einflussgrößen \mathbf{x} mit Annahme von Koeffizienten mit Proportional-Odds Struktur als auch Kovariablen mit kategoriespezifischen Koeffizienten. Letztere werden im Folgenden im Vektor \mathbf{u} zusammengefasst. Das Modell wird dann nach Agresti (2010) S. 78 als

$$\text{logit}(P(Y \leq r)) = \alpha_r + \beta^T \mathbf{x} + \gamma_r^T \mathbf{u} \quad (14)$$

mit $r = 1, \dots, q$ definiert. Zur Identifizierbarkeit wird angenommen, dass $\gamma_1 = 0$ gilt, wobei ein anderes γ_r auch gewählt werden kann. Bei der Interpretation sollte beachtet werden, dass die Parameter des Vektors β (ohne Proportionalitätsannahme) wie bisher im kumulativen Logit-Modell für jede Kategorie identisch interpretiert werden. Bei Annahme kategoriespezifischer Parameter γ_r ergibt sich für jede Kategorie eine eigene Interpretation der Wirkung der betrachteten Einflussgrößen. Gilt $\gamma_2 = \dots = \gamma_q$, so wird das gewöhnliche Proportional-Odds Modell angenommen.

3.2 Adjacent-Type Modell

In Kapitel 3.1. wurde das kumulative Logit-Modell vorgestellt, welches durch Verwendung der kumulierten Wahrscheinlichkeiten die Ordnung der Responsekategorien beachtet. Eine Alternative dazu ist das Adjacent-Category Logit-Modell. Dieses verwendet nicht die kumulierten Logits, sondern paarweise, lokale Logits (vgl. Agresti (2010): S. 96). Im Folgenden wird dieses Modell zunächst mit und ohne Annahme von Proportional-Odds sowie deren Verbindung zum multinomialen Logit-Modell vorgestellt. Danach wird das resultierende Modell für eine kategoriale Einflussgröße und die Inferenz erläutert. Die Definitionen und ihre Ergänzungen dieses Kapitels basieren auf Agresti (2010) S. 88-96.

3.2.1 Definition des Adjacent-Type Logit-Modells

Sind die multinomialen Wahrscheinlichkeiten $\pi_r(\mathbf{x})$ für die Responsevariable Y mit Kategorien $1, \dots, k$, wie beim kumulativen Modell (10), als

$$\pi_r(\mathbf{x}) = P(Y = r|\mathbf{x}) = \frac{\exp(\alpha_r + \boldsymbol{\beta}^T \mathbf{x})}{1 + \exp(\alpha_r + \boldsymbol{\beta}^T \mathbf{x})} \text{ mit } r = 1, \dots, k$$

gegeben, dann sind die Adjacent-Category Logits mit Annahme der Proportionalität definiert als:

$$\text{logit}(P(Y = r|Y = r \text{ oder } Y = r + 1, \mathbf{x})) = \log\left(\frac{\pi_{r+1}(\mathbf{x})}{\pi_r(\mathbf{x})}\right) = \alpha_r + \boldsymbol{\beta}^T \mathbf{x}$$

$$\text{mit } r = 1, \dots, k - 1$$

In diesem Modell werden also zwei benachbarte Kategorien r und $r + 1$ beachtet, sodass nicht von einem kumulativen Odds-Ratio, sondern von einem lokalen Odds-Ratio gesprochen wird. $\boldsymbol{\beta}^T \mathbf{x}$ gibt die Präferenz für die Kategorie $r + 1$ im Gegensatz zu der Kategorie r an. Für jeden Logit ergibt sich, unter Annahme, dass die erklärende Variable einen ähnlichen Effekt für jede Responsekategorie hat, ein einziges $\boldsymbol{\beta}$, statt $k - 1$ unterschiedlichen Parametervektoren für jede Betrachtung benachbarter Kategorien.

Zur Interpretation des Einflusses einer Variable x_j mit $j = 1, \dots, p$ auf die Chance einer Responsekategorie $r + 1$ zu r , wird $\exp(\beta_j)$ betrachtet. Die Chance auf ein höhere Kategorie $r + 1$ zu einer niedrigeren Stufe r entspricht dem Faktor $\exp(\beta_j)$, wenn die j -te Einflussvariable um eine Einheit steigt. Dieses Chanceverhältnis bzw. dieser Odds-Ratio ist für alle Paare benachbarter Kategorien von Y identisch, also unabhängig von r , sodass $\beta_j = \beta_{j1} = \dots = \beta_{jk}$ gilt. Das Vorzeichen von β_j gibt an, ob bei steigenden x_j niedrigere oder höhere Kategorien getroffen werden. Gilt $\beta_j > 0$ so fällt Y eher in höhere Kategorien und für $\beta_j < 0$ eher in niedrigere.

Eine Erweiterung des Modells mit Proportionalitätsannahme ist die Annahme eines eigenen Koeffizientenvektors β_r für jede Kategorie $r \in 1, \dots, k$, sodass für das Logit-Modell folgt:

$$\log \left(\frac{\pi_{r+1}(\mathbf{x})}{\pi_r(\mathbf{x})} \right) = \alpha_r + \beta_r^T \mathbf{x} \text{ mit } r = 2, \dots, k \quad (15)$$

Proportional-Odds können aber auch nur für einen Teil der Einflussgrößen, aber nicht bei allen, angenommen werden. Dieses Modell wird als Partial-Proportional-Odds Modell bezeichnet und ist analog zum Partial-Proportional-Odds Modell (14) bei der Verwendung kumulierter Logits. Zum Test der Proportionalitätsannahme können die gleichen Tests wie beim kumulativen Modell verwendet werden.

3.2.2 Verbindung zum multinomialen Logit-Modell

In Kapitel 2.1.3 wurde das multinomiale Logit-Modell vorgestellt. Dieses missachtet die Ordnung der Kategorien bei ordinalen Responsevariablen. Zur Definition benötigt es die Wahl einer Referenzkategorie. Wird als Referenz die erste Ausprägung gewählt, so erhält man als Logits $\log \frac{\pi_2}{\pi_1}, \log \frac{\pi_3}{\pi_1}, \dots, \log \frac{\pi_{k-1}}{\pi_1}, \log \frac{\pi_k}{\pi_1}$ für das multinomiale Logit-Modell. Folgende Umformung zeigt die Verbindung zu den Adjacent-Category Logits (vgl. Agresti (2010): S. 91):

$$\log \frac{\pi_r}{\pi_1} = \log \frac{\pi_2}{\pi_1} + \log \frac{\pi_3}{\pi_2} + \dots + \log \frac{\pi_r}{\pi_{r-1}} \text{ mit } r = 2, \dots, k$$

Zusammen mit (15) folgt das mehrkategoriale Logit-Modell:

$$\begin{aligned} \log \frac{\pi_r(\mathbf{x})}{\pi_1(\mathbf{x})} &= \alpha_1 + \beta_1^T \mathbf{x} + \alpha_2 + \beta_2^T \mathbf{x} + \dots + \alpha_{r-1} + \beta_{r-1}^T \mathbf{x} \\ &= \sum_{s=1}^{r-1} \alpha_s + \left(\sum_{s=1}^{r-1} \beta_s^T \right) \mathbf{x} = \alpha_r^* + \beta_r^{*T} \mathbf{x} \text{ mit } r = 2, \dots, k \end{aligned} \quad (16)$$

Unter Proportionalitätsannahme für alle Parameter vereinfacht sich das Modell zu

$$\log \frac{\pi_r(\mathbf{x})}{\pi_1(\mathbf{x})} = \sum_{s=1}^{r-1} \alpha_s + (r-1)\beta^T \mathbf{x} = \alpha_r^* + \beta^T \mathbf{z}_r, \quad r = 2, \dots, k \quad (17)$$

mit $\mathbf{z}_r = (r-1)\mathbf{x}$. Die Logits des mehrkategorialen Logit-Modells können also durch Adjacent-Type Logits ersetzt werden, allerdings wird dann die Designmatrix \mathbf{x} durch den Vektor \mathbf{z}_r ersetzt.

3.2.3 Adjacent-Type Logit-Modell für Kontingenztabelle

Wird eine einzige Einflussgröße X mit l Kategorien betrachtet, so ergibt sich eine $(l \times k)$ -Kontingenztabelle (vgl. Agresti (2010): S. 90). Die konditionale Wahrscheinlichkeit ist dann gegeben durch $\pi_{r|j} = P(Y = r|X = j)$ mit $r = 1, \dots, k - 1$ und $j = 1, \dots, l - 1$.

Im ersten Fall wird die Einflussgröße X als ordinal angenommen. Die geordneten Reihenscores von X werden im Folgenden als u_j mit $j = 1, \dots, l - 1$ bezeichnet. Um die Ordnung der Reihen nun in der proportionalen Form zu beachten, kann ähnlich wie bei Formel (12)

$$\text{logit}(P(Y = r|Y = r \text{ oder } Y = r + 1)) = \log \frac{\pi_{r+1|j}}{\pi_{r|j}} = \alpha_r + \beta u_j \quad r = 1, \dots, k - 1$$

verwendet werden. Für die Interpretation von β werden zwei Zeilen $j+1$ und j und die Differenz ihrer Logits betrachtet:

$$\text{logit}(P(Y = r|X = j + 1)) - \text{logit}(P(Y = r|X = j)) = \beta(u_{j+1} - u_j)$$

Ist die Distanz zwischen den Reihenscores identisch, so erfüllt das Modell die Uniform Association für lokale Odds-Ratios. Falls gilt $u_{j+1} - u_j = 1$, so ist der uniforme lokale Odds-Ratio gleich $\exp(\beta)$.

Im zweiten Fall wird von einem nominalen X mit ungeordneten Kategorien $1, \dots, l$ ausgegangen. Es kann ein generelleres Row Effects Modell definiert werden, bei dem statt den Parametern βu_j die ungeordneten Parameter τ_j auftreten mit $j = 1, \dots, l$:

$$\text{logit}(P(Y = r|Y = r \text{ oder } Y = r + 1)) = \log \frac{\pi_{r+1|j}}{\pi_{r|j}} = \alpha_r + \tau_j \quad \text{mit } r = 2, \dots, k$$

Für die Interpretation des Einflusses wird wieder die Differenz betrachtet:

$$\text{logit}(P(Y = r|X = j + 1)) - \text{logit}(P(Y = r|X = j)) = \tau_{j+1} - \tau_j$$

Für zwei Paare von Reihen sind die $k - 1$ Log-Odds-Ratios identisch.

3.2.4 Inferenz

Die Spezifikationen der Inferenz basieren auf Agresti (2010) S. 92 ff.. Für Subjekt i ist y_{ir} der binäre Indikator, ob der Response sich in der Kategorie r befindet und x_{ij} mit $j \in 1, \dots, p$ ist der Wert der erklärenden Variable mit $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$. Mit der Annahme von n unabhängigen

Beobachtungen ist die Log-Likelihood-Funktion über alle n Beobachtungen definiert als:

$$\begin{aligned} \log \prod_{i=1}^n \left[\prod_{r=1}^k \pi_r(\mathbf{x}_i)^{y_{ir}} \right] &= \sum_{i=1}^n \left[\sum_{r=2}^{k-1} y_{ir} \log \pi_r(\mathbf{x}_i) + \left(1 - \sum_{r=2}^k y_{ir} \right) \log \pi_1(\mathbf{x}_i) \right] \\ &= \sum_{i=1}^n \left[\sum_{r=2}^k y_{ir} \log \frac{\pi_r(\mathbf{x}_i)}{\pi_1(\mathbf{x}_i)} + \log \pi_1(\mathbf{x}_i) \right] \end{aligned}$$

Für das Modell unter Annahme der Proportionalität ergibt sich die Log-Likelihood für das Adjacent-Category Modell mit (17) zu:

$$\begin{aligned} \log \prod_{i=1}^n \left[\prod_{r=1}^k \pi_r(\mathbf{x}_i)^{y_{ir}} \right] &= \sum_{i=1}^n \left[\sum_{r=2}^k y_{ir} (\alpha_r^* + \boldsymbol{\beta}^T \mathbf{z}_{ir}) + \log \left(1 + \sum_{r=2}^k \alpha_r^* + \boldsymbol{\beta}^T \mathbf{z}_{ir} \right) \right] \\ &\quad \text{mit } \mathbf{z}_{ir} = (r-1)\mathbf{x}_i \end{aligned}$$

und mit kategoriespezifischen Parametern (siehe (16)) als:

$$\log \prod_{i=1}^n \left[\prod_{r=1}^k \pi_r(\mathbf{x}_i)^{y_{ir}} \right] = \sum_{i=1}^n \left[\sum_{r=2}^k y_{ir} (\alpha_r^* + \boldsymbol{\beta}_r^{*T} \mathbf{x}_i) + \log \left(1 + \sum_{r=2}^k \alpha_r^* + \boldsymbol{\beta}_r^{*T} \mathbf{x}_i \right) \right]$$

Wie beim kumulativen Modell kann auch hier ein Test auf Anpassungsgüte für Kontingenzta-bellen durch einen Pearson-Test oder durch einen Likelihood-Ratio-Test erfolgen.

Fittet das Modell die Daten schlecht, so können weitere Terme zum Beispiel Interaktionsterme hinzugefügt oder im Fall der Annahme von Proportionalität kategoriespezifische Parameter zu-gelassen werden.

Eine Besonderheit des Adjacent-Category Logit-Modells ist, dass bei ihm im Gegensatz zu dem kumulativen Modell die erwartete Fisher-Matrix für β (bzw. β_r) gleich der beobachte-ten Fisher-Matrix ist, also $F(\beta) = F_{\text{obs}}(\beta)$ gilt. Die Fisher-Matrix wird anhand von $F(\beta) = -(\partial^2 l(\beta)) / (\partial \beta \partial \beta^T) = E(-(\partial^2 l(\beta)) / (\partial \beta \partial \beta^T))$ erhalten. Zur Schätzung der Parameter wird wie beim kumulativen Modell der Fisher-Scoring-Algorithmus verwendet. Dieser entspricht durch die Äquivalenz der erwarteten und beobachteten Fisher-Information dem Newton-Raphson-Verfahren.

Weitere Gemeinsamkeiten und Unterschiede des kumulativen und des Adjacent-Category Logit-Modells werden im nächsten Kapitel anhand des **Allbus**-Datensatzes veranschaulicht.

4 Vergleich der Methoden

Anhand der Definitionen des kumulativen und des Adjacent-Category Modells aus Kapitel 3 zeigen sich einige Parallelen und Unterschiede der beiden Verfahren. Während das kumulative Modell auf kumulierten Logits basiert, werden beim Adjacent-Category Modell die lokalen Logits benachbarter Kategorien betrachtet. Für letzteres Modell gilt, im Gegensatz zum kumulativen Modell, dass die beobachtete Fisher-Informationsmatrix äquivalent zu ihrem Erwartungswert ist. Für beide Modelle können flexible Annahmen über die Proportionalität der Koeffizienten getroffen werden, sodass Partial-Proportional-Odds Modelle gebildet werden können. Für den Test der Proportionalitätsannahme und des Modellfits können die gleichen Teststatistiken und Kennzahlen verwendet werden.

Um weitere Unterschiede und Gemeinsamkeiten der beiden Verfahren aufzudecken, wurde das Vertrauen in die Regierung und das Vertrauen in die Mitmenschen des sogenannten **Allbus**-Datensatzes sowohl durch ein kumulatives als auch durch ein Adjacent-Category Logit-Modell gefittet. Dazu wurde zunächst durch einen Likelihood-Ratio-Test die Proportional-Odds Annahme überprüft. Die Anpassungsgüte wird anhand des AIC Kriteriums und der Devianz verglichen. Zur Berechnung wurde das **VGAM** Package in **R** verwendet, welches nun zunächst erläutert wird. Der **Allbus**-Datensatz und die Ergebnisse der Modellierung der Vertrauensskalen durch beide Verfahren werden in Abschnitt 4.2 vorgestellt.

4.1 VGAM Package

Für die Berechnung des Adjacent-Type und des kumulativen Modells wurde die Funktion `vglm()` verwendet, welche in dem Package **VGAM** implementiert ist. Die folgenden Erläuterungen basieren auf Yee (2010) S. 1 f. und Yee (2016) S.727 f.. Die Funktion `vglm()` bietet den Vorteil, dass es durch die flexible Spezifikation des „family“-Arguments für das Fitten beider Modelle verwendet werden kann. In **R** sind zwar noch weitere Funktionen zur Modellierung von kategorialen Daten vorhanden, diese sind aber entweder nur für eine Linkfunktion implementiert oder benötigen vorherige Prozesse oder Startwerte zur Berechnung. Beispiele wären die Funktionen `polr()` aus dem Package **MASS**, welches nur Proportional-Odds Modelle fittet, und `multinom()` aus dem Paket **nnet** für multinomiale Logit-Modelle. Diese sind im Comprehensive R Archive Network (kurz: CRAN) zu finden. Zudem gibt es noch weitere Funktionen von nicht-CRAN Paketen wie die Funktion `nordr()` aus dem Package **gnlm**. Dieses fittet zwar das Proportional-Odds und das Adjacent-Category Modell, zur Anwendung werden aber Startwerte benötigt. Deshalb wird das **VGAM** Package mit der Funktion `vglm()` bevorzugt.

Das Argument „family=cumulative (parallel=TRUE)“ wird zur Berechnung des kumulativen Modells verwendet, sodass die Logits

$$\text{logit}P(Y \leq r) = \log(P(Y \leq r)/P(Y > r)) \quad (18)$$

modelliert werden. Dieser Befehl ist äquivalent zu „family=propodds“. Dass die Proportional-Odds Annahme verwendet wird, besagt der Ausdruck „(parallel=TRUE)“. Falls kategoriespezifische Koeffizienten für alle Kovariablen angenommen werden, so wird „(parallel=FALSE)“ verwendet. Wird ein Partial-Proportional-Odds Modell gefittet, so wird „(parallel=FALSE~1+Name der Einflussvariable)“ gesetzt.

Der Befehl „family=acat (parallel=TRUE)“ modelliert das Adjacent-Category Logit-Modell unter Proportionalitätsannahme. Dieses verwendet die Logits

$$\text{logit}(P(Y = r|Y = r \text{ oder } Y = r + 1)) = \log(P(Y = r + 1)/P(Y = r)) \quad (19)$$

Äquivalent zum kumulativen Modell kann über das „parallel“-Argument die Proportionalitätsannahme variiert werden. Um zu testen, welche Annahme getroffen werden sollte, kann ein Likelihood-Ratio- bzw. Likelihood-Quotienten-Test durchgeführt werden, indem die Devianzen des Modells mit Proportionalitätsannahme mit dem Modell ohne dieser Annahme verglichen werden. Um die Anpassungsgüte der beiden Verfahren zu vergleichen, wurde das AIC und die Devianz beider Modelle berechnet. Deren Werte werden durch die Befehle AIC() bzw. deviance() erhalten.

4.2 Allbus-Datenbeispiel

Zum Vergleich der beiden Verfahren wurde die bisherige Theorie und die im letzten Abschnitt erläuterten **R**-Funktionen auf die Modellierung zweier Vertrauensskalen des **Allbus**-Datensatzes angewendet. Nach einer kurzen Einführung in die Erhebung der Daten und die Variablenbeschreibung, werden die Analyseergebnisse vorgestellt und deren Gemeinsamkeiten sowie Unterschiede erläutert.

4.2.1 Variablenbeschreibung und Modellspezifikationen

Der Name **Allbus** ist eine Abkürzung für „Allgemeine Bevölkerungsumfrage der Sozialwissenschaften“. Die Daten wurden zwischen April und September 2012 von der Gesellschaft TNS Infratest Sozialforschung München erhoben. Das Ziel der Studie war Trends von Einstellungen,

Verhalten und sozialem Wandel in Deutschland aufzudecken. Dazu wurden 3480 in Deutschland lebende und vor dem 01.01.1994 geborene Personen befragt. Nachdem eine Gemeinde zufällig mit einer Wahrscheinlichkeit proportional zur Einwohnergröße gezogen wurde, erfolgte eine zufällige Auswahl der Personen aus der Einwohnerkartei der Gemeinde. Die Befragung wurde mündlich zu den Themenbereichen Wichtigkeit von Lebensbereichen, individuelle und kollektive Wertorientierungen, Abtreibung, politische Einstellungen, Religion und Weltanschauung, Ethnozentrismus und Minoritäten, Gesundheit, Familie und Sonstiges ausgeführt. So entstand ein Datensatz mit 752 Variablen. Diese und weitere Informationen sind in Terwey and Baltzer (2013) zu finden. Für die Gegenüberstellung der beiden Verfahren wird eine Auswahl an Variablen verwendet. Eine Übersicht dazu befindet sich in Tabelle 1.

Name der Variable	Beschreibung
<i>government</i>	Vertrauen in die Bundesregierung
<i>fellow</i>	Vertrauen zu den Mitmenschen
<i>sex</i>	Geschlecht (Frau=1/ Mann=0)
<i>partner</i>	Partnerschaft (ja=1/ nein=0)
<i>employment</i>	Erwerbstätigkeit (ja=1/ nein=0)
<i>age</i>	Um Mittelwert 50 zentriertes Alter in Jahren
<i>age2</i>	Quadriertes um 50 zentriertes Alter
<i>income</i>	Einkommen in Euro (in 100ter Schritten)
<i>life_kat</i>	Lebenszufriedenheit
<i>interest_kat</i>	Politisches Interesse
<i>leftright_kat</i>	Politische Präferenz

Tabelle 1: Übersicht über die verwendeten Variablen des **Allbus**-Datenbeispiels

Die ersten beiden Variablen *government* und *fellow* sind die Zielvariablen, welche jeweils durch die restlichen aufgeführten Variablen gefittet werden sollen. Zur genaueren Übersicht über die Daten und Verteilungen der Einflussgrößen erfolgt die graphische Darstellung der verwendeten Variablen in den Abbildungen 3 bis 10. Zur Modellierung werden nur Beobachtungen mit vollständigen Angaben beachtet. Dies trifft auf 1422 Personen zu.

Die Variable *government* besteht aus den Kategorien 1 (kein Vertrauen) bis 7 (großes Vertrauen). Ihre Verteilung ist dargestellt in Graphik 3. Die meisten Befragten gaben zur Einschätzung ihres Vertrauens in die Regierung die Kategorie 4 (mittleres Vertrauen) an. Mit zunehmender Abweichung von diesem Modus, sinkt auch die Anzahl der Personen, die diese Werte gewählt haben. Kategorie 7 (sehr großes Vertrauen) gaben am wenigsten Personen an.



Abbildung 3: Verteilung der Variable *government* des **Allbus**-Datenbeispiels

Die Variable *fellow* ist eingeteilt in Kategorie 1 für „eher nein“, Stufe 2 für „kommt drauf an“ und Kategorie 3 für „eher ja“. Abbildung 4 zeigt, dass etwa 600 Personen eher kein Vertrauen in ihre Mitmenschen haben. Etwas weniger Befragte gaben an, dass es darauf an kommt, ob sie Anderen vertrauen oder nicht. Etwa 250 Personen sagten, sie würden ihren Mitmenschen eher vertrauen.

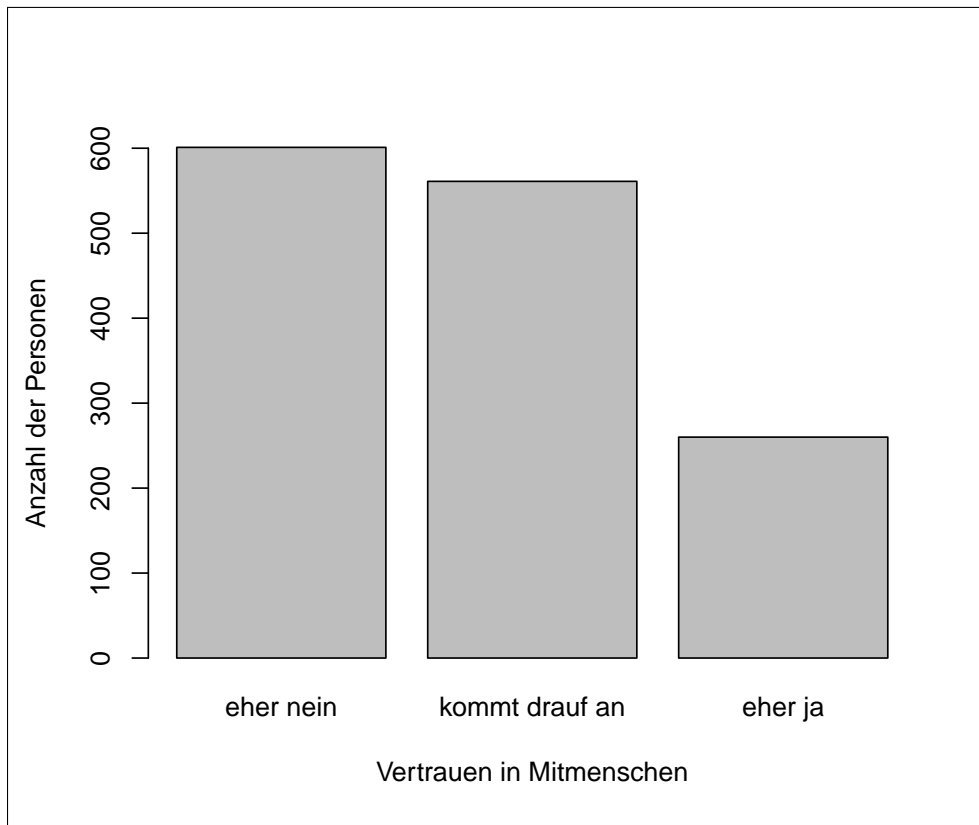


Abbildung 4: Verteilung der Variable *fellow* des *Allbus*-Datenbeispiels

Für eine Person i gilt somit für die Zielvariable *government* oder *fellow* folgende Spezifikation:

$$y_{ir} = \begin{cases} 1, & \text{Responsekategorie } r \text{ in Antwort genutzt} \\ 0, & \text{sonst} \end{cases}$$

mit $r \in \{1, 2, 3, 4, 5, 6\}$ (Ausprägung 7 als Referenzkategorie) für die Variable *government* und $r \in \{1, 2\}$ (Stufe 3 als Referenzkategorie) für die Zielgröße *fellow*.

Die drei Kreisdiagramme in Abbildung 5 zeigen die Verteilung der Variable *sex*, *partner* und *employment*. In dem betrachteten Datensatz befinden sich mehr männliche Probanden als weibliche. Fast dreiviertel der Befragten befinden sich in einer festen Partnerschaft und etwa zweidrittel gehen einer Erwerbstätigkeit nach.

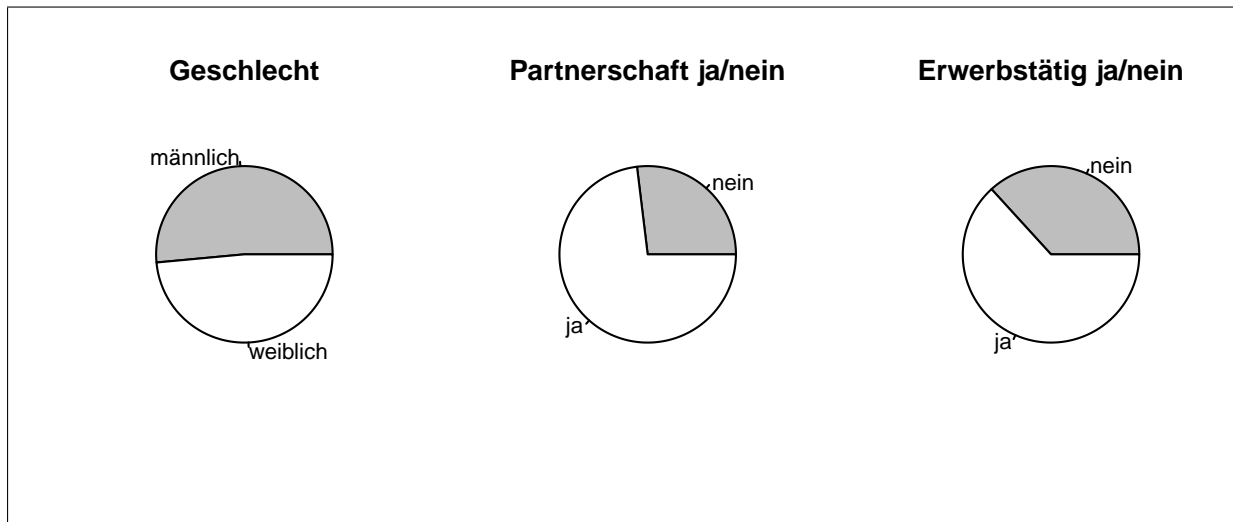


Abbildung 5: Kreisdiagramme zu den Variablen *sex*, *partner* und *employment* des **Allbus-**Datenbeispiels

Abbildung 6 zeigt die Verteilung des Alters. Die jüngste Person ist 18 Jahre und die älteste 96 Jahre alt. Die Linie, welche den Mittelwert darstellt, zeigt, dass die befragten Personen im Mittel etwa 50 Jahre alt sind. Zur besseren Interpretation wurde die Variable *age* um den Wert 50 zentriert. Anschließend wurde sie quadriert, sodass die Variable *age2* entstand.

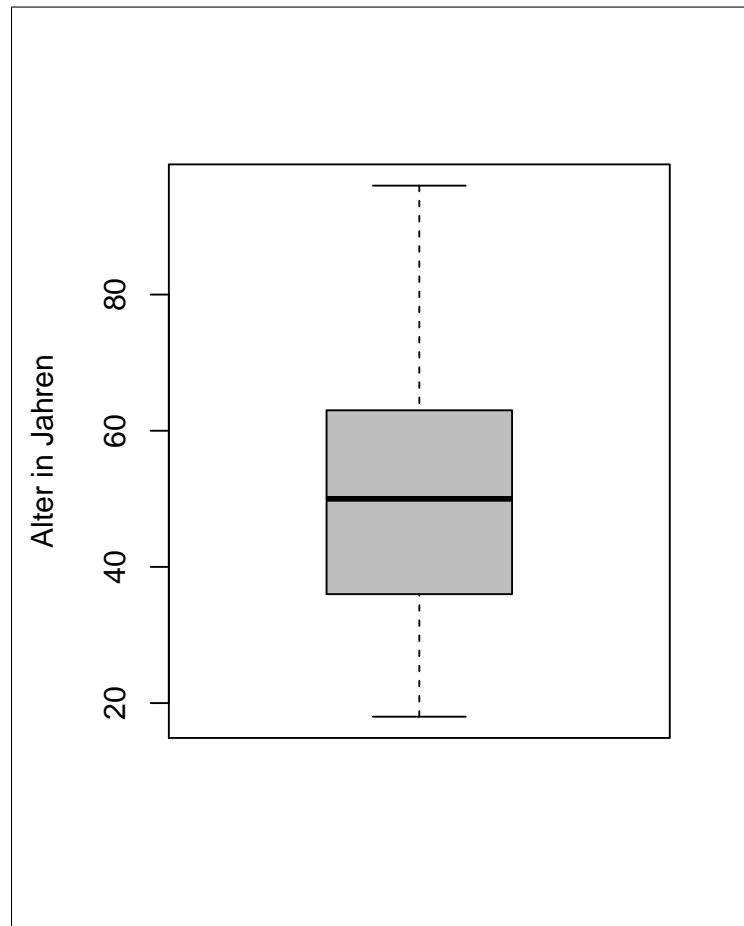


Abbildung 6: Boxplot zur Variable *age* des **Allbus**-Datenbeispiels

Die Verteilung des Einkommens pro Monat in Euro ist in Abbildung 7 zu sehen. Im Mittel wurde ein Einkommen von 1517 Euro angegeben. Mehrere Personen gaben ein höheres Einkommen als 4000 Euro an. Diese sind als Ausreißer in dem Boxplot markiert. Da der Median mit 1300 Euro geringer als der Mittelwert ist, handelt es sich um eine rechtsschiefe Verteilung. 10000 Euro pro Monat ist das höchste Einkommen, das angegeben wurde. Für eine sinnvollere Interpretation der Wirkung des Einkommens wurde die Einflussgröße durch 100 geteilt.

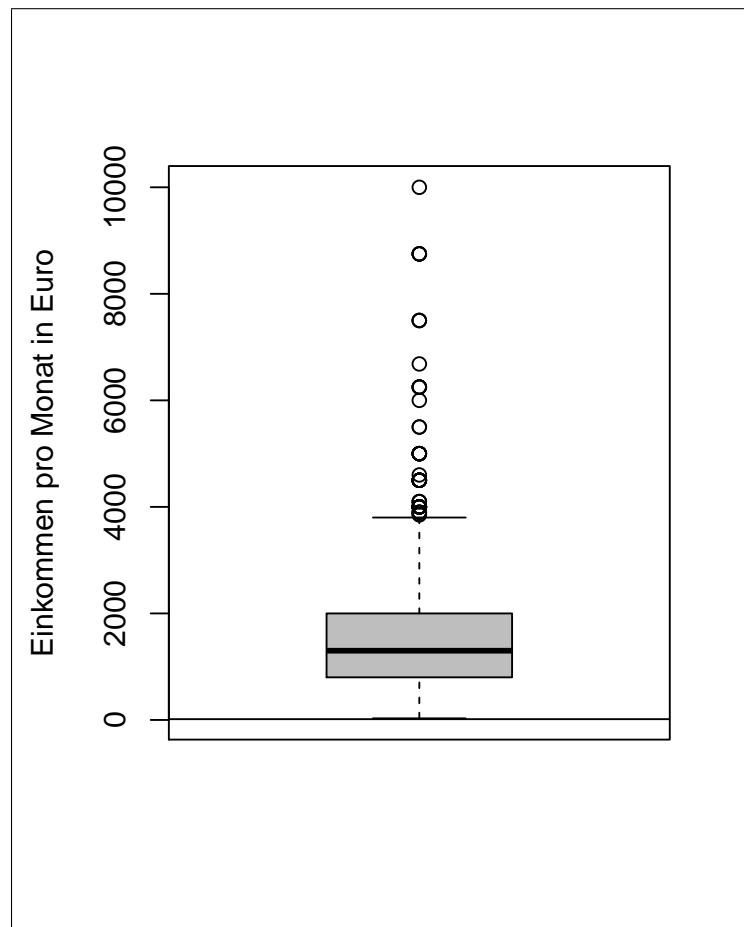


Abbildung 7: Boxplot zur Variable *income* des **Allbus**-Datenbeispiels

Die Variable *life_kat* basiert auf einer Skala von 0 bis 10, auf der die Befragten ihre allgemeine Lebenszufriedenheit einschätzen sollten. Diese wird kategorial in das Modell aufgenommen, eingeteilt in „unzufrieden“ (Werte von 0 bis 3 auf der ursprünglichen Skala), „mittel“ (von 4 bis 6) und „zufrieden“ (von 7 bis 10). Abbildung 8 zeigt, dass die meisten Personen angaben, dass sie zufrieden mit ihrem Leben sind. Etwa 40 Befragte sagten, sie wären unzufrieden.

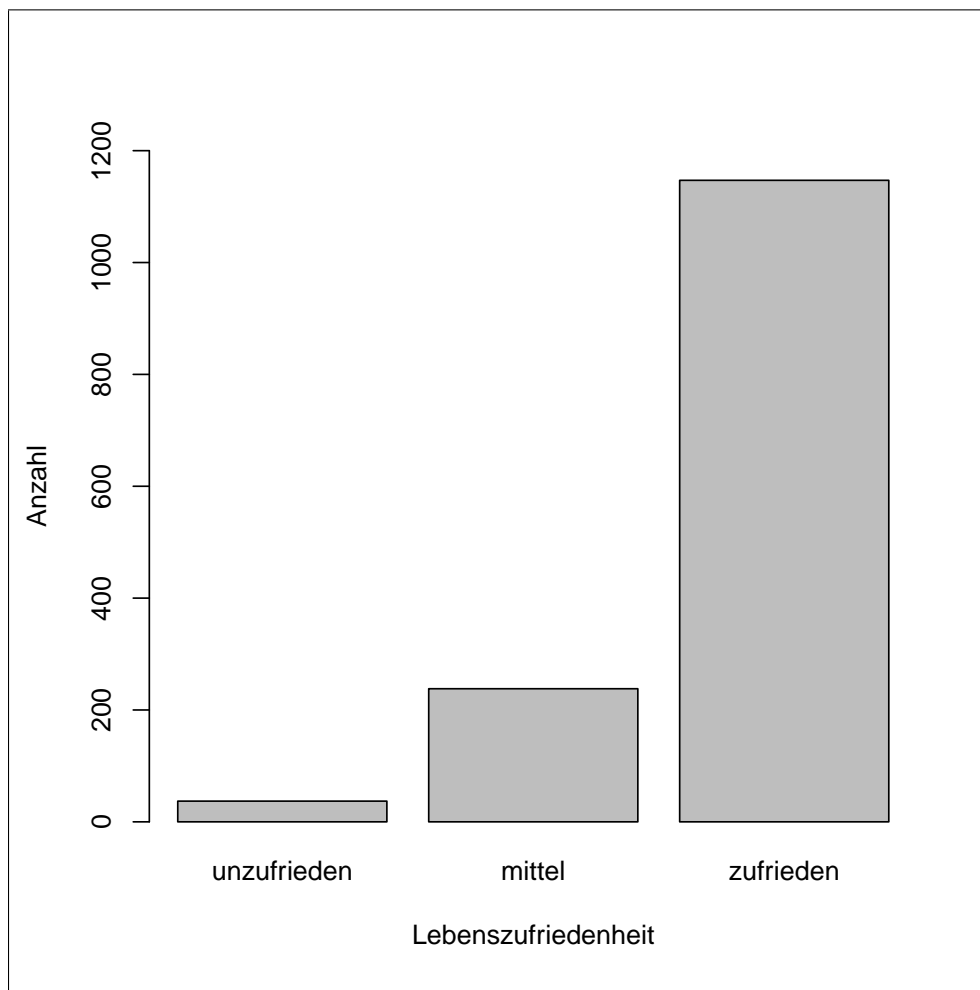


Abbildung 8: Verteilung der Variable *life_kat* des **Allbus**-Datenbeispiels

Abbildung 9 zeigt die Verteilung des politischen Interesses. Die meisten haben ein mittleres Interesse, dies entspricht der Zahl 3. Bei sinkendem (Kategorie 1 oder 2) oder steigendem Interesse (Kategorien 4 und 5) nimmt die Anzahl der Personen ab.

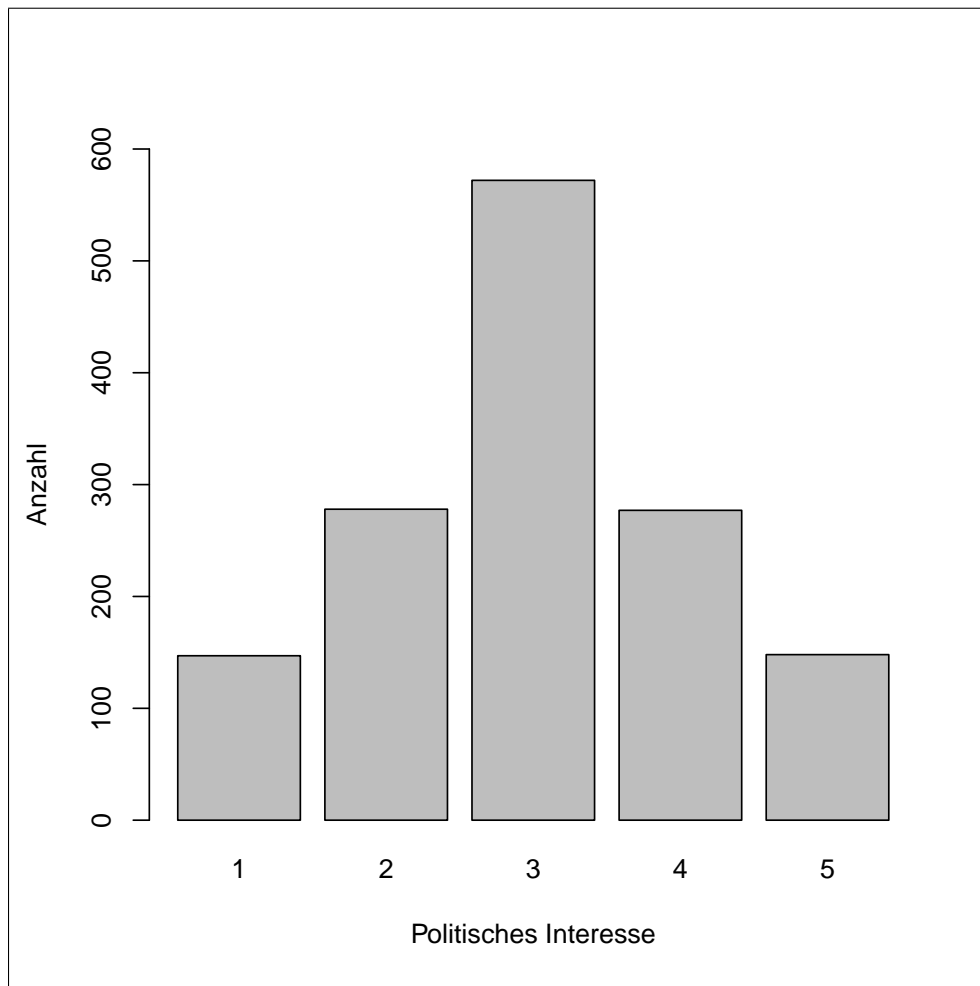


Abbildung 9: Verteilung der Variable *interest.kat* des **Allbus**-Datenbeispiels

Die Größe *leftright_kat* gibt an, zu welcher politischen Ausrichtung der Befragte zu zählen ist. Sie ist eingeteilt in eher „links“, „Mitte“ und eher „rechts“. Ursprünglich konnte der Proband dies auf einer Skala von 1 (links) bis 10 (rechts) angeben. In Abbildung 10 ist zu sehen, dass die meisten Leute der politischen Mitte zugeordnet werden können. Dies entspricht Werten zwischen 4 und 7 auf der ursprünglichen Skala, welche von 1 bis 10 reicht. Etwa 275 Personen gaben an, dass sie eher links orientiert sind. Dies entspricht Werten zwischen 1 und 3 auf der ursprünglichen Skala. Am wenigstens Personen gaben Zahlen zwischen 8 und 10 und damit eine eher rechte politische Präferenz an.

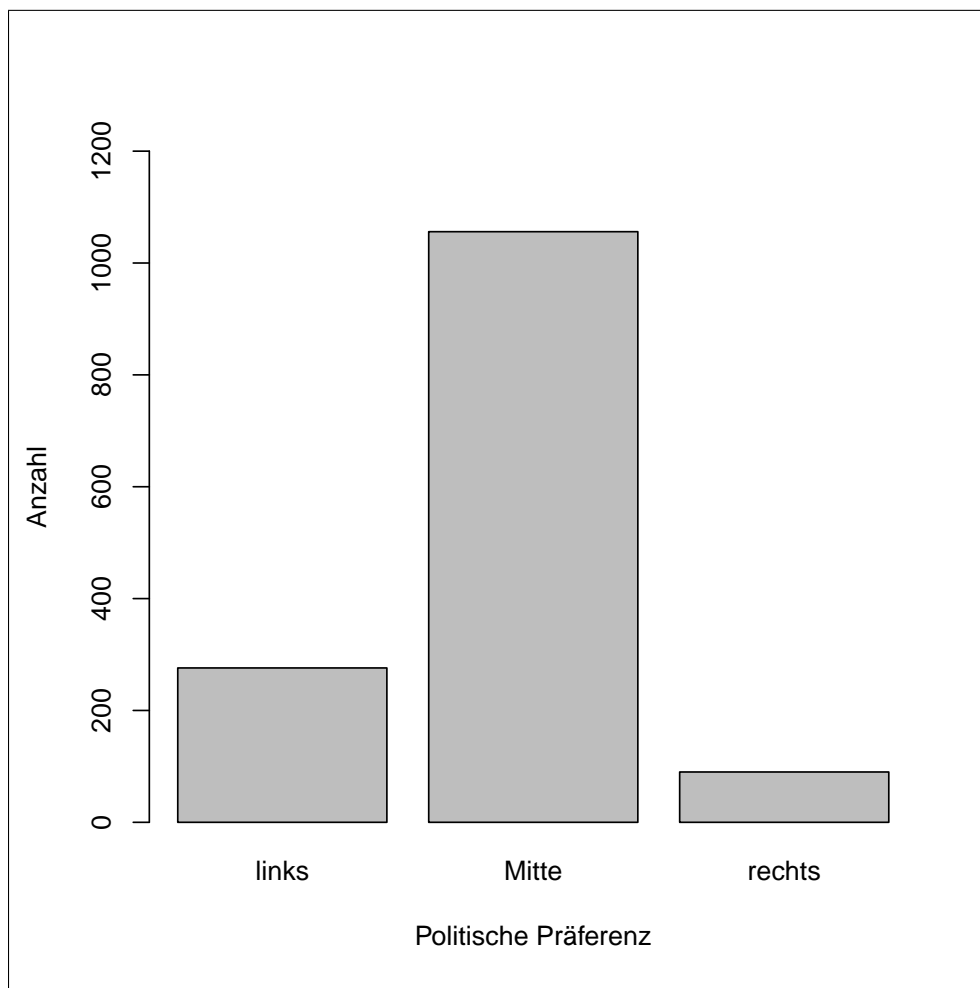


Abbildung 10: Verteilung der Variable *leftright_kat* des *Allbus*-Datenbeispiels

Für die beiden Zielgrößen *government* und *fellow* wird zunächst der gleiche lineare Prädiktor unter Proportional-Odds Annahme verwendet:

$$\begin{aligned}
 \eta_{ir} = & \beta_{0r} + \beta_{\text{sex}1}x_{\text{sex}1,i} + \beta_{\text{partner}1}x_{\text{partner}1,i} + \beta_{\text{employment}1}x_{\text{employment}1,i} + \beta_{\text{age}}x_{\text{age},i} \\
 & + \beta_{\text{age}2}x_{\text{age}2,i} + \beta_{\text{income}}x_{\text{income},i} + \beta_{\text{life_kat}1}x_{\text{life_kat}1,i} + \beta_{\text{life_kat}3}x_{\text{life_kat}3,i} \\
 & + \beta_{\text{interest_kat}1}x_{\text{interest_kat}1,i} + \beta_{\text{interest_kat}2}x_{\text{interest_kat}2,i} + \beta_{\text{interest_kat}4}x_{\text{interest_kat}4,i} \\
 & + \beta_{\text{interest_kat}5}x_{\text{interest_kat}5,i} + \beta_{\text{leftright_kat}1}x_{\text{leftright_kat}1,i} + \beta_{\text{leftright_kat}r}x_{\text{leftright_kat}r,i}
 \end{aligned} \tag{20}$$

mit $r \in \{1, \dots, 6\}$ für die Zielvariable *government*

bzw. $r \in \{1, 2\}$ für die Zielvariable *fellow*

$$x_{\text{sex}1,i} = \begin{cases} 1, & \text{weiblich} \\ 0, & \text{männlich} \end{cases}$$

$$x_{\text{partner}1,i} = \begin{cases} 1, & \text{Partnerschaft} \\ 0, & \text{keine Partnerschaft} \end{cases}$$

$$x_{\text{employment}1,i} = \begin{cases} 1, & \text{erwerbstätig} \\ 0, & \text{nicht erwerbstätig} \end{cases}$$

$$x_{\text{life_kat}1,i} = \begin{cases} 1, & \text{Lebenszufriedenheit = unzufrieden} \\ 0, & \text{sonst} \end{cases}$$

$$x_{\text{life_kat}3,i} = \begin{cases} 1, & \text{Lebenszufriedenheit = zufrieden} \\ 0, & \text{sonst} \end{cases}$$

$$x_{\text{interest_kat}1,i} = \begin{cases} 1, & \text{politisches Interesse = gar nicht} \\ 0, & \text{sonst} \end{cases}$$

$$x_{\text{interest_kat}2,i} = \begin{cases} 1, & \text{politisches Interesse = weniger} \\ 0, & \text{sonst} \end{cases}$$

$$x_{\text{interest_kat}4,i} = \begin{cases} 1, & \text{politisches Interesse = stark} \\ 0, & \text{sonst} \end{cases}$$

$$x_{\text{interest_kat}5,i} = \begin{cases} 1, & \text{politisches Interesse = sehr stark} \\ 0, & \text{sonst} \end{cases}$$

$$x_{\text{leftright_katl},i} = \begin{cases} 1, & \text{politisches Ausrichtung = eher links} \\ 0, & \text{sonst} \end{cases}$$

$$x_{\text{leftright_katr},i} = \begin{cases} 1, & \text{politisches Ausrichtung = eher rechts} \\ 0, & \text{sonst} \end{cases}$$

Der Prädiktor zeigt, dass für das Geschlecht ein Mann, für die Partnerschaft eine Person ohne Partner, für die Erwerbstätigkeit eine Person ohne Arbeit, für die Zufriedenheit eine mittelmäßig zufriedene (*life_kat2*), für das politische Interesse eine mittel interessierte (*interest_kat3*) und für die politische Ausrichtung eine zur Mitte orientierte Person als Referenzgruppen gewählt werden.

4.2.2 Vergleich der Testergebnisse Zielvariable *government*

Zuerst werden die Ergebnisse der Berechnung der Modelle für die Variable *government* erläutert. Um die Proportionalitätsannahme zu testen, wird ein Likelihood-Ratio-Test (siehe Formel (13)) durchgeführt. Dieser gibt Aufschluss darüber, ob das Modell mit Proportional-Odds Annahme dem Modell mit kategoriespezifischen Koeffizienten vorzuziehen ist. Die Likelihoodfunktion bzw. die Schätzung der Parameter kann bei dem kumulativen Modell mit Annahme kategoriespezifischer Koeffizienten für jede Einflussgröße nicht sinnvoll berechnet werden, wie die Warning-Meldungen in Anhang B.1 zeigen. Gründe dafür sind die hohen Anzahlen an Kategorien der Zielvariable *government* und der qualitativen Einflussgrößen, welche zu starken Restriktionen des Prädiktors führen. Für das kumulative Modell kann deshalb nur die Proportional-Odds Annahme für jede Einflussgröße einzeln überprüft werden. Dazu werden die Nullhypothesen $H_0 : \hat{\beta}_{j1} = \dots = \hat{\beta}_{jk}$ gegenüber den Alternativhypothesen $H_1 : \hat{\beta}_{j1} \neq \dots \neq \hat{\beta}_{jk}$ mit $j \in \{sex, partner, employment, age, age2, income, life_kat, interest_kat, leftright_kat\}$ betrachtet. Die Ergebnisse der Tests sind in Tabelle 2 dargestellt. Für die Einflussvariablen *sex* und *interest_kat* wird die Nullhypothese zum Signifikanzniveau 0.05 abgelehnt und kategoriespezifische Parameter angenommen. Dies zeigen auch die Symbole hinter den p-Werten. Ist der p-Wert kleiner also 0.05, so befindet sich ein *,* hinter der Variable. Werte unter 0.01 entsprechen **,** und Werte unter 0.001 ,***. Für das Adjacent-Category Logit-Modell kann die Proportionalitätsannahme für alle Einflussgrößen gleichzeitig anhand eines Likelihood-Ratio-Tests überprüft und aufgrund eines p-Wertes von 0.001 abgelehnt werden. Die p-Werte des Tests der Proportional-Odds Annahme für jede einzelne Einflussvariable befinden sich in Tabelle 3. Die Nullhypothese des Likelihood-Ratio-Tests wird bei der Verwendung benachbarter, lokaler Logits für die Variablen *sex* und *interest_kat* zum Signifikanzniveau 0.05 abgelehnt.

Variablenname	p-Wert
sex	0.026 *
partner	0.417
employment	0.121
age	0.655
age2	0.272
income	0.237
life_kat	0.374
interest_kat	0.002 **
leftright_kat	0.231

Tabelle 2: Testergebnis der Partial-Proportional-Odds Annahme für das kumulative Logit-Modell der Zielvariable *government* des **Allbus**-Datenbeispiels

Variablenname	p-Wert
sex	0.043 *
partner	0.429
employment	0.111
age	0.575
age2	0.356
income	0.283
life_kat	0.269
interest_kat	0.001 ***
leftright_kat	0.122

Tabelle 3: Testergebnis der Partial-Proportional-Odds Annahme für das Adjacent-Type Logit-Modell der Zielvariable *government* des **Allbus**-Datenbeispiels

Bei beiden Modellen wird unter Partial-Proportional-Odds Annahme folgender Prädiktor für die Zielvariable *government* verwendet:

$$\begin{aligned} \eta_{ir} = & \beta_{0r} + \beta_{\text{sex1},r}x_{\text{sex1},i} + \beta_{\text{partner1}}x_{\text{partner1},i} + \beta_{\text{employment1}}x_{\text{employment1},i} + \beta_{\text{age}}x_{\text{age},i} \\ & + \beta_{\text{age2}}x_{\text{age2},i} + \beta_{\text{income}}x_{\text{income},i} + \beta_{\text{life_kat1}}x_{\text{life_kat1},i} + \beta_{\text{life_kat3}}x_{\text{life_kat3},i} \\ & + \beta_{\text{interest_kat1},r}x_{\text{interest_kat1},i} + \beta_{\text{interest_kat2},r}x_{\text{interest_kat2},i} + \beta_{\text{interest_kat4},r}x_{\text{interest_kat4},i} \\ & + \beta_{\text{interest_kat5},r}x_{\text{interest_kat5},i} + \beta_{\text{leftright_kat1}}x_{\text{leftright_kat1},i} + \beta_{\text{leftright_katr}}x_{\text{leftright_katr},i} \end{aligned}$$

mit $r \in \{1, \dots, 6\}$ für die Zielvariable *government*. Die Variablen sind wie beim zuvor verwendeten Prädiktor (20) definiert.

Beim kumulativen Modell ergeben sich nach der Berechnung die Koeffizienten der ersten Spalte von Tabelle 4. Für Adjacent-Category Logits sind diese in Tabelle 5 aufgelistet. Zur besseren Interpretation sind zudem die exponierten Koeffizienten und deren berechnete p-Werte zur Nullhypothese $H_0 : \hat{\beta}_j = 0$ bei Annahme von Proportional-Odds mit $j \in \{\text{sex1}, \text{partner1}, \text{employment1}, \text{age}, \text{age2}, \text{income}, \text{life_kat1}, \text{life_kat3}, \text{interest_kat1}, \text{interest_kat2}, \text{interest_kat4}, \text{interest_kat5}, \text{leftright_kat1}, \text{leftright_katr}\}$ gegeben. Bei categoriespezifischen Parametern wird die Nullhypothese $H_0 : \hat{\beta}_{jr} = 0$ betrachtet. Hinter den p-Werten sind die Indikatoren als Sterne für die Signifikanz der Tests äquivalent zu den Sternen in den Tabellen 2 und 3 gegeben.

Variable	Estimate	exp(Estimate)	p-Wert	
Intercept:1	-1.932	0.145	0.000	***
Intercept:2	-1.079	0.340	0.000	***
Intercept:3	-0.083	0.920	0.689	
Intercept:4	1.063	2.895	0.000	***
Intercept:5	2.518	12.404	0.000	***
Intercept:6	4.253	70.316	0.000	***
sex1:1	-0.577	0.562	0.007	**
sex1:2	-0.375	0.687	0.013	*
sex1:3	-0.223	0.800	0.069	
sex1:4	-0.060	0.942	0.624	
sex1:5	0.150	1.162	0.391	
sex1:6	0.076	1.079	0.816	
partner1	0.100	1.105	0.373	
employment1	-0.010	0.990	0.946	
age	0.006	1.006	0.108	

age2	-0.001	0.999	0.000	***
income	-0.012	0.988	0.021	*
life_kat1	0.099	1.104	0.754	
life_kat3	-0.420	0.657	0.001	***
interest_kat1:1	1.515	4.549	0.000	***
interest_kat1:2	1.575	4.831	0.000	***
interest_kat1:3	1.375	3.955	0.000	***
interest_kat1:4	1.200	3.320	0.000	***
interest_kat1:5	1.011	2.748	0.009	**
interest_kat1:6	-0.202	0.817	0.699	
interest_kat2:1	0.087	1.091	0.779	
interest_kat2:2	0.413	1.511	0.038	*
interest_kat2:3	0.459	1.582	0.003	**
interest_kat2:4	0.587	1.799	0.000	***
interest_kat2:5	0.473	1.605	0.056	
interest_kat2:6	0.227	1.255	0.638	
interest_kat4:1	-0.277	0.758	0.408	
interest_kat4:2	-0.201	0.818	0.376	
interest_kat4:3	-0.012	0.988	0.941	
interest_kat4:4	-0.112	0.894	0.472	
interest_kat4:5	-0.091	0.913	0.675	
interest_kat4:6	-0.058	0.944	0.896	
interest_kat5:1	0.699	2.012	0.024	*
interest_kat5:2	0.695	2.004	0.003	**
interest_kat5:3	0.069	1.071	0.735	
interest_kat5:4	0.061	1.063	0.759	
interest_kat5:5	-0.448	0.639	0.078	
interest_kat5:6	-0.244	0.783	0.649	
leftrigh_katl	0.355	1.426	0.004	**
leftrigh_katr	0.233	1.262	0.236	

Tabelle 4: Koeffizienten, exponierte Koeffizienten und p-Werte des Wald-Tests für das kumulative Logit-Modell für die Zielvariable *government* des **Allbus**-Datenbeispiels

Variable	Estimate	exp(Estimate)	p-Wert	
Intercept:1	-0.112	0.894	0.680	
Intercept:2	0.531	1.701	0.020	*
Intercept:3	0.222	1.249	0.202	
Intercept:4	-0.260	0.771	0.104	
Intercept:5	-1.140	0.320	0.000	***
Intercept:6	-1.587	0.205	0.000	***
sex1:1	0.449	1.567	0.093	
sex1:2	0.142	1.153	0.515	
sex1:3	0.081	1.084	0.630	
sex1:4	-0.021	0.979	0.891	
sex1:5	-0.188	0.829	0.380	
sex1:6	-0.071	0.931	0.851	
partner1	-0.041	0.960	0.360	
employment1	-0.012	0.988	0.833	
age	-0.003	0.997	0.070	
age2	0.000	1.000	0.000	***
income	0.005	1.005	0.016	*
life_kat1	-0.039	0.962	0.753	
life_kat3	0.155	1.168	0.003	**
interest_kat1:1	-0.387	0.679	0.269	
interest_kat1:2	-0.786	0.456	0.013	*
interest_kat1:3	-0.569	0.566	0.048	*
interest_kat1:4	-0.472	0.624	0.134	
interest_kat1:5	-0.902	0.406	0.159	
interest_kat1:6	1.788	5.977	0.023	*
interest_kat2:1	0.409	1.505	0.276	
interest_kat2:2	-0.288	0.750	0.303	
interest_kat2:3	-0.135	0.874	0.522	
interest_kat2:4	-0.446	0.640	0.030	*
interest_kat2:5	-0.159	0.853	0.610	
interest_kat2:6	0.213	1.237	0.699	
interest_kat4:1	0.184	1.202	0.670	
interest_kat4:2	0.219	1.245	0.506	
interest_kat4:3	-0.155	0.856	0.493	

interest_kat4:4	0.134	1.143	0.506	
interest_kat4:5	-0.004	0.996	0.988	
interest_kat4:6	-0.075	0.928	0.879	
interest_kat5:1	-0.137	0.872	0.731	
interest_kat5:2	-1.263	0.283	0.002	**
interest_kat5:3	0.753	2.123	0.035	*
interest_kat5:4	-0.320	0.726	0.245	
interest_kat5:5	0.738	2.092	0.024	*
interest_kat5:6	-0.192	0.825	0.740	
leftright_katl	-0.123	0.884	0.011	*
leftright_katr	-0.094	0.910	0.231	

Tabelle 5: Koeffizienten, exponierte Koeffizienten und p-Werte des Wald-Tests für das Adjacent-Type Logit-Modell für die Zielvariable *government* des **Allbus**-Datenbeispiels

Zunächst werden die Intercepts miteinander verglichen. Da für das kumulative Modell die kumulierten Wahrscheinlichkeiten betrachtet werden, steigt die Wahrscheinlichkeit für höheres r wie die Koeffizienten von *Intercept:r* mit $r = 1, \dots, 6$ zeigen. Die exponierten Koeffizienten von *Intercept:r* stehen dabei für die Chance auf eine Kategorie r oder niedriger im Gegensatz zu einer höheren Kategorie, wenn die stetigen Variablen (*age*, *age2*, *income*) auf die Mittelwerte und die kategorialen Einflussgrößen auf ihre Referenzkategorien gesetzt werden. Diese Chance steigt besonders stark an, wenn statt der Chance von Kategorie 5 oder niedriger zu höheren Kategorien die Chance von Kategorie 6 oder kleiner zu höheren Kategorien betrachtet wird. Dies zeigt der Vergleich der Werte von *Intercept:5* und *Intercept:6*. Beim Adjacent-Category Modell ergibt sich eine andere Interpretation für die Intercepts. Die exponierten *Intercept:r* geben die Chance von Kategorie $r + 1$ im Gegensatz zu Kategorie r an bei Festsetzen der stetigen Variablen auf die Mittelwerte und der diskreten Größen auf ihre Referenzkategorien. Der Faktor $\exp(\text{Intercept:2})$ ist mit 1.701 am größten unter allen Intercepts. Die Chance auf Kategorie 3 im Gegensatz zu Kategorie 2 beträgt also 1.701, wenn ein Mann im Alter von 50 Jahren ohne Partnerin und Erwerbstätigkeit, einem Einkommen von 1517 Euro und mittlerer Lebenszufriedenheit, politisch der Mitte zuzuordnen ist und mittleres politisches Interesse hat. Werden die Intercepts von $r = 3$ bis $r = 6$ betrachtet, so sinkt die Chance auf $r + 1$ zu der benachbarten Kategorie r .

Nun folgt die Betrachtung der Einflüsse der Kovariablen auf die Zielvariable *government*. Die Tabellen zeigen, dass die Koeffizienten des Adjacent-Category Modell umgekehrte Vorzeichen

zum kumulativen Modell haben, wenn nach dem Wald-Test für die betreffenden Variablen ein signifikanter Einfluss auf die Vertrauensskala angenommen wird. Dies liegt an der unterschiedlichen Form der verwendeten Logits, welche zu anderen Interpretationsrichtungen führen (siehe Formeln (18) und (19)).

Zunächst werden die Einflüsse der Variablen mit Proportional-Odds Annahme betrachtet. Für das kumulative Modell wird angenommen, dass sich die Chancen auf Kategorie r oder niedriger zu einer höheren Kategorie für eine Person mit Partner im Gegensatz zu einer Person ohne Partner und für einen Befragten mit Job zu einem Befragten ohne Job nicht ändern. Das gleiche Ergebnis liefert auch das Adjacent-Category Modell. Diese und alle weiteren Interpretation gelten nur dann, wenn die anderen Einflussgrößen festgehalten werden.

Zudem ändert sich die Chance nicht, wenn statt einer Person, die sich zur politischen Mitte zählt, eine Person betrachtet wird, die eher rechts orientiert ist. Ist die Person jedoch eher links, statt der politischen Mitte zuzuordnen, so kann für beide Modelle eine Änderung der Chance angenommen werden. Dies zeigen die Signifikanzen des Wald-Tests. Werden kumulierte Logits verwendet, steigt die Chance auf Vertrauenskategorien r oder niedriger in die Regierung im Gegensatz zu höheren Vertrauenskategorien um den Faktor $\exp(0.355) = 1.426$, wenn statt einer zur Mitte zählenden Person eine eher links orientierte Person betrachtet wird. Äquivalent dazu ergibt sich für das Adjacent-Category Modell, dass die Chance auf eine nächsthöhere Vertrauenskategorie um den Faktor $\exp(-0.123) = 0.884$ sinkt, wenn eine Person eher links als zur politischen Mitte zu zählen ist.

Für die Wirkung des Einkommens wird für beide Verfahren ein von 0 verschiedener Einfluss angenommen. Je höher das Einkommen ist, desto geringer ist die Chance auf geringe Vertrauenskategorien in die Regierung bei der Verwendung kumulativer Logits. Äquivalent steigt für das Adjacent-Category Modell die Chance auf eine nächsthöhere Vertrauenskategorien um den Faktor $\exp(0.005) = 1.005$, wenn das Einkommen um 100 Euro steigt.

Den Einfluss des Alters stellt Abbildung 11 dar. Auf der y-Achse ist der Wert des linearen Prädiktors $\eta_i = \beta_{age}x_{age,i} + \beta_{age2}x_{age2,i}$ nach dem Alter der Befragten abgebildet. Dieser liegt für eine Person von 50 Jahren bei 0, da die Variable *age* um diesen Wert zentriert wurde. Steigt das Alter bis 50 Jahre an, so nimmt der negative Beitrag zum Prädiktor bei Verwendung von kumulierten Logits ab. Ist der Befragte älter, so nimmt der negative Einfluss wieder zu. Beim Adjacent-Type Modell ist der Kurvenverlauf umgekehrt, da andere Logits verwendet werden. Der Beitrag sinkt vom Positiven auf 0 bei 50 Jahren und steigt dann wieder an.

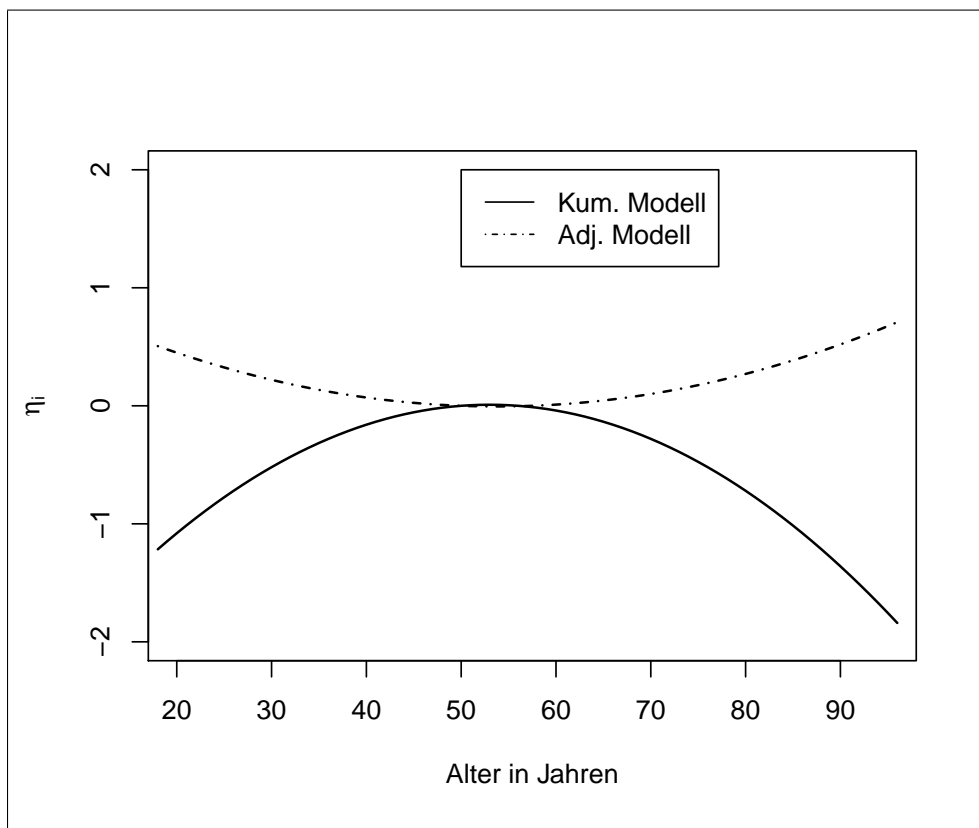


Abbildung 11: Werte des linearen Prädiktors $\eta_i = \beta_{age}x_{age,i} + \beta_{age2}x_{age2,i}$ für die Zielvariable *government* nach dem Alter in Jahren des kumulativen Modells (durchgezogene Linie) und Adjacent-Category Modells (gestrichelte Linie)

Für die Variable *life_kat* wird bei beiden Modellen angenommen, dass für Personen die zufrieden mit ihrem Leben sind im Gegensatz zu mittelmäßig zufriedenen Personen die Chance auf geringes Vertrauen sinkt (kumulatives Modell) bzw. die Chance auf eine nächsthöhere Kategorie steigt (Adjacent-Type Modell).

Für die kategoriespezifischen Koeffizienten von *sex* ergeben sich ähnliche Trends. Je höher die Vertrauensstufe r ist, desto höher ist die Chance auf niedrigere Kategorien als r , wenn eine Frau statt einem Mann betrachtet wird (kumulatives Modell). Beziehungswise sinkt die Chance auf eine höhere Kategorie $r + 1$ für eine Frau im Gegensatz zu einem Mann bei höheren Vertrauenskategorien von $r = 1, \dots, 5$ (Adjacent-Type Modell). Für die Koeffizienten von *interest_kat* konnten keine ähnlichen Trends festgestellt werden. Die Chancen schwanken bei Änderung der Betrachtung von $r = 1$ zu $r = 6$ beim Adjacent-Category Modell stark, während beim kumulativen Modell, die Chancen nur geringfügig steigende oder fallende Trends aufweisen, wie die ersten Spalten der Ergebnistabellen zeigen. Anhand des Wald-Tests kann in vielen Fällen nicht

angenommen werden, dass sich die Koeffizienten von 0 unterscheiden.

Zum Vergleich des Modellfits eignet sich das AIC. Beim kumulativen Modell beträgt es 4917.836 und beim Adjacent-Category Modell 4922.468. Das kumulative Modell fittet die Vertrauensskala also besser als das Adjacent-Type Modell. Da beiden der gleiche lineare Prädiktor mit identischer Anzahl an Koeffizienten zugrunde liegt, kann auch die Devianz als Vergleichskriterium verwendet werden. Dieses beträgt beim kumulativen Modell 4827.836 und ist damit kleiner als beim Adjacent-Type Modell mit 4832.468.

4.2.3 Vergleich der Testergebnisse Zielvariable *fellow*

Nun folgt die Betrachtung der Koeffizienten für die Zielgröße *fellow*. Sowohl für das Adjacent-Type Modell als auch für das kumulative Modell kann ein Likelihood-Ratio-Test (siehe Formel (13)) für die gleichzeitige Überprüfung der Proportionalitätsannahme aller Kovariablen berechnet werden. Der Test lehnt die Annahme für beide Modelle ab. Für den anschließenden Test der Proportional-Odds Annahme jeder einzelnen Einflussgröße ergeben sich bei Verwendung kumulativer Logits die p-Werte aus Tabelle 6 und für das Adjacent-Type Modell die Werte aus Tabelle 7. Für beide Modelle kann die Proportional-Odds Annahme für den Einfluss der Variablen *age* und *sex* abgelehnt werden, da die p-Werte kleiner als 0.05 sind.

Variablenname	p-Wert
sex	0.011 *
partner	0.553
employment	0.202
age	0.000 ***
age2	0.243
income	0.950
life_kat	0.985
interest_kat	0.159
leftright_kat	0.758

Tabelle 6: Testergebnis des Likelihood-Ratio-Tests zur Partial-Proportional-Odds Annahme für das kumulative Logit-Modell der Zielvariable *fellow* des **Allbus**-Datenbeispiels

Variablenname	p-Wert
sex	0.022 *
partner	0.704
employment	0.071
age	0.000 ***
age2	0.180
income	0.600
life_kat	0.873
interest_kat	0.267
leftright_kat	0.940

Tabelle 7: Testergebnis des Likelihood-Ratio-Tests zur Partial-Proportional-Odds Annahme für das Adjacent-Type Logit-Modell der Zielvariable *fellow* des Allbus-Datenbeispiels

Somit wird für beide Modelle folgender Prädiktor angenommen:

$$\begin{aligned} \eta_{ir} = & \beta_{0r} + \beta_{\text{sex}1,r}x_{\text{sex}1,i} + \beta_{\text{partner}1}x_{\text{partner}1,i} + \beta_{\text{employment}1}x_{\text{employment}1,i} + \beta_{\text{age},r}x_{\text{age},i} \\ & + \beta_{\text{age}2}x_{\text{age}2,i} + \beta_{\text{income}}x_{\text{income},i} + \beta_{\text{life_kat}1}x_{\text{life_kat}1,i} + \beta_{\text{life_kat}3}x_{\text{life_kat}3,i} \\ & + \beta_{\text{interest_kat}1}x_{\text{interest_kat}1,i} + \beta_{\text{interest_kat}2}x_{\text{interest_kat}2,i} + \beta_{\text{interest_kat}4}x_{\text{interest_kat}4,i} \\ & + \beta_{\text{interest_kat}5}x_{\text{interest_kat}5,i} + \beta_{\text{leftright_kat}1}x_{\text{leftright_kat}1,i} + \beta_{\text{leftright_kat}r}x_{\text{leftright_kat}r,i} \end{aligned}$$

mit $r \in \{1,2\}$ für die Zielvariable *fellow*.

Die berechneten Koeffizienten, deren exponierte Werte, zugehörige p-Werte des Wald-Tests und die Indikatoren der Signifikanz der beiden Modelle befinden sich in den Tabellen 8 und 9.

Variable	Estimate	exp(Estimate)	p-Wert
Intercept:1	0.577	1.781	0.009 **
Intercept:2	2.394	10.957	0.000 ***
sex1:1	-0.151	0.860	0.210
sex1:2	0.208	1.231	0.175
partner1	-0.003	0.997	0.983
employment1	-0.053	0.948	0.730
age:1	0.021	1.021	0.000 ***
age:2	0.001	1.001	0.819
age2	0.000	1.000	0.407
income	-0.036	0.965	0.000 ***

life_kat1	0.084	1.088	0.815	
life_kat3	-0.345	0.708	0.018	*
interest_kat1	0.610	1.840	0.001	***
interest_kat2	0.343	1.409	0.017	*
interest_kat4	-0.219	0.803	0.130	
interest_kat5	-0.663	0.515	0.000	***
leftright_katl	-0.288	0.750	0.029	*
leftright_katr	-0.228	0.796	0.285	

Tabelle 8: Koeffizienten, exponierte Koeffizienten und p-Werte des Wald-Tests für das kumulative Modell mit Partial-Proportional-Odds Annahme für die Zielvariable *follow* des **Allbus**-Datenbeispiels

Variable	Estimate	exp(Estimate)	p-Wert	
Intercept:1	-0.774	0.461	0.000	***
Intercept:2	-1.438	0.237	0.000	***
sex1:1	0.209	1.232	0.093	
sex1:2	-0.279	0.757	0.087	
partner1	-0.008	0.992	0.930	
employment1	0.097	1.102	0.408	
age:1	-0.022	0.978	0.000	***
age:2	0.011	1.011	0.027	*
age2	-0.000	1.000	0.681	
income	0.026	1.026	0.000	***
life_kat1	-0.049	0.952	0.859	
life_kat3	0.252	1.287	0.023	*
interest_kat1	-0.477	0.621	0.001	***
interest_kat2	-0.274	0.760	0.012	*
interest_kat4	0.135	1.145	0.199	
interest_kat5	0.451	1.570	0.001	***
leftright_katl	0.209	1.232	0.030	*
leftright_katr	0.144	1.155	0.361	

Tabelle 9: Koeffizienten, exponierte Koeffizienten und p-Werte des Wald-Tests für das Adjacent-Type Modell mit Partial-Proportional-Odds Annahme für die Zielvariable *follow* des **Allbus**-Datenbeispiels

Da beim kumulativen Modell kumulierte Logits verwendet werden, ist der Wert von *Intercept:2* im Vergleich zu *Intercept:1* höher. Beim Adjacent-Type Modell liegt die Chance auf Kategorie 2 statt Kategorie 1 bei 0.461, wenn alle stetigen Variablen auf ihre Mittelwerte und alle diskreten Variablen auf ihre Referenz gesetzt werden. Die Chance ist damit höher als die Chance von Stufe 3 zu Stufe 2, wie der Vergleich der Werte von *Intercept:1* zu *Intercept: 2* belegt.

Wird die Wirkung der Kovariablen verglichen, so zeigen sich, aufgrund der Verwendung unterschiedlicher Logits, umgekehrte Vorzeichen der Koeffizienten, welche signifikant von 0 verschieden angenommen werden (siehe Formeln (18) und (19)). Für beide Modelle gilt, dass die Einflüsse von *income*, *life_kat3* im Gegensatz zur Referenzgruppe *life_kat2*, *interest_kat1*, *interest_kat2* und *interest_kat5* zur Gruppe *interest_kat3* sowie *leftright_kat1* zur Referenz politische „Mitte“ als signifikant von 0 verschieden angenommen werden können. Jedoch unterscheiden sich die Interpretationsformen. Der Koeffizientenwert der Variable *income* zeigt zum Beispiel beim kumulativen Modell, dass die Chance auf Vertrauenskategorie *r* oder niedriger zu höheren Vertrauensstufen um den Faktor 0.965 sinkt, wenn sich das Einkommen pro Monat um 100 Euro erhöht. Der Koeffizient beim Adjacent-Category Logit-Modell besagt, dass die Chance wegen des positiven Vorzeichens auf Kategorie *r+1* zu Ausprägung *r* um den Faktor $\exp(0.026) = 1.026$ steigt. Beide Interpretationen gelten nur unter Festhalten der anderen Einflussgrößen und besagen, dass die Chance auf mehr Vertrauen in die Mitmenschen steigt je mehr die Person verdient.

Den Einfluss des Alters stellt Abbildung 12 dar. Auf der y-Achse ist der lineare Prädiktor $\eta_i = \beta_{age,r}x_{age,i} + \beta_{age2}x_{age2,i}$ nach dem Alter der Befragten abgebildet. Da für das Alter kategoriespezifische Koeffizienten angenommen werden, wird nach den Kategorien unterschieden. Alle vier Graphen schneiden sich bei 50 Jahren im Punkt 0, da die Variable *age* um 50 Jahre zentriert wurde. Bei der Betrachtung von Kategorie 1 zeigen sich deutlich die unterschiedlichen Interpretationsrichtungen der beiden Modelle. Beim Adjacent-Type Modell ist der Einfluss zunächst positiv und ab 50 sinkt er immer weiter in einen negativen Bereich wie die gestrichelte schwarze Linie zeigt. Beim kumulativen Modell steigt der Beitrag vom negativen kontinuierlich in einen immer höheren positiven Bereich. Bei Kategorie 2 ist der Beitrag zum Prädiktor, wie die beiden grauen Linien darstellen, beim kumulativen Modell stets größer gleich 0, während er beim Adjacent-Category Modell zunächst bis 50 Jahre negativ ist. Dann kehrt er sich in das Positive um und wächst mit geringer werdender Steigung weiter an.

Nun folgt die Betrachtung der Wirkung kategorialer Variablen. Die Chance auf die nächsthöhere Vertrauensstufe steigt, wenn die Person sehr starkes (Kat. 5) statt mittleres (Kat. 3) politisches Interesse hat.

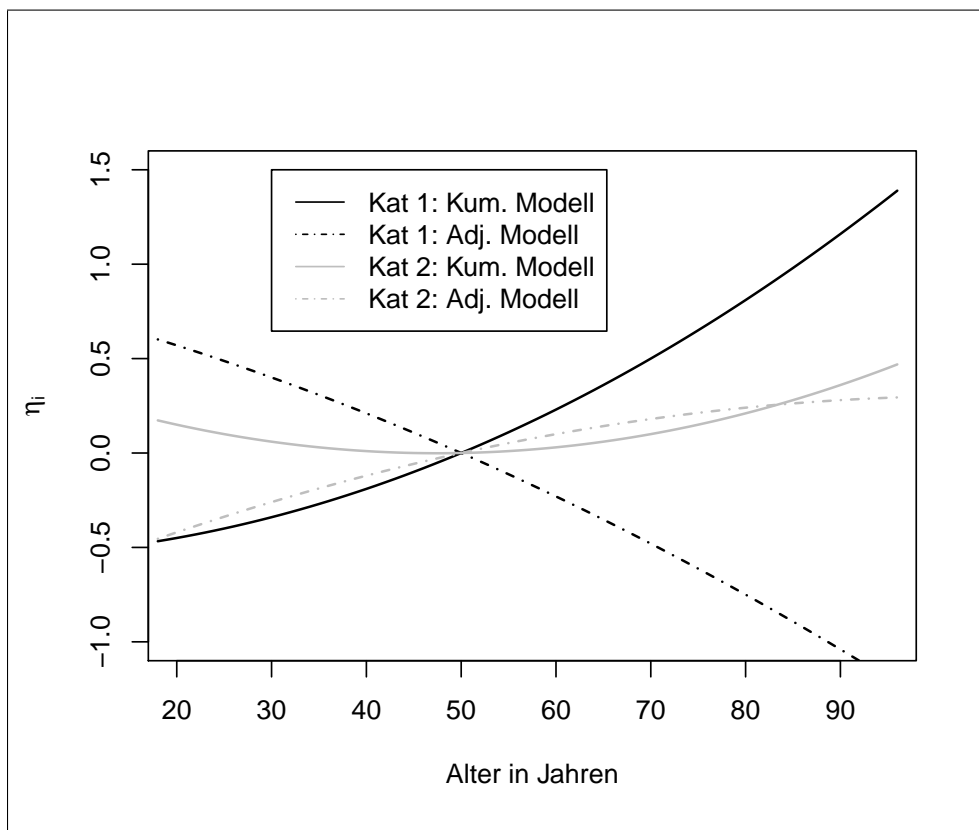


Abbildung 12: Werte des linearen Prädiktors $\eta_i = \beta_{age,r}x_{age,i} + \beta_{age2}x_{age2,i}$ für die Zielvariable *fellow* nach dem Alter in Jahren des kumulativen Modells (durchgezogene Linie) und Adjacent-Category Modells (gestrichelte Linie)

Diese Chance sinkt jedoch, wenn die Person weniger (Kat. 2) oder gar kein (Kat. 1) politisches Interesse hat im Gegensatz zu einer Person mit mittlerem Interesse. Dies zeigen die Koeffizienten des Adjacent-Type Modells. Äquivalent gilt für das kumulative Modell, dass die Chance auf Vertrauenskategorien niedriger gleich r im Gegensatz zu höheren Kategorien sinkt, wenn die Person statt mittleres Interesse sehr starkes politisches Interesse hat. Es steigt, wenn die Person weniger oder gar kein politisches Interesse aufweist. Für jemanden mit starkem Interesse (Kat. 4) wird angenommen, dass sich dessen Chance auf eine Kategorie niedriger gleich r zu einer mittel interessierten Person nicht verändert.

Ist der Befragte eher links statt in der Mitte politisch orientiert oder ist er zufrieden mit seinem Leben statt nur mittelmäßig zufrieden, so hat er eine höhere Chance auf eine nächsthöhere Kategorie, bzw. seine Chance sinkt auf niedrigere Kategorien niedriger gleich r . Dies zeigen sowohl die Koeffizientenwerte *leftrigh katl* und *life kat3* des Adjacent-Type als auch des kumulativen Modells. Für eher rechts orientierte Personen wird angenommen, dass sich die Chance auf

Kategorien niedriger gleich r oder auf die nächsthöhere Kategorie r zu einer in der Mitte zuzuordnenden Person nicht verändert. Für den Beziehungsstatus und den Erwerbstätigkeitsstatus wird bei beiden Modellen angenommen, dass sie keine Wirkung auf die Wahl der Vertrauens-kategorie haben.

Zum Vergleich der Modellgüte wird wieder das AIC verwendet. Wie beim Modell mit Zielgröße *government* ist das AIC des kumulativen Modells mit 2796.782 kleiner als das AIC des Adjacent-Category Modells mit 2798.103. Mit 2760.782 ist auch die Devianz beim kumulativen Modell kleiner als beim Adjacent-Category Modell mit 2762.103.

4.3 Zusammenfassung des Vergleichs

Anhand des Datenbeispiels hat sich gezeigt, dass die beiden Verfahren ähnliche Ergebnisse liefern, wobei die jeweilige Interpretationsrichtung beachtet werden sollte. Vergleichte man die Koeffizienten, so waren die Vorzeichen umgedreht, da die Interpretationen eine andere Richtung aufwiesen. Beide Verfahren führten beim Test der Partial-Proportional-Odds Annahme zu gleichen Testergebnissen, wodurch gleiche Prädiktoren verwendet wurden. Ein großer Vorteil des Adjacent-Type Modells ist, dass Modelle mit kategoriespezifischen Koeffizienten berechnet werden können, auch wenn die Zielvariable bereits aus vielen Kategorien besteht und mehrere kategoriale Einflussgrößen betrachtet werden. Dies war nicht der Fall bei der Modellierung des Vertrauens in die Regierung unter Verwendung kumulativer Logits. Aufgrund der hohen Anzahl der Restriktionen konnte kein Likelihood-Ratio-Test durchgeführt werden. Dies galt auch für die Modellierung des Vertrauens in die Justiz, Zielvariable *justice*, des **Allbus**-Datensatzes. Die Ergebnistabellen zu den Partial-Proportional-Odds Tests sowie die geschätzten Werte der Koeffizienten befinden sich in Anhang B.2. Zur Modellierung konnten kategorieunspezifische Koeffizienten für alle Einflussgrößen angenommen werden.

Bei den stetigen Variablen der Modelle zu den Zielvariablen *government* und *fellow* zeigten die p-Werte zum Wald-Test ähnliche Signifikanzen, während sich bei den kategorialen Größen häufigere Abweichungen abzeichneten. Diese Abweichungen wurden größer, wenn kategorie-spezifische kategoriale Einflüsse betrachtet wurden. Beim Vergleich des Modellfits zeigte sich, dass das kumulative Modell ein niedrigeres AIC und eine niedrigere Devianz als das Adjacent-Category Modell hat. Die Differenzen der Vergleichsgrößen waren bei den beiden Zielgrößen aber nicht größer als der Wert 5. Es ist nicht die Regel, dass das AIC und die Devianz des kumulativen Modells kleiner als beim Adjacent-Type Modell sind. Bei der Modellierung der Zielvariable *justice*, dessen Ergebnisse im Anhang B.2 zu finden sind, sind das AIC und die Devianz des Adjacent-Type Modells geringer als beim kumulativen Modell.

Im Allgemeinen gilt, dass sowohl das kumulative Logit-Modell als auch das Adjacent-Category Logit-Modell in ähnlichen Situation einen vergleichbaren Fit aufweisen. Dies bestätigt auch Agresti (2010) S. 95 f.. Wie kann trotzdem zwischen den beiden Verfahren entschieden werden? Wird die Betrachtung des Effekts auf individuelle Responsekategorien präferiert, so eignet sich das Adjacent-Type Modell. Wird aber die Betrachtung gruppierter Kategorien über die gesamte Skala der zugrundeliegenden latenten Variable bevorzugt, so sollten kumulierte Logits und damit das kumulative Modell verwendet werden.

5 Longitudinale Daten

Das Multinomialmodell für ordinale Responses kann auf ordinale gemischte Modelle erweitert werden, welche neben den festen Effekten auch zufällig Effekte in ihren linearen Prädiktoren beinhalten. Diese Modellform sollte für longitudinale Daten verwendet werden. Zu jedem Subjekt $i \in 1, \dots, n$ liegen T_i Messwiederholungen vor, welche als Repeated Measurements bezeichnet werden. Im Datensatz über Knieverletzungen, der in Sektion 5.2 erläutert wird, sollten zum Beispiel 127 Personen an vier verschiedenen Messzeitpunkten jeweils die Empfindung der Schmerzen ihres Knies auf einer Skala von 1 bis 5 bewerten. Da jede Person ein eigenes Schmerzempfinden hat und die vier erfassten Messpunkte pro Patient stark korrelieren, werden subjektspezifische Parameter benötigt. Wie die Therapieform, das Alter und das Geschlecht auf die kategoriale Zielvariable Schmerzempfinden einwirken, wird im Abschnitt 5.2 vorgestellt. Als Grundlage dazu werden zunächst die ordinalen gemischten Modelle erläutert.

5.1 Ordinale gemischte Modelle

Die folgenden Spezifikationen basieren auf Tutz (2012) S. 405-408. Zunächst wird ein zusätzlicher Index t für den Messzeitpunkt eingeführt. Die Responsevariable für eine Person i zum Messpunkt t wird dann als Y_{it} definiert mit $i = 1, \dots, n$ und $t = 1, \dots, T_i$. Die in dem Modell betrachteten Kovariablen für die Person i zum Zeitpunkt t werden im Vektor \mathbf{x}_{it} zusammengefasst. Variablen mit einem subjektspezifischen, zufälligen Effekt \mathbf{b}_i werden im Vektor \mathbf{z}_{it} gebündelt. Für den Erwartungswert $\mu_{it} = E(y_{it} | \mathbf{x}_{it}, \mathbf{z}_{it}, \mathbf{b}_i)$ gilt dann:

$$u_{it} = h(\mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{b}_i)$$

Die äquivalente Form mit der Linkfunktion $g = h^{-1}$ dazu ist:

$$g(u_{it}) = \mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{b}_i$$

Der Parametervektor $\boldsymbol{\beta}$ beinhaltet alle festen Effektparameter des Modells und \mathbf{b}_i alle zufälligen Effekte, wobei für den Erwartungswert von \mathbf{b}_i $E(\mathbf{b}_i) = 0$ und für die Kovarianz $cov(\mathbf{b}_i) = \mathbf{Q}$ gelten. \mathbf{Q} ist eine unbekannte Kovarianzmatrix.

Werden alle Messzeitpunkte $t \in 1, \dots, T_i$ in den Vektoren $\mathbf{y}_i^T = (y_{i1}, \dots, y_{iT_i})$, $\mathbf{x}_i^T = (1, \mathbf{x}_{i1}, \dots, \mathbf{x}_{iT_i})$, $\mathbf{z}_i^T = (\mathbf{z}_{i1}, \dots, \mathbf{z}_{iT_i})$ zusammengefasst, so ergibt sich das Modell:

$$u_i = h(\mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{b}_i) \quad \text{oder} \quad g(u_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{b}_i$$

Gilt $z_{it} = 1$, so erhält man ein Modell mit Random Intercept:

$$g(\mu_i) = b_i + \mathbf{x}_i^T \boldsymbol{\beta} \quad (21)$$

Durch den Random Intercept variiert nur das Response Level über das betrachtete Subjekt i und nicht die Kovariablen.

Da für die Beobachtungen $y_{11}, y_{12}, \dots, y_{it}, \dots, y_{nT_i}$ und $\mathbf{b}_1, \dots, \mathbf{b}_n$ gilt, dass sie unabhängig sind, können diese in den Vektoren $\mathbf{y}^T = (\mathbf{y}_1^T, \dots, \mathbf{y}_n^T)$ und $\mathbf{b}^T = (\mathbf{b}_1^T, \dots, \mathbf{b}_n^T)$ zusammengefasst werden. Somit ergibt sich das Modell für den Erwartungswertvektor $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$ in der Matrixschreibweise

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}$$

mit $\mathbf{X}^T = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ und $\mathbf{Z}^T = (\mathbf{Z}_1, \dots, \mathbf{Z}_n)$.

Die Schätzung der Koeffizienten \mathbf{b} und $\boldsymbol{\beta}$ erfolgt zum Beispiel durch das Maximieren der Log-Likelihood-Funktion

$$\begin{aligned} l(\boldsymbol{\beta}, \mathbf{Q}) &= \sum_{i=1}^n l_i(\boldsymbol{\beta}, \mathbf{Q}) = \sum_{i=1}^n \log \left(\int f(\mathbf{y}_{it} | \mathbf{b}_i, \mathbf{Q}) p(\mathbf{b}_i, \mathbf{Q}) d\mathbf{b}_i \right) \\ &= \sum_{i=1}^n \log \left(\int \prod_{t=1}^{T_i} f(\mathbf{y}_i | \mathbf{b}_i, \mathbf{Q}) p(\mathbf{b}_i, \mathbf{Q}) d\mathbf{b}_i \right) \end{aligned}$$

Für die Maximierung kann das Gauss-Hermite Quadrature-Verfahren verwendet werden. Dieses wird ebenfalls zur Schätzung der Koeffizienten für den **Knee**-Datensatz im nächsten Abschnitt verwendet. Dazu werden statt den ursprünglichen T_i Beobachtungen für Subjekt i mT_i Beobachtungen verwendet, wobei m der Anzahl der vorher spezifizierten Quadrature Points entspricht. Nun werden die unbekannt Koeffizienten von $\boldsymbol{\beta}$ und $\mathbf{Q}^{-1/2}$ durch die Quadrature Points d_j ersetzt. Dies führt zu einer gewichteten Score-Funktion in Form eines generalisierten linearen Modells. Durch den Fisher-Algorithmus oder das Newton-Raphson-Verfahren kann nun über d_j auf $\boldsymbol{\beta}$ und $\mathbf{Q}^{-1/2}$ geschlossen werden. Welches Verfahren verwendet wird hängt davon ab, ob die beobachtete oder erwartete Fisher-Matrix für den Algorithmus genutzt wird. Ist die Anzahl der Quadrature Points groß genug, so wird die Approximation der Log-Likelihood sehr genau. Mit zunehmender Zahl kann diese Prozedur jedoch sehr lange dauern. Im Allgemeinen gilt, dass 3 bis 7 Quadrature Points reichen, um eine genaue Approximation zu erhalten (vgl. Christensen (2015): S. 25). Alternativen zum Gauss-Hermite Verfahren sind der EM-Algorithmus oder bei niedrig-dimensionalen b_i die Approximation durch die numerische Monte-Carlo Quadrature Technik.

Dieses Modell wird im Folgenden nach Tutz (2012) S. 416 f. auf mehrkategoriale Responses übertragen. Dazu wird $Y_{it} \in 1, \dots, k$ angenommen, wobei die Kategorien $1, \dots, k$ geordnet sind. Für den linearen Prädiktor des ordinalen gemischten Modells gilt äquivalent zu (21)

$$\eta_{it} = b_i + \beta_{0r} + \mathbf{x}_{it}^T \boldsymbol{\beta}$$

mit b_i als subjektspezifischen Random-Effekt, β_{0r} als kategoriespezifischen Intercept und \mathbf{x}_{it}^T als Designvektor. Der Parametervektor $\boldsymbol{\beta}$ ist für alle betrachteten Personen $i = 1, \dots, n$ identisch. Das einfache ordinale gemischte Modell kumulativen Typs ist dann definiert als

$$P(Y_{it} \leq r | \mathbf{x}_{it}) = F(b_i + \beta_{0r} + \mathbf{x}_{it}^T \boldsymbol{\beta})$$

Die Ordnung der Kategorien einer Zielvariable fließen in die Wahl der Funktion F ein. Bei Verwendung kumulativer Logits ergibt sich das kumulative Logit-Modell mit Random Intercepts:

$$P(Y_{it} \leq r | \mathbf{x}_{it}) = \frac{\exp(b_i + \beta_{0r} + \mathbf{x}_{it}^T \boldsymbol{\beta})}{1 + \exp(b_i + \beta_{0r} + \mathbf{x}_{it}^T \boldsymbol{\beta})}$$

Werden Adjacent-Category Logits verwendet, so wird folgendes Modell erhalten:

$$P(Y_{it} = r | Y_{it} = r \text{ oder } Y_{it} = r + 1, \mathbf{x}_{it}) = F(b_i + \beta_{0r} + \mathbf{x}_{it}^T \boldsymbol{\beta}) = \frac{\exp(b_i + \beta_{0r} + \mathbf{x}_{it}^T \boldsymbol{\beta})}{1 + \exp(b_i + \beta_{0r} + \mathbf{x}_{it}^T \boldsymbol{\beta})}$$

5.2 Methodenvergleich anhand des **Knee**-Datensatzes

Der **Knee**-Datensatz basiert auf einer klinischen Studie über $n = 127$ Patienten mit Verletzungen am Knie. Dieser wurde auch im Buch „Regression for Categorical Data“ von Tutz (siehe Tutz (2012) S. 3) verwendet. Die Personen wurden zufällig in zwei Gruppen unterteilt. Bei der einen Gruppe, bestehend aus 64 Personen, wurde zur Therapie ein Spray verwendet. Die restlichen Probanden bekamen keine Therapie. Am ersten, dritten, siebten und zehnten Tag schätzten die Befragten unter normierter Belastung ihr Schmerzempfinden auf einer Skala von 1 wie „geringe Schmerzen“ bis 5 „sehr starke Schmerzen“ ein. Keine der Personen gab an, dass sich die Schmerzen in dem betrachteten Zeitraum verstärkt haben. Da die Skala des Schmerzempfindens aus geordneten Kategorien besteht, kann die Veränderung der Einschätzung der Schmerzen am ersten Tag zum zehnten Tag durch Berechnung der Differenz dargestellt werden. In Graphik 13 ist der Abstand zwischen den Einschätzungen nach der Therapieform abgebildet.

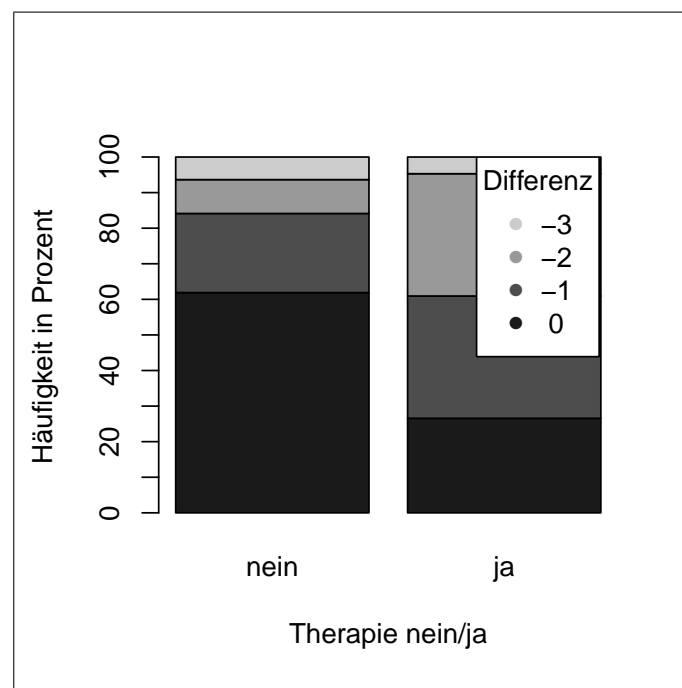


Abbildung 13: Verteilung der Bewertungsabstände nach Therapie ja/nein des **Knee**-Datenbeispiels

Es zeigt sich, dass bei etwa 60 % der Personen ohne Therapiemaßnahmen die Differenz 0 beträgt und somit keine Verbesserung der Schmerzen nach zehn Tagen wahrgenommen wurde. Bei den Personen mit Behandlung durch das Spray gaben nur circa 27% an, dass ihre Schmerzempfinden nach zehn Tagen gleich geblieben ist.

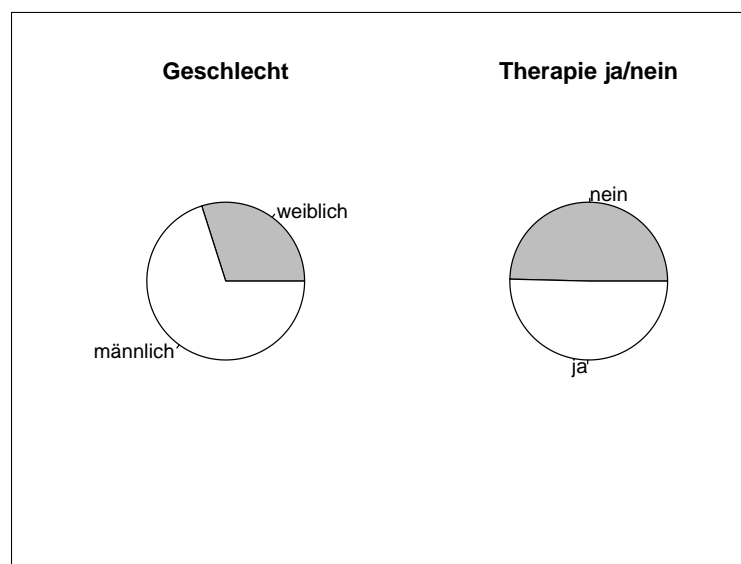
Um die Schmerzkategorie zu modellieren, wurde ein Datensatz erstellt, der in seinen Zeilen die einzelnen Beobachtungen der Personen beinhaltet. Als Indikator, um welche Person es sich handelt, wurde die Variable *Person* verwendet, deren Wertebereich von 1, ..., 127 reicht und jeweils vier mal vorkommt.

Eine Übersicht über die verwendeten Variablen befindet sich in Tabelle 10. Als Einflussgrößen dienen die Variablen *Th*, *Age*, *Age2* und *Sex*. Graphik 14 stellt die Verteilung des Geschlechts und der Therapieform dar. Fast dreiviertel der Befragten sind männlich und etwa die Hälfte der Befragten wurden der Therapiegruppe zugeordnet. Zum Zeitpunkt der Befragung waren die Probanden zwischen 14 und 52 Jahre alt. Zur besseren Interpretation wird die Variable *Age* durch das mittlere Alter zentriert. Dieses liegt bei 30 Jahren wie Abbildung 15 zeigt.

Name der Variable	Beschreibung
<i>y</i>	Schmerzempfinden an Tag 1, 3, 7 oder 10
<i>Th</i>	Therapie (ja=1/ nein=0)
<i>Age</i>	Alter in Jahren (um den Mittelwert (30 Jahre) zentriert)
<i>Age2</i>	Quadriertes, zentriertes Alter
<i>Sex</i>	Geschlecht (Mann=1/ Frau=0)
<i>Person</i>	Personennummer

Tabelle 10: Übersicht über die verwendeten Variablen des **Knee**-Datenbeispiels

Zur Modellierung wird die im **R**-Package **vcrpart** implementierte Funktion `olmm()` verwendet. Die folgenden Erläuterungen dieses Pakets und Spezifikationen der Funktion basieren auf Buegin and Ritschard (2016) S. 9 ff.. Das kumulative Logit-Modell resultiert, wenn „`family=cumulative(link=„logit“)`“ gesetzt wird und das Adjacent-Category Logit-Modell bei „`family=adjacent(link=„logit“)`“. Zur Schätzung der Koeffizienten wird das Gauss-Hermite Quadrature-Verfahren mit 7 Quadrature Points, dem Defaultwert, verwendet. Die Anzahl der Quadrature Points kann über den „`ghx`“-Parameter verändert werden. Die Variable *Person* wird als Random Intercept durch den Befehl „`re(1|Person)`“ in das Modell aufgenommen. Es wird angenommen, dass dieser normalverteilt ist mit Erwartungswert 0.

Abbildung 14: Verteilung der Einflussvariable *Sex* und *Th* des **Knee**-Datenbeispiels

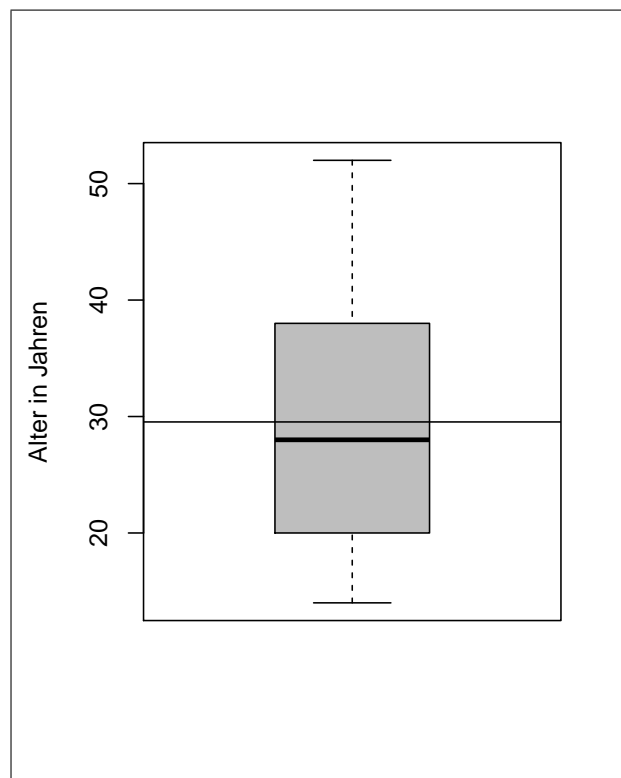


Abbildung 15: Verteilung der Einflussvariable *Age* des **Knee**-Datenbeispiels

Bei der Interpretation ist zu beachten, dass beim kumulativen Modell der Logit Link

$$\text{logit}(P(Y \leq r)) = \log \left(\frac{P(Y \leq r)}{P(Y > r)} \right)$$

verwendet wird, wie in Kapitel 3.1 und in dem Beispiel zu den **Allbus**-Daten in Kapitel 4.2. Für das Adjacent-Type Logit-Modell werden die Logits für die Wahrscheinlichkeit der niedrigeren von zwei benachbarten Kategorien

$$\text{logit}(P(Y = r | Y = r \text{ oder } Y = r + 1)) = \log \left(\frac{P(Y = r)}{P(Y = r + 1)} \right)$$

genutzt. Insbesondere in Kapitel 3.2 und bei der Modellierung der Vertrauensskalen in Kapitel 4.2 wurde jedoch die Chance auf die höhere von zwei benachbarten Kategorien betrachtet. Die daraus resultierende Veränderung der Interpretationsrichtung des Adjacent-Type Modells zu den Interpretationen der vorherigen Kapitel sollte deshalb dringend bei der folgenden Erläuterung der Ergebnisse beachtet werden.

Zunächst wird für beide Modelle ein Likelihood-Ratio-Test durchgeführt. Er testet, ob das Mo-

dell, bei dem für alle Einflussgrößen Proportional-Odds angenommen werden, dem Modell ohne Proportionalitätsannahme für alle Größen zu bevorzugen ist. Dazu wird die im Package **vcrrpart** implementierte Funktion `anova()` verwendet (siehe Buergin and Ritschard (2016): S. 17 ff.). Der Test ergibt für beide Modelle, dass kategoriespezifische Koeffizienten in das Modell aufgenommen werden sollten. Die Tabellen 11 und 12 zeigen die p-Werte für den Likelihood-Ratio-Test der Proportional-Odds Annahme jeder einzelnen Einflussgröße. Sowohl für das kumulative als auch das Adjacent-Type Logit-Modell kann die Annahmen kategorieunspezifischer Koeffizienten für die Kovariablen *Th* und *Sex* abgelehnt werden. Um die Variablen *Th* und *Sex* mit kategoriespezifischen Koeffizienten in das Modell aufzunehmen, wird das Argument „`ce(Variablenname)`“ verwendet.

Aufgrund der gleichen Testergebnisse der Proportionalitätsannahmen liegt beiden der gleiche Prädiktor zugrunde:

$$\eta_{ir} = b_i + \beta_{0r} + \beta_{Th1,r} x_{Th1,i} + \beta_{age} x_{age,i} + \beta_{age2} x_{age2,i} + \beta_{Sex1,r} x_{Sex1,i}$$

mit $r \in 1, \dots, 4$

$$x_{Th1,i} = \begin{cases} 1, & \text{Therapie} \\ 0, & \text{keine Therapie} \end{cases}$$

Variablenname	p-Wert
Th	0.000 ***
Age	0.136
Age2	1.000
Sex	0.012 *

Tabelle 11: Testergebnis für die Partial-Proportional-Odds Annahme für das kumulative Logit-Modell des longitudinalen **Knee**-Datensatzes

Variablenname	p-Wert
Th	0.000 ***
Age	0.229
Age2	0.397
Sex	0.001 ***

Tabelle 12: Testergebnis für die Partial-Proportional-Odds Annahme für das Adjacent-Category Logit-Modells des longitudinalen **Knee**-Datensatzes

$$x_{\text{Sex1},i} = \begin{cases} 1, & \text{weiblich} \\ 0, & \text{männlich} \end{cases}$$

Für den Personeneffekt b_i wird angenommen, dass er normalverteilt mit Erwartungswert 0 ist. Die berechneten Koeffizienten und deren exponierte Werte sind für das kumulative Modell in Tabelle 13 und für das Adjacent-Category Modell in Tabelle 14 zu finden. Die dritte Spalte zeigt die p-Werte des Wald-Tests und die Indikatoren der Signifikanz.

Zunächst werden die kategoriespezifischen Intercepts der beiden Modelle betrachtet. Bei Verwendung von Adjacent-Category Logits zeigt Tabelle 14, dass die Chance auf Schmerzkategorie 1 im Gegensatz zu 2 bei 0.070 liegt, wenn ein 30 jähriger Mann der Placebogruppe betrachtet wird. Die Wahrscheinlichkeit auf Kategorie 2 zu 3 für diesen Mann sinkt dann auf den Faktor 0.039 und steigt dann wieder wie der Wert von *Kat3:Intercept* zeigt. Die Chance auf Kategorie 4 zu Stufe 5 beträgt dann 97.417.

Variable	Estimate	exp(Estimate)	p-value	
Kat1:Intercept	-4.519	0.011	0.000	***
Kat2:Intercept	-3.204	0.041	0.000	***
Kat3:Intercept	0.517	1.677	0.171	
Kat4:Intercept	5.627	277.827	0.000	***
Kat1:Th1	1.242	3.463	0.039	*
Kat2:Th1	2.499	12.170	0.000	***
Kat3:Th1	1.610	5.003	0.000	***
Kat4:Th1	0.292	1.339	0.611	
Kat1:Sex1	1.329	3.777	0.078	
Kat2:Sex1	1.230	3.421	0.018	*
Kat3:Sex1	-0.041	0.960	0.936	
Kat4:Sex1	-0.820	0.440	0.162	
Age	0.065	1.067	0.002	**
Age2	0.013	1.013	0.000	***

	Varianz	Standardabweichung
Personeneffekt	17.419	4.174

Tabelle 13: Koeffiziententabelle für das kumulative Modell mit Partial-Proportional-Odds Annahme für die Modellierung der Schmerzempfindung des **Knee**-Datenbeispiels

Beim kumulativen Modell steigt die Chance auf eine Kategorie niedriger gleich r im Gegensatz zu höheren Kategorien für einen 30 Jahre alte Mann ohne Therapie an, wenn r steigt. Dies liegt an der Verwendung kumulierter Wahrscheinlichkeiten. Besonders stark ist der Zuwachs der Chance von Kategorie 3 oder niedriger zu Kategorie 4 oder kleiner, wie der Vergleich der Werte von *Kat3:Intercept* zu *Kat4:Intercept* zeigt.

Nun werden die Einflüsse der Kovariablen interpretiert. Die Vorzeichen der Koeffizienten sind identisch. Die Wirkung der Therapieform auf die Wahl der Schmerzkategorie variiert über die Stufen von y . Für das kumulative Modell zeigt sich, dass die Chance für Schmerzstufen niedriger gleich r zu höheren Stufen für $r = 1$ für eine Person mit Therapie zu einer der Placebogruppe steigt, dieser Trend sich für $r = 2$ fortsetzt, dann aber für $r = 3$ und $r = 4$ wieder abfällt unter Festhalten der anderen Variablen. Ein ähnliches Bild ergibt sich auch bei der Betrachtung der Koeffizienten der Variable *Th* des Adjacent-Type Modells.

Variable	Estimate	exp(Estimate)	p-value	
Kat1:Intercept	-2.660	0.070	0.000	***
Kat2:Intercept	-3.248	0.039	0.000	***
Kat3:Intercept	0.214	1.239	0.756	
Kat4:Intercept	4.579	97.417	0.000	***
Kat1:Th1	0.246	1.279	0.674	
Kat2:Th1	2.502	12.207	0.019	*
Kat3:Th1	1.286	3.618	0.102	
Kat4:Th1	0.090	1.094	0.828	
Kat1:Sex1	1.502	4.491	0.036	*
Kat2:Sex1	1.317	3.732	0.271	
Kat3:Sex1	-0.315	0.730	0.704	
Kat4:Sex1	-0.929	0.395	0.050	*
Age	0.051	1.052	0.001	***
Age2	0.010	1.010	0.000	***

	Varianz	Standardabweichung
Personeneffekt	10.110	3.180

Tabelle 14: Koeffiziententabelle für das Adjacent-Category Modell mit Partial-Proportional-Odds Annahme für die Modellierung der Schmerzempfindung

Der Therapieeffekt scheint also für kleine Kategorien stark anzusteigen, hat die Person jedoch große Schmerzen, so ist dieser Effekt niedriger. Die Chance auf die Wahl einer Schmerzkategorie r zur nächsthöheren Stufe $r + 1$ fällt beim Adjacent-Type Logit-Modell für eine Frau zu einem Mann für steigendes r ab. Dieser Trend zeigt sich auch bei der Verwendung kumulativer Logits, wie in Spalte 1 von Tabelle 13 zu sehen ist. Die Chance für eine Frau im Gegensatz zu einem Mann liegt für Schmerzstufe 1 im Gegensatz zu höheren Schmerzstufen beim Faktor 3.777. Wird jedoch die Chance für eine Frau zu einem Mann auf Schmerzstufe 4 oder niedriger zu Schmerzstufe 5 betrachtet, so liegt dieser bei dem Faktor 0.440. Beide Interpretationen gelten nur unter Festhalten der Variablen *Age*, *Age2* sowie der Therapieform.

Abbildung 16 stellt die Werte des linearen Prädiktors $\eta_i = \beta_{age}x_{age,i} + \beta_{age2}x_{age2,i}$ für verschiedene Altersstufen dar. Es zeigt sich im ähnlichen Verlauf der beiden Graphen, dass der Beitrag zum Prädiktor beider Modelle positiv bleibt, jedoch bis 30 Jahre sinkt und dann wieder ansteigt.

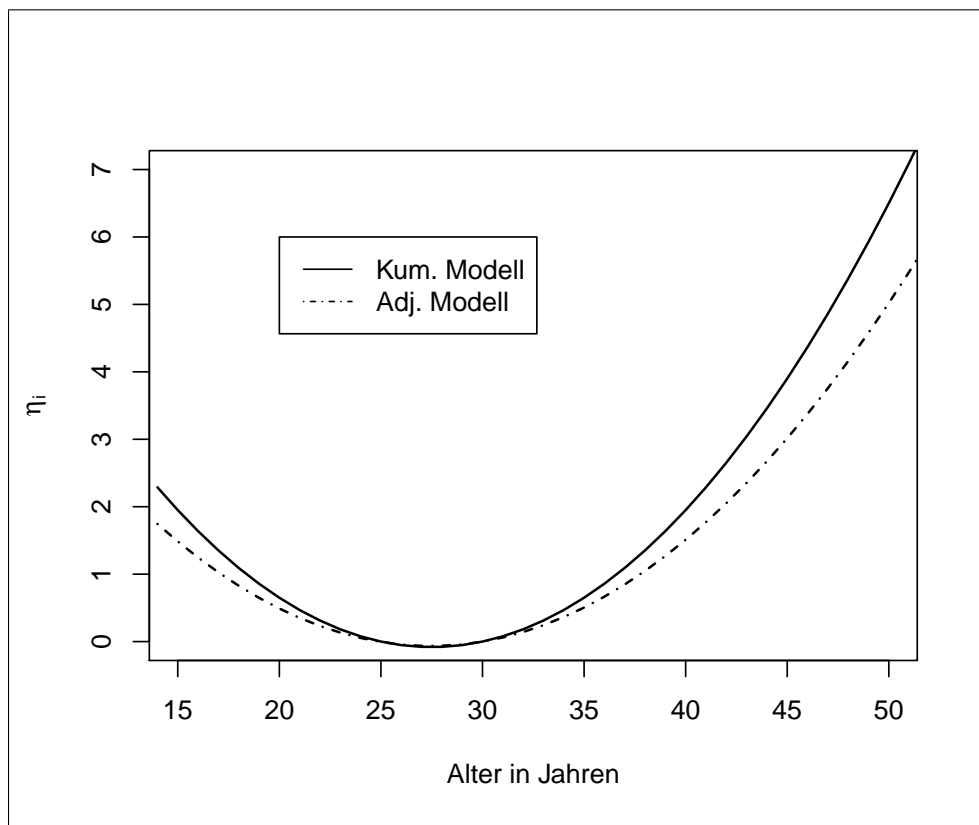


Abbildung 16: Werte des linearen Prädiktors $\eta_i = \beta_{age}x_{age,i} + \beta_{age2}x_{age2,i}$ zur Modellierung der Schmerzkategorien nach dem Alter in Jahren des kumulativen Modells und Adjacent-Category Modells

Unterhalb der Koeffizientenwerte in den Tabellen 13 und 14 ist die Streuung des Personeneffekts angegeben. Dieser ist für das kumulative Modell mit einer Standardabweichung von 4.174 höher als beim Adjacent-Type Modell mit 3.180. Aufgrund der hohen Streuung kann angenommen werden, dass sich der Personeneffekt zwischen den Personen stark unterscheidet und ungleich 0 ist.

Zum Vergleich des Modellfits wurde wieder das AIC und die Devianz als Kriterien gewählt. Beim kumulativen Modell wurde ein AIC von 1105.894 und eine Devianz von 1075.894 erhalten. Diese sind geringer als beim Adjacent-Type Modell mit einem AIC von 1113.626 und einer Devianz von 1083.626. Es wird angenommen, dass das kumulative Partial-Proportional-Odds Modell die Datensituation besser fittet. Da der Unterschied aber nur bei einer Differenz von etwa 8 liegt, kann angenommen werden, dass die beiden Modelle einen ähnlichen Fit aufweisen. Dies bestätigten auch die ausführlichen Interpretationen der Koeffizienten.

6 Fazit

Ziel der Arbeit war der empirische Vergleich ordinaler Regressionsmodelle. Dabei lag der Fokus auf dem kumulativen und Adjacent-Category Logit-Modell. Beide Verfahren eignen sich für das Fitten ordinaler, kategorialer Zielvariablen besser als das Multinomialmodell, da sie die zugrundeliegende Ordnung der Kategorien beachten. Die Anwendung beider Verfahren auf den **Allbus**- und **Knee**-Datensatz sowie der anschließende Vergleich sollten Aufschluss darüber geben, welche Methode zu bevorzugen ist.

Ein Vorteil des Adjacent-Type Logit-Modells ist, dass trotz hoher Restriktion des Prädiktors eine sinnvolle Modellberechnung möglich ist. Für das kumulative Modell ist das nicht immer der Fall, wie das Beispiel in Abschnitt 4.2.2 zeigte. Tests der Partial-Proportional-Odds Annahme konnten für beide Modelle durchgeführt werden und lieferten ähnliche Ergebnisse. Bei den angeführten Beispielen wurde deshalb immer der gleiche Prädiktor verwendet. Nach der Berechnung des Modells sollte die Art der verwendeten Logits beachtet werden, da dies zu unterschiedlichen Werten und Vorzeichen der Koeffizienten führt unter Beachtung verschiedener Interpretationsrichtungen. Unter Betrachtung der einzelnen Koeffizienten ergaben sich geringfügige Unterschiede, welche aber im Mittel ähnliche Trends aufwiesen. Um die Modellgüte zu prüfen wurden die AICs und Devianzen verglichen. Das AIC hat gegenüber der Devianz den Vorteil, dass es auch bei einer unterschiedlichen Anzahl an Parametern vergleichbare Werte liefert, wenn die gleiche Verteilung zugrunde liegt. Das AIC des kumulativen Logit-Modells war bei der Modellierung des Vertrauens in die Regierung und in die Mitmenschen niedriger als das AIC des Adjacent-Type Modells. Sie unterschieden sich aber nur geringfügig. Bei den Longitudinaldaten, für die zusätzlich noch der Personeneffekt in Form von Random Intercepts in die Modelle aufgenommen wurde, agierten beide Verfahren ebenfalls ähnlich. Für das kumulative Modell streuten die Personeneffekte aber höher als für das Adjacent-Type Modell.

Im Allgemeinen gilt, dass sowohl das kumulative Logit-Modell als auch das Adjacent-Category Logit-Modell in ähnlichen Situation einen vergleichbaren Fit aufweisen. Wird die Betrachtung des Effekts auf individuelle Responsekategorien präferiert, so eignet sich das Adjacent-Type Modell. Wird aber die Betrachtung gruppierter Kategorien über die gesamte Skala der zugrundeliegenden latenten Variable bevorzugt, so sollte das kumulative Modell verwendet werden.

Literaturverzeichnis

- Agresti, A. (2010). *Analysis of Ordinal Categorical Data*, Wiley series in probability and statistics, 2. ed. edn, Wiley, Hoboken, NJ.
- Agresti, A. (2013). *Categorical Data Analysis*, Vol. 792 of *Wiley series in probability and statistics*, 3rd ed. edn, Wiley, Hoboken, NJ.
- Buergin, R. and Ritschard, G. (2016). Package ‘vcrpart’. (Stand: 18.04.2016).
URL: <https://cran.r-project.org/web/packages/vcrpart/vcrpart.pdf>
- Christensen, R. (2015). Package ‘ordinal’. (Stand: 19.04.2016).
URL: <https://cran.r-project.org/web/packages/ordinal/ordinal.pdf>
- Fahrmeir, L., Kneib, T. and Lang, S. (2009). *Regression: Modelle, Methoden und Anwendungen*, Statistik und ihre Anwendungen, Springer, Berlin, Heidelberg.
- Fahrmeir, L., Künstler, R., Pigeot, I. and Tutz, G. (2011). *Statistik - Der Weg zur Datenanalyse*, 7. edn, Springer, Heidelberg.
- Lipsitz, S. R., Fitzmaurice, G. M. and Molenberghs, G. (1996). Goodness-of-Fit Tests for Ordinal Response Regression Models, *Applied Statistics* **45**(2): 175–190.
- Peter McCullagh (1980). Regression Models for Ordinal Models, *Journal of the Royal Statistical Society. Series B (Methodological)* (Volume 42): 109–142.
- Terwey, M. and Baltzer, S. (2013). ALLBUS 2012 - Variable Report: GESIS-Variable Reports Nr. 2013/16.
- Tutz, G. (2012). *Regression for Categorical Data*, Cambridge series in statistical and probabilistic mathematics, Cambridge University Press, Cambridge.
- WHO - World Health Organization (2000). Global Database on Body Mass Index. (Stand: 30.03.2016).
URL: http://apps.who.int/bmi/index.jsp?introPage=intro_3.html
- Yee, T. W. (2010). The VGAM Package for Categorical Data Analysis, *Journal of Statistical Software* (Volume 32 (10)): 1–34.

Yee, T. W. (2016). Package 'VGAM': Vector Generalized Linear and Additive Models. (Stand: 28.04.2016).

URL: *<https://cran.r-project.org/web/packages/VGAM/VGAM.pdf>*

Abbildungsverzeichnis

1	Schwellenwerte und Dichten latente Variable zwei Ausprägungen von x_i	10
2	Darstellung Wahrscheinlichkeiten für das kumulative Proportional-Odds Modell	13
3	Verteilung der Variable <i>government</i> des Allbus -Datenbeispiels	27
4	Verteilung der Variable <i>fellow</i> des Allbus -Datenbeispiels	28
5	Kreisdiagramme zu den Variablen <i>sex</i> , <i>partner</i> und <i>employment</i> des Allbus - Datenbeispiels	29
6	Boxplot zur Variable <i>age</i> des Allbus -Datenbeispiels	30
7	Boxplot zur Variable <i>income</i> des Allbus -Datenbeispiels	31
8	Verteilung der Variable <i>life_kat</i> des Allbus -Datenbeispiels	32
9	Verteilung der Variable <i>interest_kat</i> des Allbus -Datenbeispiels	33
10	Verteilung der Variable <i>leftright_kat</i> des Allbus -Datenbeispiels	34
11	Werte des exponierten linearen Prädiktors nach dem Alter des kumulativen und Adjacent-Category Modells für die Zielvariable <i>government</i>	43
12	Werte des exponierten linearen Prädiktors nach dem Alter des kumulativen und Adjacent-Category Modells für die Zielvariable <i>fellow</i>	48
13	Verteilung der Bewertungsabstände nach Therapie ja/nein des Knee -Datenbeispiels	54
14	Verteilung der Einflussvariable <i>Sex</i> und <i>Th</i> des Knee -Datenbeispiels	55
15	Verteilung der Einflussvariable <i>Age</i> des Knee -Datenbeispiels	56
16	Werte des exponierten linearen Prädiktors nach dem Alter des kumulativen und Adjacent-Category Modells im Knee -Datensatz	60
17	Verteilung der Variable <i>justice</i> des Allbus -Datenbeispiels	72

Tabellenverzeichnis

1	Übersicht über die verwendeten Variablen des Allbus -Datenbeispiels	26
2	Testergebnis Partial-Proportional-Odds Annahme kumulatives Logit-Modell Zielvariable <i>government</i> des Allbus -Datenbeispiels	37
3	Testergebnis Partial-Proportional-Odds Annahme Adjacent-Type Logit-Modell Zielvariable <i>government</i> des Allbus -Datenbeispiels	37
4	Koeffiziententabelle kumulatives Logit-Modell für die Zielvariable <i>government</i> des Allbus -Datenbeispiels	39
5	Koeffiziententabelle Adjacent-Type Logit-Modell für die Zielvariable <i>government</i> des Allbus -Datenbeispiels	41
6	Testergebnis zur Partial-Proportional-Odds Annahme kumulatives Logit-Modell Zielvariable <i>fellow</i> des Allbus -Datenbeispiels	44
7	Testergebnis zur Partial-Proportional-Odds Annahme Adjacent-Type Logit-Modell Zielvariable <i>fellow</i> des Allbus -Datenbeispiels	45
8	Koeffiziententabelle kumulatives Logit-Modell für die Zielvariable <i>fellow</i> des Allbus -Datenbeispiels	46
9	Koeffiziententabelle Adjacent-Type Logit-Modell für die Zielvariable <i>fellow</i> des Allbus -Datenbeispiels	46
10	Übersicht über die verwendeten Variablen des Knee -Datenbeispiels	55
11	Testergebnis für die Partial-Proportional-Odds Annahme für das kumulative Logit-Modell des Knee -Datenbeispiels	57
12	Testergebnis für die Partial-Proportional-Odds Annahme für das Adjacent-Type Logit-Modell des Knee -Datenbeispiels	57
13	Koeffiziententabelle kumulative Logit-Modell mit Partial-Proportional-Odds Annahme des Knee -Datenbeispiels	58
14	Koeffiziententabelle Adjacent-Type Logit-Modell mit Partial-Proportional-Odds Annahme des Knee -Datenbeispiels	59
15	Testergebnis Partial-Proportional-Odds Annahme kumulatives Logit-Modell zur Zielvariable <i>justice</i> des Allbus -Datenbeispiels	73
16	Testergebnis Partial-Proportional-Odds Annahme Adjacent-Category Logit-Modell zur Zielvariable <i>justice</i> des Allbus -Datenbeispiels	73
17	Koeffiziententabelle kumulatives Logit-Modell der Zielvariable <i>justice</i> des Allbus -Datenbeispiels	74

18 Koeffiziententabelle Adjacent-Type Logit-Modell der Zielvariable *just-ice* des
Allbus-Datenbeispiels 75

A Verfügbare Dateien

Die dieser Arbeit beiliegende CD beinhaltet neben der verfassten Arbeit als pdf-Datei zwei weitere Ordner:

1. „R-Beispiele“ mit den Unterordnern Programme mit den Source-Dateien der aufgeführten R-Beispiele des **Allbus**- und **Knee**-Datensatzes, einem Ergebnisordner und Ordner der Datensätze
2. „Graphiken“ mit den R-Code der für diese Arbeit erstellten Graphiken sowie den Graphiken als pdf-Dateien

Es wurden die R-Pakete **vcrpart**, **VGAM** und **xtable** verwendet. Nun folgt eine kurze alphabetische Auflistung der verfügbaren Source- und RData-Dateien.

Source-Dateien

Dateiname	Beschreibung	Funktionen
allbus.R	Erstellung des Datensatzes zur Analyse der Zielvariablen <i>government</i> und <i>fellow</i> des Allbus -Datensatzes	
allbus3.R	Erstellung des Datensatzes zur Analyse der Zielvariable <i>justice</i> des Allbus -Datensatzes	
fellow.R	Tests und Modellberechnung kumulatives und Adjacent-Type Logit-Modell für die Zielvariable <i>fellow</i> des Allbus -Datensatzes	AIC(), vglm()
government.R	Tests und Modellberechnung kumulatives und Adjacent-Type Logit-Modell für die Zielvariable <i>government</i> des Allbus -Datensatzes	AIC(), vglm()
graphiken.R	Erstellung der Graphiken für den theoretischen Teil der Arbeit	

hilffkt_allbus.R	Hilfsfunktionen für den Allbus -Datensatz	tabelle(), xtable()
hilffkt_knee.R	Hilfsfunktionen für den Knee -Datensatz	tabelle(), xtable()
justice.R	Tests und Modellberechnung kumulatives und Adjacent-Type Logit-Modell für die Zielvariable <i>justice</i> des Allbus -Datensatzes	AIC(), vglm()
knee.R	Erstellung des Datensatzes zur Analyse der Schmerzkategorien des Knee -Datensatzes	
knee_modelle.R	Tests und Modellberechnung kumulatives und Adjacent-Type Logit-Modell für die Modellierung der Schmerzkategorien des Knee -Datensatzes	anova(), olmm()

RData-Dateien

Dateiname	Beschreibung	Variablen
Allbus.RData	Datensatz zur Analyse der Zielvariablen <i>government</i> und <i>fellow</i> des Allbus -Datenbeispiels	<i>government, fellow, interest, leftright, income, employment, partner, sex, age, life, income_old, age_old, age2, life_kat, leftright_kat, interest_kat</i>
Justice.RData	Datensatz zur Analyse der Zielvariablen <i>justice</i> des Allbus -Datenbeispiels	<i>justice interest, leftright, income, employment, partner, sex, age, life, income_old, age_old, age2, life_kat, leftright_kat, interest_kat</i>

Knee.RData	Ursprünglicher Knee -Datensatz	<i>N, Th, Age, Sex, R1, R2, R, R4</i>
Knee_kum.RData	Datensatz zur Analyse der Schmerzkategorien des Knee -Datenbeispiels	<i>Person, y, Sex, Th, Age, Age_old, Age2</i>
Urdata.RData	Ursprünglicher <i>Allbus</i> -Datensatz	752 Variablen

B R-Code und Ergebnistabellen

B.1 Likelihood-Ratio-Test kumulatives Logit-Modell Zielvariable *government* des Allbus-Datensatzes

Modell unter Proportional-Odds Annahme:

```
> cum1<-vglm(allbus$government~allbus$interest_kat+
+           allbus$income+allbus$employment+
+           allbus$partner+allbus$sex+allbus$age+
+           allbus$age2+allbus$life_kat+
+           allbus$leftright_kat,
+           family=cumulative(parallel=TRUE))
```

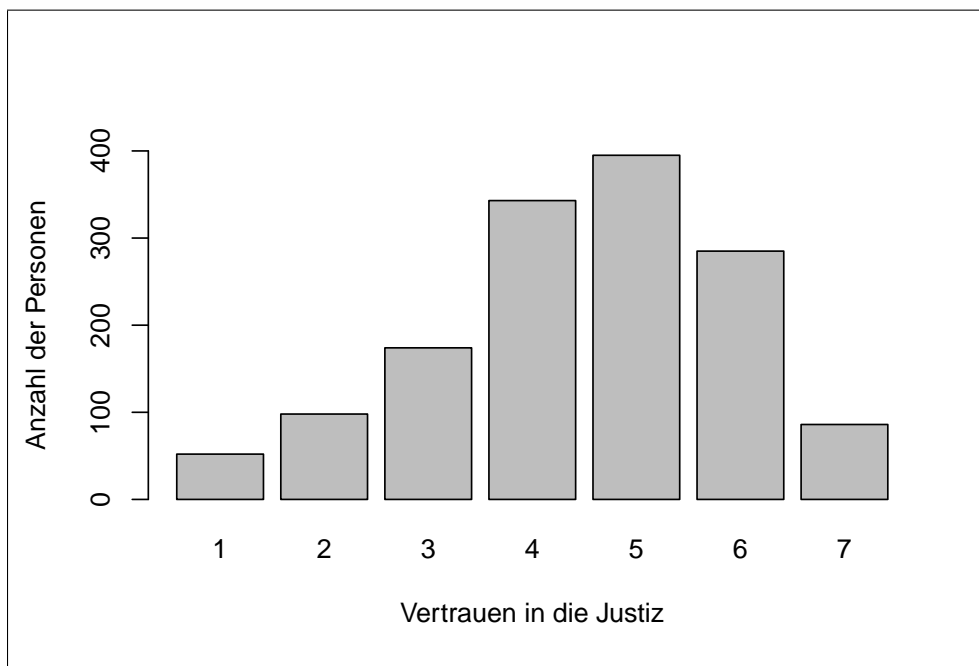
Modell unter Annahme kategoriespezifischer Koeffizienten aller Einflussgrößen:

```
> cum1<-vglm(allbus$government~allbus$interest_kat+
+           allbus$income+allbus$employment+
+           allbus$partner+allbus$sex+allbus$age+
+           allbus$age2+allbus$life_kat+
+           allbus$leftright_kat,
+           family=cumulative(parallel=FALSE))
```

Es wurden folgende Warnungen von R ausgegeben:

Warnmeldungen:

```
1: In Deviance.categorical.data.vgam(mu = mu, y = y, w = w,
... : fitted values close to 0 or 1
2: In Deviance.categorical.data.vgam(mu = mu, y = y, w = w,
... : fitted values close to 0 or 1
:
39: In Deviance.categorical.data.vgam(mu = mu,
y = y, w = w, ... : fitted values close to 0 or 1
40: In vglm.fitter(x = x, y = y, w = w, offset = offset,
Xm2 = Xm2, ... : convergence not obtained in 30 iterations
41: In log(prob) : NaNs produced
```

B.2 Auswertung Zielvariable *justice* des Allbus-DatensatzesAbbildung 17: Verteilung der Variable *justice* des **Allbus**-Datenbeispiels

Die Tabellen 15 und 16 zeigen, dass sowohl für das kumulative als auch das Adjacent-Type Logit-Modell die Proportional-Odds Annahme ausreichend ist.

Variablenname	p-Wert
sex	0.788
partner	0.497
employment	0.120
age	0.778
age2	0.604
income	0.069
life_kat	0.108
interest_kat	0.103
leftright_kat	0.389

Tabelle 15: Testergebnis Partial-Proportional-Odds Annahme kumulatives Logit-Modell zur Zielvariable *justice* des **Allbus**-Datenbeispiels

Variablenname	p-Wert
sex	0.761
partner	0.494
employment	0.089
age	0.485
age2	0.654
income	0.157
life_kat	0.136
interest_kat	0.210
leftright_kat	0.429

Tabelle 16: Testergebnis Partial-Proportional-Odds Annahme Adjacent-Category Logit-Modell zur Zielvariable *justice* des **Allbus**-Datenbeispiels

Variable	Estimate	exp(Estimate)	p-Wert	
Intercept:1	-2.744	0.064	0.000	***
Intercept:2	-1.583	0.205	0.000	***
Intercept:3	-0.631	0.532	0.002	**
Intercept:4	0.517	1.677	0.009	**
Intercept:5	1.757	5.795	0.000	***
Intercept:6	3.482	32.525	0.000	***
sex1	-0.116	0.890	0.260	
partner1	-0.072	0.931	0.518	
employment1	-0.248	0.780	0.080	
age	0.010	1.010	0.005	**
age2	-0.000	1.000	0.008	**
income	-0.013	0.987	0.007	**
life_kat1	0.298	1.347	0.343	
life_kat3	-0.238	0.788	0.070	
interest_kat1	0.909	2.482	0.000	***
interest_kat2	0.155	1.168	0.235	
interest_kat4	-0.064	0.938	0.632	
interest_kat5	-0.130	0.878	0.444	
leftright_katl	0.132	1.141	0.278	
leftright_katr	0.434	1.543	0.027	*

Tabelle 17: Koeffiziententabelle des kumulativen Logit-Modells mit Proportional-Odds Annahme der Zielvariable *justice* des **Allbus**-Datenbeispiels

Variable	Estimate	exp(Estimate)	p-Wert	
Intercept:1	0.484	1.623	0.010	**
Intercept:2	0.383	1.467	0.009	**
Intercept:3	0.452	1.571	0.000	***
Intercept:4	-0.117	0.890	0.271	
Intercept:5	-0.613	0.542	0.000	***
Intercept:6	-1.511	0.221	0.000	***
sex1	0.050	1.051	0.227	
partner1	0.026	1.026	0.553	
employment1	0.086	1.090	0.127	
age	-0.004	0.996	0.004	**
age2	0.000	1.000	0.010	**
income	0.006	1.006	0.003	**
life_kat1	-0.112	0.894	0.351	
life_kat3	0.085	1.089	0.101	
interest_kat1	-0.349	0.705	0.000	***
interest_kat2	-0.060	0.942	0.247	
interest_kat4	0.036	1.037	0.507	
interest_kat5	0.022	1.022	0.746	
leftright_katl	-0.046	0.955	0.343	
leftright_katr	-0.156	0.856	0.043	*

Tabelle 18: Koeffiziententabelle des Adjacent-Type Logit-Modells mit Proportional-Odds Annahme der Zielvariable *justice* des **Allbus**-Datenbeispiels

Die Vorzeichen der Koeffizienten von Tabelle 17 zu 18 sind umgekehrt wegen der unterschiedlichen Art der verwendeten Logits (siehe Abschnitt 4.2). Jedoch wird bei beiden Modellen für die gleichen Kovariablen angenommen, dass sie einen Effekt ungleich 0 auf die Zielvariable *justice* haben.

```
> AIC(justice_kum)
[1] 4955.918
> AIC(justice_adj)
[1] 4954.457
> deviance(justice_kum)
[1] 4915.918
```

```
> deviance(justice_adj)
[1] 4914.457
```

Das AIC und die Devianz zeigen, dass der Modellfit des Adjacent-Type Logit-Modells besser ist als der des kumulativen Logit-Modells.