

Technische Universität München

ZENTRUM MATHEMATIK

Regressionsverfahren mit nominaler Zielvariable

Diplomarbeit

von

Kathleen E H R L I C H

Themensteller: Prof. Dr. Claudia Czado

Betreuer: Prof. Dr. Claudia Czado

Dipl. Math. Sergiy Prokopenko

Abgabetermin: 1. Mai 2002

Inhaltsverzeichnis

Abbildungsverzeichnis	ii
Tabellenverzeichnis	iii
Abkürzungen und Zeichenerklärungen	vii
1 Einleitung	1
2 Grundlagen	3
2.1 Einführung in die Regressionsanalyse	3
2.1.1 Statistisches Modell	3
2.1.2 Merkmalstypen	5
2.1.3 Kodierung qualitativer Einflussgrößen	6
2.2 Gruppierte und ungruppierte Daten	8
2.3 Prinzip der Maximum-Likelihood-Schätzung	9
2.4 Das generalisierte lineare Modell (GLM)	10
2.4.1 Maximum-Likelihood-Schätzung in GLMs	15
2.4.2 Maximum-Quasi-Likelihood	17
3 Multinomiale Modelle	21
3.1 Modellbildung	21
3.2 Die Multinomialverteilung	21
3.3 Das multinomiale Logit-Modell	22
3.4 Das bedingte Logit-Modell	31
3.5 Das multinomiale Logit-Modell als verallgemeinertes lineares Modell	31
3.6 Explorative Datenanalyse	32
3.6.1 Explorative Datenanalyse zum Erkennen marginaler Einflüsse einer Kovariable	32
3.6.2 Explorative Datenanalyse zum Erkennen von Interaktionen	37
3.7 Schätzen und Testen für multinomiale Modelle	40
3.7.1 Maximum-Likelihood-Schätzung	40
3.7.2 Anpassungstests und Residuen	55
3.7.3 Akaike-Kriterium	60

3.7.4	Hypothesentest und Wald-Statistik	61
3.7.5	Modellverifikation	63
3.7.6	Modellwahlstrategie bei Daten mit vielen Kovariablen	66
3.7.7	Effektivitätsmaße vom Typ R^2	68
3.8	Alternativen und Vorteile des Logit-Modells	69
4	Zufällige Effekte	71
4.1	Zufällige Effekte in generalisierten linearen Modellen	71
4.2	Penalisierte Quasi-Likelihood Methode	73
5	Anwendung auf Mobilitätsdaten	79
5.1	Erhebung und Beschreibung der Daten	79
5.1.1	Struktur der Daten	81
5.1.2	Versuchsplanung und Datenaufbereitung	95
5.2	Datenanalyse	97
5.2.1	Explorative Datenanalyse zum Erkennen von Haupteffekten	97
5.2.2	Haupteffektmodell	100
5.2.3	Explorative Datenanalyse zum Erkennen von Interaktionen	101
5.2.4	Modell mit Interaktionen	103
5.3	Interpretation der Parameter	107
5.4	Profile der Verkehrsmittelwahl	126
6	Zusammenfassung und Ausblick	139
A	Numerische Bestimmung der ML-Schätzung	141
B	S-Plus Funktionen	143
B.1	S-Plus Funktion zur explorativen Datenanalyse von Haupteffekten	143
B.2	S-Plus Funktion zum Berechnen der relativen Häufigkeiten	144
B.3	S-Plus Funktion zur explorativen Datenanalyse von Interaktionen	145
B.4	S-Plus Funktion zum Vergleich empirischer und geschätzter Logits	147
C	Datensatzbeschreibung	149
D	Ergebnisse der explorativen Datenanalyse	153
D.1	Plots einiger Haupteffekte	153
D.2	Interaktionsplots	160
	Literaturverzeichnis	163

Abbildungsverzeichnis

1	Alter als Haupteffekt	35
2	Netzkarte als Haupteffekt	36
3	Interaktion der Regressoren Nutzung und Alter	39
4	Vergleich empirischer Logits mit geschätzten Paramtern $\hat{\beta}_{ir}$	65
5	Zufrieden mit der Anbindung an ÖV	87
6	Dauer des Weges/der Fahrt	88
7	Wege zu Fuß	89
8	Wege mit ÖV	89
9	Streckentyp	89
10	Gründe für die Verkehrsmittelwahl	90
11	Anzahl der zurückgelegten Wege in den verschiedenen Altersgruppen nach Verkehrsbereichen	93
12	Anzahl der zurückgelegten Wege in den verschiedenen Einkommensgruppen	94
13	Anteil der Wege (n=2554 Wege) nach Verkehrsbereichen	97
14	Geschätzte Chancenverhältnisse in Abhängigkeit vom Mobilitätstag	109
15	Geschätzte Chancenverhältnisse in Abhängigkeit von der Haushaltsgröße .	110
16	Geschätzte Chancenverhältnisse in Abhängigkeit von der Entfernung zur nächstgelegenen Haltestelle	111
17	Geschätzte Chancenverhältnisse in Abhängigkeit von der Entfernung [km] .	112
18	Geschätzte Chancenverhältnisse in Abhängigkeit von der Wegedauer [min]	113
19	Geschätzte Chancenverhältnisse in Abhängigkeit vom Eindruck der Fahrt / des Weges	114
20	Geschätzte odds in Abhängigkeit von Nutzung und Haushaltseinkommen .	116
21	Geschätztes Chancenverhältnis in Abhängigkeit von Nutzung und HH- Einkommen	118
22	Geschätzte odds in Abhängigkeit von Haushaltseinkommen, persönlichen Einkommen und Bahncard-Besitz	119
23	Geschätzte log-odds in Abhängigkeit von Bahncard-Besitz und Einkommen	120
24	Geschätzte Chancenverhältnisse in Abhängigkeit von Haushalts- und persönli- chen Einkommen	121
25	Geschätzte Chancenverhältnisse in Abhängigkeit von persönlichen Einkom- men und Bahncard-Besitz	122
26	Geschätzte odds in Abhängigkeit von Alter und Nutzung	123
27	Geschätzte Chancenverhältnisse in Abhängigkeit von Alter und Nutzung .	124
28	Residuenplot in Abhängigkeit von $\eta_r = \mathbf{x}^T \boldsymbol{\beta}_r$	125
29	Residuenplot in Abhängigkeit von den Beobachtungen	125

30	Profil – Pkw-Hauptnutzer alleine unterwegs	129
31	Profil – Pkw-Mitnutzer alleine unterwegs	130
32	Profil – Nutzer der Kategorie sonstiges alleine unterwegs	130
33	Profilklasse 1 mit mehreren unterwegs	131
34	Profilklasse 2 für Pkw-Hauptnutzer bei unterschiedlicher Entfernung und Wochentag	132
35	Profilklasse 2 für Pkw-Mitnutzer bei unterschiedlicher Entfernung und Wo- chentag	133
36	Profilklasse 2 für Kategorie sonstige bei unterschiedlicher Entfernung und Wochentag	134
37	Profilklasse 3 bei unterschiedlicher Wegedauer, falls der Weg als nicht be- quem empfunden wurde	135
38	Profilklasse 3 bei unterschiedlicher Wegedauer, falls der Weg als bequem empfunden wurde	135
39	Profilklasse 4 bei unterschiedlichem Alter	137
40	Haupteffekt des Regressors Mobilitätstag	153
41	Haupteffekt des Regressors Haushaltgröße	153
42	Haupteffekt des Regressors Haushaltseinkommen	154
43	Haupteffekt des Regressors Anbindung an schienengeb. ÖV	154
44	Haupteffekt des Regressors Entfernung zur nächstgelegenen Haltestelle . .	154
45	Haupteffekt des Regressors gutes Wetter	155
46	Haupteffekt des Regressors Entfernung [km]	155
47	Haupteffekt des Regressors Streckentyp	155
48	Haupteffekt des Regressors Weg alleine oder mit mehreren zurückgelegt . .	156
49	Haupteffekt des Regressors vorher mit diesem Verkehrsmittel unterwegs . .	156
50	Haupteffekt Gewohnheit	156
51	Haupteffekt des Regressors gute Entfernung	157
52	Haupteffekt des Regressors Bequemlichkeit	157
53	Haupteffekt des Regressors kostengünstig	157
54	Haupteffekt des Regressors Wegedauer	158
55	Haupteffekt des Regressors Eindruck des Weges / der Fahrt	158
56	Haupteffekt des Regressors Nutzung	158
57	Haupteffekt des Regressors Rad-Verfügung	159
58	Haupteffekt des Regressors persönliches Einkommen	159
59	Haupteffekt des Regressors Wegzweck	159
60	Interaktion der Regressoren Haushaltseinkommen und Pkw-Nutzung	160
61	Interaktion der Regressoren Haushalts- und persönliches Einkommen	160
62	Interaktion der Regressoren Bahncard und persönliches Einkommen	161

63	Interaktion der Regressoren Alter und Haushaltseinkommen	161
----	--	-----

Tabellenverzeichnis

1	Daten aus der Mobilitätsstudie '97	8
2	Daten aus der Mobilitätsstudie '97	27
3	logarithmierte relative Chancen	28
4	Daten aus der Mobilitätsstudie '97	29
5	Multinomiales Modell für die Verkehrsmittelwahl	30
6	Steigungsparameter für Alter der Population Führerschein	30
7	Parameterschätzer für Alter	54
8	AIC für Alter	61
9	Ergebnisse der Explorative Analyse für [-0.5ex] marginale Effekte einzelner Kovariablen	100
10	Likelihood-Quotienten-Tests der Haupteinflussgrößen	101
11	Likelihood-Quotienten-Tests der Interaktionen	103
12	Parameterschätzungen für das Modell mit Interaktionen	107
13	Odds-ratios der Mobilitätstage	108
14	Odds-ratios der Haushaltsgröße	110
15	Geschätzte odds in Abhängigkeit von der Nutzung und vom HH-Einkommen	116
16	Geschätzte odds-ratios in Abhängigkeit von der Nutzung und vom HH- Einkommen	117
17	Geschätzte odds-ratios in Abhängigkeit von Alter und Nutzung	123
18	Profile	128

Abkürzungen

GLM	Generalisiertes lineares Modell
GLMM	Generalisiertes lineares Mischungsmodell
IV	Individualverkehr
LQ	Likelihood-Quotient
ÖV	öffentlicher Verkehr
Q	Quadratsumme
SS	sums of squares
VM	Verkehrsmittel

Zeichenerklärungen

$A \iff B$	A gilt genau dann, wenn B gilt
\mathbb{R}	Menge der reellen Zahlen
$\text{rg}(\mathbf{A})$	Rang der Matrix \mathbf{A}
$ \mathbf{A} $	Determinante einer quadratischen Matrix \mathbf{A}
\mathbb{E}	Erwartungswert
Var	Varianz
Cov	Kovarianz
\mathbf{x}^T	Transposition eines Vektors \mathbf{x}
$\text{diag}(\mathbf{x})$	$= \begin{pmatrix} x_1 & & 0 \\ & \ddots & \\ 0 & & x_p \end{pmatrix}$
\mathbf{I}_n	$= \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} n \times n\text{-dimensionale Einheitsmatrix}$
\sim	verteilt wie
$\overset{\text{a}}{\sim}$	asymptotisch verteilt
\propto	proportional zu
$N(\mu, \sigma^2)$	Normalverteilung mit Erwartungswert μ und Varianz σ^2
χ_p^2	χ^2 -Verteilung mit p Freiheitsgraden
$\chi_{p,\alpha}^2$	α -Quantil der χ^2 -Verteilung mit p Freiheitsgraden
$M(n, \boldsymbol{\pi})$	Multinomialverteilung mit n Parametern und Wahrscheinlichkeitsvektor $\boldsymbol{\pi}$
$B(n, \pi)$	Binomialverteilung mit n Parametern und Wahrscheinlichkeit π
$\text{Poi}(\lambda)$	Poissonverteilung mit Parameter λ
H_0	Nullhypothese
H_1	Alternative

1 Einleitung

“Each peculiarity in a man is shared by his kinsman but on the average in a less degree.”

Gesetz der universalen Regression, Galton (1889)

Regressionsanalysen gehören in vielen Bereichen der Wissenschaft, zum Beispiel in den Wirtschafts- und Sozialwissenschaften, den Ingenieurwissenschaften, der Umweltforschung, der Biologie, der Medizin oder der Informatik, zu den etablierten Techniken. Die Statistik stellt Methoden und Verfahren zur Verfügung, die der Aufbereitung, tabellarischen und graphischen Darstellung und der Auswertung komplexer Datenstrukturen dienen. Ziel dabei ist es, den Einfluss unabhängiger Variablen auf eine oder mehrere abhängige Variablen zu erfassen. Mit Hilfe bewährter Methoden lassen sich selbst bei relativ kleinen Datenbasen gute Schätzungen der Modellparameter erreichen. Die Parameter werden dann so angepasst, dass eine möglichst gute Beschreibung der Daten erreicht werden kann.

Viele statistische Untersuchungen basieren auf ungeordneten qualitativen (oder polychotomen) Variablen. So sind beispielsweise Volkswirte am Mobilitätsverhalten privater Haushalte, Mediziner an der Wirksamkeit eines Medikamentes und Wirtschaftswissenschaftler am Kaufverhalten eines bestimmten Produktes interessiert.

Das Hauptziel der vorliegenden Arbeit liegt auf der Entwicklung des multinomialen Logit-Modells für die Analyse des Mobilitätsverhaltens privater Haushalte in Alltag und Freizeit. Grundlage ist die Mobilitätsstudie '97 (vgl. Zängler (1999) [29]). Im zweiten Kapitel werden einige Grundlagen zusammengefasst. Das im Mittelpunkt dieser Arbeit stehende multinomiale Logit-Modell wird in Kapitel 3 eingeführt. Von Interesse sind die Parameterschätzungen sowie Tests über diese Parameter. Im vierten Kapitel wird auf die Theorie der zufälligen Effekte eingegangen und im fünften Kapitel das Mobilitätsverhalten anhand der vorher eingeführten Theorie analysiert. Im Anhang ist ein Methode zur numerischen Bestimmung der Maximum-Likelihood-Schätzung beschrieben. Darüber hinaus finden sich dort S-Plus Funktionen zum Erkennen von Haupteinflussgrößen und Interaktionen, zum Vergleich empirischer Logits mit den geschätzten Parametern und zur Berechnung der relativen Häufigkeiten.

Die Analyse wurde mit dem Statistikprogramm S-Plus 2000 Professional Edition for Windows, Release 2 (October 1999) durchgeführt.

2 Grundlagen

Dieses Kapitel enthält wichtige Grundlagen, auf denen die nachfolgende Arbeit aufbaut.

2.1 Einführung in die Regressionsanalyse

Der folgende Abschnitt stellt eine kurze Einführung in die Regressionsanalyse für metrische Responsevariable dar. Für eine ausführliche Darstellung dieses Themengebietes sei auf Fahrmeir, Hamerle & Tutz (1996) [11], Kapitel 4 verwiesen.

2.1.1 Statistisches Modell

Die Regressionsanalyse ist ein statistisches Verfahren zur Bestimmung des linearen Zusammenhangs zwischen einer abhängigen Variable y (Zielvariable, *Response*) und unabhängigen Variablen x_1, x_2, \dots, x_p (Kovariablen, *Regressoren*). Der Einfluss eines Vektors $\mathbf{x}^T = (1, x_1, x_2, \dots, x_p)$ auf y wird im Rahmen der klassischen Regression durch den linearen Ansatz

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \varepsilon = \mathbf{x}^T \boldsymbol{\beta} + \varepsilon \quad (2.1)$$

modelliert, wobei β_k , $k = 0, \dots, p$ die unbekanntes Regressionskoeffizienten darstellen. Der Vektor \mathbf{x}^T beschreibt die Einflussgrößen, ε die Störgröße. Der Intercept β_0 kann als der Mittelwert von y interpretiert werden, falls alle Variablen x_i , $i = 1, \dots, p$ gleich Null sind. ε ist häufig normalverteilt mit $\mathbb{E}(\varepsilon) = 0$ und $Var(\varepsilon) = \sigma^2$. Damit kann (2.1) auch dargestellt werden durch

$$\mathbb{E}(y | \mathbf{x}) = \eta(\mathbf{x}) = \mathbf{x}^T \boldsymbol{\beta},$$

d.h. es wird nur die Abhängigkeit des zu erwartenden Response von den Prädiktoren wiedergegeben. Der Erwartungswert von y gegeben den Einflussgrößen \mathbf{x} muss linear von diesen Größen abhängen. Dabei stellt $\eta(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$ den *linearen Prädiktor* dar.

Ziele der Regressionsanalyse können zum Beispiel der Nachweis einer bekannten Beziehung, das Schätzen der Parameter einer bekannten funktionalen Beziehung, das Erkennen eines funktionalen Zusammenhangs oder die Prognose zukünftiger Ausprägungen der Responsevariable bei gegebenen Regressoren sein.

Für die Beobachtungen $(y_i, x_{i1}, \dots, x_{ip})$, $i = 1, \dots, n$ lautet das Modell

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \dots, n$$

bzw. in Matrixschreibweise

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

In geschlossener Schreibweise erhält man folglich

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

wobei \mathbf{y} den Responsevektor und \mathbf{X} die Designmatrix darstellen. Der Parametervektor $\boldsymbol{\beta}$ kann mit Hilfe der Kleinst-Quadrate-Methode geschätzt werden. Dabei wird die Summe der quadratischen Abweichungen

$$\sum_{i=1}^n (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2$$

minimiert. Jeder minimierende Wert $\hat{\boldsymbol{\beta}}$ erfüllt die Normalgleichungen

$$\mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = \mathbf{X}^T \mathbf{y}.$$

Hat \mathbf{X} vollen Rang ergibt sich als Lösung der lineare Schätzer

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}.$$

Die zugehörigen geschätzten Werte sind durch $\hat{y}_i = \mathbf{x}_i^T \hat{\boldsymbol{\beta}}$ gegeben.

Falls das Modell eine erklärende Variable enthält, kann mit der Methode der kleinsten Quadrate formal an beliebige Daten eine Gerade angepasst werden. Bei zwei erklärenden Variablen kann an die Beobachtungen eine Ebene angepasst werden, usw. Um die Güte der Anpassung zu beurteilen, zerlegt man die Summe der quadratischen Abweichungen wie folgt:

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$Q_{\text{gesamt}} = Q_{\text{Regression}} + Q_{\text{Residuum}},$$

wobei $\bar{y} = \sum_i y_i / n$ der Mittelwert von y ist. Q steht für die Quadratsumme. Häufig schreibt man statt Q auch SS (engl. „sums of squares“). Die Anpassung ist um so besser, je größer der Anteil der Summe der quadratischen Abweichungen ist, der durch die Regressionsgerade erklärt wird. Dazu wird das *Bestimmtheitsmaß*

$$R^2 = \frac{Q_{\text{Regression}}}{Q_{\text{gesamt}}} = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2}$$

verwendet. Das Bestimmtheitsmaß kann folglich zur Überprüfung, ob durch den linearen Ansatz individuelle Beobachtungen vorhersagbar sind herangezogen werden.

2.1.2 Merkmalstypen

Merkmale werden nach der Art ihrer Ausprägungen klassifiziert. Zum einen unterscheidet man quantitative und qualitative Merkmale und zum anderen stetige und diskrete Merkmale.

Die Ausprägungen *quantitativer* Merkmale wie Alter und Gewicht einer Person unterscheiden sich in ihrer Größe, die Ausprägungen *qualitativer* Merkmale wie Geschlecht und Familienstand in ihrer Art.

Geht man von der Zahl der möglichen Ausprägungen eines Merkmals aus, so gelangt man zur Aufteilung in diskrete und stetige Merkmale. Ein *diskretes* Merkmal liegt vor, wenn nur endlich viele Ausprägungen (Kategorien) existieren, sog. Responsekategorien. Es wird daher auch als *kategoriales* Merkmal bezeichnet. Beispiele sind Gehaltsklassen, Familienstand oder die Nutzungshäufigkeit eines Verkehrsmittels. Viele interessierende Variablen besitzen nur Kategorien als Ausprägungen. Die Häufigkeitsuntersuchung wird oft mit den Kategorien „täglich“, „mehrmals pro Woche“, „seltener“ oder „nie“ erhoben. Dann wird auch von qualitativen Variablen gesprochen, da Kategorien wie „ledig“, „verheiratet“, „geschieden“, „verwitwet“ oder „angenehm“, „durchschnittlich“, „unangenehm“ nur eine Qualität mitteilen, nicht aber das Ausmaß einer Größe wie „Alter“ wiedergeben. *Stetige* Merkmale (z.B. Alter oder Temperaturmessungen) können beliebige Werte in einem bestimmten Bereich annehmen.

Eine weitere Unterscheidung von Merkmalstypen ergibt sich dadurch, dass die zu Grunde liegende Meß-Skala betrachtet wird. Man unterscheidet dabei in Nominalskalen, Ordinalskalen und metrischen Skalen. Im Fall einer *Nominalskala* stellen die Zahlenwerte nur Etiketten dar. Es handelt sich um ungeordnete qualitative Variablen. Beispielsweise wird den Ausprägungen, „Kfz vorhanden“, „kein Kfz vorhanden“ mit den Zahlen 1, 2 dieselbe Bedeutung zugeordnet, wie den Zahlen 100, 10 oder beliebigen anderen zwei Zahlen. Die Größe der Zahl selbst dient nur der Identifizierung bzw. der Kodierung eines Sachverhalts. Bei der Analyse des Kaufverhaltens eines Kraftfahrzeuges mit den Kategorien „Volkswagen“, „Ford“, „General Motors“, „Daimler Chrysler“ und „BMW“ kann von ungeordneten Kategorien ausgegangen werden. Die Zuordnung von Zahlen zu den einzelnen Automarken ist beliebig. Im Fall einer *Ordinalskala* erfolgt die Zuordnung so, dass die Kategorien zumindest geordnet sind. Den Kategorien „bis 2000 DM“, „2000+ bis 4000 DM“ und „mehr als 4000 DM“ der Einflussgröße „Haushaltseinkommen“ lässt sich somit 1, 2, 3 oder aber 10, 20, 30 zuordnen. Empirisch sind diese Werte allerdings nicht sinnvoll interpretierbar. Im Fall eines *ordinalen Response* sind Modelle erforderlich, die

keine beliebige Vertauschung der Kategorien zulässt. Die Werte einer *metrischen Skala* sind derart, dass sich zudem Abstände zwischen den Werten interpretieren lassen.

2.1.3 Kodierung qualitativer Einflussgrößen

Die Art und Weise wie eine erklärende Variable in ein Modell eingeht, ist entscheidend für die Eigenschaften des Modells, z.B. für die Interpretation der Parameter des Modells auf die in Abschnitt 3.3 eingegangen wird. *Metrische Einflussgrößen*, die bei vielen physikalischen Größen wie Masse, Gewicht, Volumen oder auch Alter und Einkommen vorliegen, als erklärende Variable sind dem Anwender oft vertraut. Die Einbeziehung von *Faktoren*, d.h. Merkmalen mit wenigen, möglicherweise ungeordneten Ausprägungen wie Geschlecht sind häufig weniger verständlich. Qualitative Merkmale in Form eines Terms $x\beta$ mit $x \in \{1, \dots, I\}$ direkt in das Modell aufzunehmen führt zu uninterpretierbaren Parametern. Es ist daher notwendig, diese Kategorien durch Dummy- (Stellvertreter-) Variablen darzustellen.

Im Rahmen der linearen Regressionsanalyse wird davon ausgegangen, dass die Einflussgrößen in linearer Form den zu erwartenden Response bestimmen. Die Formulierung des linearen Regressionsmodells, bei der metrische, sinnvoll gewichtete, Einflussgrößen, vorausgesetzt werden, ist irreführend. Ist beispielsweise x_1 eine qualitative Größe, die das Geschlecht einer Person kodiert (1: weiblich, 2: männlich), dann ist die Gewichtung $\beta_1 x_1$ offensichtlich nicht sinnvoll. Um diese Willkürlichkeit auszuschließen, wird eine passende Kodierung von x gewählt. Es haben sich zwei Arten der Kodierung durchgesetzt, die *(0-1)-* oder *Dummy-Kodierung* und die *Effektkodierung*.

(0-1)-Kodierung

Die qualitativen Größen (Dummy-Variablen) haben folgende Gestalt

$$x_{A(i)} = \begin{cases} 1 & A = i \\ 0 & \text{sonst,} \end{cases}$$

wobei $i = 1, \dots, I - 1$ ist. Die I -te Kategorie wird implizit durch $x_1 = \dots = x_{I-1} = 0$ erfasst. Das Vorliegen von $A = i$ lässt sich in dieser einfachen Kodierung an der „1“ in der i -ten Komponente ablesen. Die Parameter des Regressionsmodells in (0-1)-Kodierung lassen sich daher interpretieren als

$$\begin{aligned} \beta_0 &= \mu_I \\ \beta_i &= \mu_i - \beta_0, \quad i = 1, \dots, I, \end{aligned}$$

wobei $\mu_i = \mathbb{E}(y | A = i)$. β_i misst also den Einfluss der Kategorie i auf den Response. Kategorie I dient als *Referenzkategorie*. Die Wahl einer Kategorie als Referenzkategorie

ist beliebig. Ist eine Faktorstufe ausgezeichnet, wird diese als Referenzkategorie (häufig die erste oder letzte) gewählt. Bei Untersuchungen im Marketingbereich beispielsweise kann die ausgezeichnete Kategorie einer neuen Marketing-Umfrage entsprechen, die $I - 1$ alten Alternativen gegenübergestellt wird.

Effektkodierung

Für eine Einflussgröße A mit I Kategorien lautet die Effektkodierung

$$x_{A(i)} = \begin{cases} 1 & A = i, \quad i = 1, \dots, I - 1 \\ -1 & A = I \\ 0 & \text{sonst.} \end{cases}$$

Damit gilt

$$\beta_I = - \sum_{i=1}^{I-1} \beta_i \quad \text{bzw.} \quad \sum_{i=1}^I \beta_i = 0.$$

Als Einflussgröße wird nun der Vektor $(x_{A(1)}, \dots, x_{A(I-1)})$ betrachtet. Die letzte Kategorie $A = I$ wird implizit durch den Vektor $(-1, \dots, -1)$ erfasst. Nach Einsetzen der verschiedenen Einflussgrößen A erhält man die Parameter des Regressionsmodells in Effektkodierung

$$\begin{aligned} \beta_0 &= \frac{1}{I} \sum_{i=1}^I \mu_i \\ \beta_i &= \mu_i - \beta_0, \quad i = 1, \dots, I. \end{aligned}$$

β_i kann als die Abweichung zwischen dem Erwartungswert in der i -ten Faktorstufe und dem mittleren Niveau β_0 interpretiert werden.

Modellierung von Interaktionen

Im Allgemeinen sind mehrere Einflussgrößen von Interesse. Insbesondere ist eine Mischung aus kategorialer und metrischer Variable, wie Geschlecht und Alter denkbar. Als Einfluss-term zum Vektor (x_1, \dots, x_p) wird

$$\eta(\mathbf{x}) = \mathbf{x}^T \boldsymbol{\beta} = \beta_1 x_1 + \dots + \beta_p x_p$$

betrachtet, wobei die x_i metrische Variablen oder Dummy-Variablen zu einem kategorialen Merkmal sind. Die x_i können auch Produkte von Variablen sein. In diesem Fall sind die Koeffizienten sogenannte *Interaktionseffekte*, die das Zusammenwirken zweier oder mehrerer Variablen beschreiben. Die Wahrscheinlichkeit bei der Wahl eines Verkehrsmittels das Auto zu benutzen hängt unter anderem von der Verfügbarkeit eines Pkw ab. Es ist beispielsweise von Interesse, zu untersuchen, ob die Variation der Verfügbarkeit eines Pkw für

verschiedene Haushaltstypen (Single, Ehepaare, usw.) identisches Verkehrsmittelwahlverhalten zur Folge hat. Neben der Einflussgröße Pkw-Verfügung wird die mehrkategoriale Größe Haushaltstyp als weitere Kovariable betrachtet. In Abschnitt 3.3 und in Abschnitt 3.6.2 wird diese Thematik anhand von Beispielen noch einmal ausführlich diskutiert.

Als Komponenten in \mathbf{x} sind möglich:

x_i, x_i^2, x_i^3, \dots	Haupteffekte und Potenzen zur metrischen Variable x_i .
$x_{A(i)}, \dots, x_{A(I-1)}$	Dummy-Variablen zur kategorialen Variable $A \in \{1, \dots, I\}$.
$x_{A(i)} \cdot x_{B(j)},$ $i = 1, \dots, I - 1,$ $j = 1, \dots, J - 1$	Zwei-Faktor-Interaktionen (Produkte von Dummy-Variablen) zu den kategorialen Variablen $A \in \{1, \dots, I\}$ und $B \in \{1, \dots, J\}$.
$x_i x_{B(j)}, j = 1, \dots, J$	Interaktion zwischen metrischer Variable x_i und kategorialem Merkmal $B \in \{1, \dots, J\}$.

2.2 Gruppierte und ungruppierte Daten

Die beobachteten Daten können einerseits in gruppierter andererseits in ungruppiert Form dargestellt werden. Gruppierte Form heißt, dass die einzelnen Beobachtungen identische Werte aufweisen. Die Daten können als Matrix mit einer Spalte für jede Kovariable und k Spalten, die die Häufigkeiten der Responsekategorien enthalten, dargestellt werden.

Beispiel 2.2.1. (gruppierte bzw. ungruppierte Daten)

In vielen Fällen besitzt die Einflussgröße \mathbf{x} nur wenige mögliche Ausprägungen. Dann ist es sinnvoll alle Beobachtungen mit gleichem Einflussgrößenvektor zusammenzufassen.

Geschlecht	Führerschein	Alter				
		18+ bis 31	31+ bis 42	42+ bis 55	55+ bis 65	> 65
männlich	ja	283	355	307	158	41
	nein	6	12	10	12	15
weiblich	ja	281	327	267	143	93
	nein	21	28	45	107	20

Tabelle 1: Daten aus der Mobilitätsstudie '97

Die Kovariablen werden wie folgt kodiert:

Geschlecht (G) 1 = männlich, 0 = weiblich

Alter (A) 1 = 18 + bis 31, 2 = 31 + bis 42, 3 = 42 + bis 55,
4 = 55 + bis 65, 5 = > 65

Die Daten in gruppierter Form lauten dann:

G	A	ja	nein
1	1	283	6
1	2	355	12
1	3	307	10
1	4	158	12
1	5	41	15
2	1	281	21
2	2	327	28
2	3	267	45
2	4	143	107
2	5	93	20

Bei ungruppierten Daten wird jede Beobachtung einzeln aufgeführt. Die Responsevariable wird wie folgt kodiert:

Führerschein (F) 1 = ja, 2 = nein

Damit ergeben sich die Daten aus Tabelle 1 in ungruppiertes Form als

G	A	F
1	1	1
1	1	1
1	1	1
	⋮	
1	1	1
1	1	2
	⋮	
1	1	2
	⋮	
2	5	2

Die meisten Statistik-Softwarepakete akzeptieren beide Formen.

2.3 Prinzip der Maximum-Likelihood-Schätzung

Definition 2.3.1. Es sei $\mathbf{y} = (y_1, \dots, y_n)^T$ eine Zufallsvariable mit der Dichtefunktion $f(\mathbf{y}; \boldsymbol{\theta})$, wobei der Parametervektor $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)^T$ in dem Parameterraum Ω der a-priori zulässigen Parameterwerte $\boldsymbol{\theta}$ liegt. Dann definiert die Dichtefunktion $f(\mathbf{y}; \boldsymbol{\theta})$ für jede Realisierung y_0 von \mathbf{y} eine Funktion von $\boldsymbol{\theta}$, d.h.

$$L(\boldsymbol{\theta}) = L(\theta_1, \dots, \theta_m) = f(y_0; \boldsymbol{\theta}),$$

die als Likelihoodfunktion von y_0 bezeichnet wird.

Falls $\hat{\boldsymbol{\theta}}$ existiert, wählt das Maximum-Likelihood-Prinzip als Schätzung von $\boldsymbol{\theta}$ in Abhängigkeit von y_0 den Wert $\hat{\boldsymbol{\theta}} \in \Omega$, für den

$$L(\hat{\boldsymbol{\theta}}) \geq L(\boldsymbol{\theta}) \quad \text{für alle } \boldsymbol{\theta} \in \Omega$$

gilt. Dabei braucht $\hat{\boldsymbol{\theta}}$ nicht eindeutig zu sein. Der Schätzwert $\hat{\boldsymbol{\theta}}$ ist dann so gewählt, dass die Realisierung y_0 im stetigen Fall den dichtesten oder im diskreten Fall wahrscheinlichsten Wert der Verteilung von \mathbf{y} darstellt. Wird die Maximierung von $L(\boldsymbol{\theta})$ für alle Realisierungen durchgeführt, so ist $\hat{\boldsymbol{\theta}}$ eine Funktion von \mathbf{y} und damit selbst eine Zufallsvariable, die als Maximum-Likelihood-Schätzung von $\boldsymbol{\theta}$ bezeichnet wird.

Im Folgenden soll die Maximum-Likelihood-Schätzung für den Fall unabhängiger Beobachtungen eingeführt werden. Seien dazu y_1, \dots, y_n unabhängige Wiederholungen einer Zufallsvariablen so, dass y_i einem Verteilungstyp mit Parameter $\boldsymbol{\theta}$ folgt, d.h. y_i besitzt im stetigen Fall eine Dichte bzw. im diskreten Fall eine Wahrscheinlichkeitsfunktion $f(y_i; \boldsymbol{\theta})$ mit bekannter Funktion f . Wegen der Unabhängigkeit der y_1, \dots, y_n lautet die gemeinsame Dichte

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n f(y_i; \boldsymbol{\theta}).$$

Die Maximum-Likelihood-Schätzung besteht darin die Likelihoodfunktion zu maximieren. Gesucht ist der Wert $\hat{\boldsymbol{\theta}}$, für den die Dichte bzw. die Auftretenswahrscheinlichkeit für die tatsächlich beobachteten Werte y_1, \dots, y_n maximal ist. Oft ist es einfacher die Log-Likelihoodfunktion $l(\boldsymbol{\theta}) = \log(L(\boldsymbol{\theta}))$ zu maximieren. Da der Logarithmus eine monoton wachsende Funktion ist, liegen die Maxima der Funktion $L(\boldsymbol{\theta})$ und $\log(L(\boldsymbol{\theta}))$ an der selben Stelle $\hat{\boldsymbol{\theta}}$. Die Log-Likelihoodfunktion lässt sich darstellen durch

$$l(\boldsymbol{\theta}) = \log(L(\boldsymbol{\theta})) = \log \prod_{i=1}^n f(y_i; \boldsymbol{\theta}) = \sum_{i=1}^n \log(f(y_i; \boldsymbol{\theta})).$$

Setzt man die erste Ableitung nach $\boldsymbol{\theta}$ der Log-Likelihoodfunktion gleich Null, erhält man eine Schätzgleichung. Ob es sich tatsächlich um ein Maximum handelt, kann mit Hilfe der zweiten Ableitung untersucht werden. Es muss geprüft werden ob die zweite Ableitung nach $\boldsymbol{\theta}$ eine negativ definite Matrix ist, d.h.

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}$$

sollte negativ definit sein. Die zweite Ableitung gibt die Krümmung wieder. Je stärker die Krümmung bei $\hat{\boldsymbol{\theta}}$, desto kleiner ist die asymptotische Varianz des Schätzers. Die Maximum-Likelihood-Schätzung wird ausführlich zum Beispiel in Bickel und Doksum (1977, Kapitel 3.3) [5] beschrieben.

2.4 Das generalisierte lineare Modell (GLM)

In diesem Abschnitt wird das generalisierte lineare Modell vorgestellt. Es dient als Grundlage für Modelle mit zufälligen Effekten, die in Kapitel 4 dargestellt werden.

Generalisierte lineare Modelle stellen einen sehr allgemeinen regressionsanalytischen Modelltyp dar. Zurückgehend auf Nelder und Wedderburn (1972) [24] wurde das Konzept in einer Vielzahl von Veröffentlichungen auf seine Möglichkeiten hin untersucht und weiterentwickelt. Eine ausführliche Darstellung in die Theorie der generalisierten linearen Modelle geben McCullagh und Nelder (1989) [21], worauf auch folgender Abschnitt basiert. Untersuchungen zur asymptotischen Theorie finden sich insbesondere bei Fahrmeir und Kaufmann (1986, 1987) [13] und [14].

Die Schätzung für binäre (dichotome) und Poisson-verteilte abhängige Variablen lässt sich im Rahmen der univariaten GLMs vereinheitlichen. Generalisierte lineare Modelle spezifizieren den Zusammenhang zwischen den k Beobachtungen der abhängigen Variable y und den Regressoren $\mathbf{x} = (x_1, \dots, x_p)^T$, welche aus den ursprünglichen Kovariablen abgeleitet wurden, durch drei Komponenten, vgl. McCullagh und Nelder (1989) [21]. Hierbei beschreibt die stochastische Komponente die Verteilungsfamilie aus der die Beobachtungen y_i , $i = 1, \dots, k$ gegeben \mathbf{x}_i generiert wurden. Ausgangspunkt für die Betrachtung der GLM in diesem Abschnitt sind klassische lineare Regressionsmodelle der Form

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \dots, k,$$

wobei $\boldsymbol{\beta}$ für einen unbekanntem Parametervektor steht. ε_i beschreibt die Störterme. Es wird angenommen, dass $\varepsilon_i \sim N(0, \sigma^2)$. Für die Verteilung von y_i gilt dann $y_i \sim N(\mu_i, \sigma^2)$. Die systematische Komponente stellt den linearen Prädiktor η_i mit

$$\eta_i = \beta_1 x_{i1} + \dots + \beta_p x_{ip}$$

in einen linearen Zusammenhang zu den Regressoren \mathbf{x}_i . Sie wird mit dem bedingten Erwartungswert $\mu_i = \mathbb{E}(y_i | \mathbf{x}_i)$ durch eine monotone Linkfunktion $g = h^{-1}$ verknüpft. Es gilt

$$\eta_i = h^{-1}(\mu_i) = g(\mu_i) \quad \text{bzw.} \quad \mu_i = h(\eta_i) = h(\mathbf{x}_i^T \boldsymbol{\beta}),$$

wobei die Responsefunktion $h(\eta)$ zweimal stetig differenzierbar ist und \mathbb{R} auf den Wertebereich von μ abbildet.

In der klassischen Definition generalisierter linearer Modelle wird die Verteilung der Responsevariable y gegeben x als einfache Exponentialfamilie der Form

$$f(y | \theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\} \quad (2.2)$$

mit spezifischen Funktionen $a(\cdot)$, $b(\cdot)$ und $c(\cdot)$ angenommen (univariate Beispiele sind die Normal-, Binomial-, Poisson-, Exponential- und Gamma-Verteilung, multivariate Beispiele die Multinormal- und Multinomial-Verteilung). Hierbei bezeichnet ϕ den Dispersionsparameter und $\theta = \theta(\mu)$ den natürlichen Parameter, der über $\mu = \partial b(\theta) / \partial \theta$ den bedingten

Erwartungswert der Responsevariable y beschreibt. Die bedingte Varianz der abhängigen Variable steht über

$$\text{Var}(y | x) = \frac{\partial^2 b(\theta)}{\partial \theta^2} a(\phi)$$

mit dem Dispersionsparameter ϕ in Zusammenhang. Im Gegensatz zu linearen Modellen wird bei GLMs die Varianz der Response-Variable mit deren Erwartungswert über die *Varianzfunktion*

$$v(\mu) := \frac{\partial^2 b(\theta)}{\partial \theta^2} = \frac{\partial^2 b(\theta(\mu))}{\partial \theta^2(\mu)} \quad (2.3)$$

verknüpft. Die Funktion $a(\phi)$ hat in der Regel die Form $a(\phi) = \phi/\omega_i$, wobei ω_i ein festes mit der Beobachtung i assoziiertes Gewicht ist. Während die Funktion $c(\cdot)$ der Normalisierung dient, ist $b(\cdot)$ für die Charakterisierung der Verteilungseigenschaften von großer Bedeutung. Ausgehend von der Exponentialfunktion erhält man die Log-Likelihoodfunktion

$$l(\theta, \phi | y) = \log(f(y | \theta, \phi)) \quad (2.4)$$

als Funktion von θ und ϕ bei gegebenen y . Der Erwartungswert und die Varianz von y können aus

$$\mathbb{E} \left(\frac{\partial l}{\partial \theta} \right) = 0 \quad (2.5)$$

und

$$\mathbb{E} \left(\frac{\partial^2 l}{\partial \theta^2} \right) + \mathbb{E} \left(\frac{\partial l}{\partial \theta} \right)^2 = 0 \quad (2.6)$$

hergeleitet werden. Gleichung (2.5) und (2.6) erfordern die Regularitätsbedingungen (Casella und Berger (1990) [8], Seite 308). Der Einfachheit halber wird im Folgenden b' für die erste Ableitung und b'' für die zweite Ableitung nach der entsprechenden Variable verwendet. Aus (2.2) bzw. (2.4) ergibt sich

$$l(\theta, \phi | y) = \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)$$

mit den Ableitungen

$$\frac{\partial l}{\partial \theta} = \frac{y - b'(\theta)}{a(\phi)} \quad (2.7)$$

und

$$\frac{\partial^2 l}{\partial \theta^2} = -\frac{b''(\theta)}{a(\phi)}. \quad (2.8)$$

Aus (2.5) und (2.7) folgt

$$0 = \mathbb{E} \left(\frac{\partial l}{\partial \theta} \right) = \frac{\mu - b'(\theta)}{a(\phi)}$$

und damit

$$\mathbb{E}(y) = \mu = b'(\theta). \quad (2.9)$$

Analog ergibt sich aus (2.6), (2.7) und (2.8)

$$\begin{aligned} 0 &= \mathbb{E}\left(-\frac{b''(\theta)}{a(\phi)}\right) + \mathbb{E}\left(\frac{y - b'(\theta)}{a(\phi)}\right)^2 = -\frac{b''(\theta)}{a(\phi)} + \frac{\mathbb{E}(y - \mathbb{E}(y))}{a^2(\phi)} \\ &= -\frac{b''(\theta)}{a(\phi)} + \frac{\text{Var}(y)}{a^2(\phi)}, \end{aligned}$$

so dass

$$\text{Var}(y) = b''(\theta)a(\phi) \stackrel{(2.3)}{=} v(\mu)a(\phi). \quad (2.10)$$

Folglich ist die Varianz von y das Produkt zweier Funktionen. Die erste Funktion, $b''(\theta)$, hängt nur vom natürlichen Parameter θ und folglich vom Erwartungswert ab, während die zweite Funktion $a(\phi)$ unabhängig von θ ist und nur von ϕ abhängt.

Beispiel 2.4.1. (*Poisson-Verteilung*)

Für die Poisson-Verteilung $\text{Poi}(\lambda)$ gilt

$$f(y) = \frac{\lambda^y}{y!} e^{-\lambda} = \exp\{y \log(\lambda) - \lambda - \log(y!)\}$$

mit

$$\begin{aligned} \theta &= \log(\lambda), \quad b(\theta) = \lambda = \exp(\theta), \\ \phi &= 1 \text{ sowie } a(\phi) = 1 \text{ und } c(y, \phi) = -\log(y!). \end{aligned}$$

Die Poisson-Verteilung gehört folglich zur exponentiellen Familie. Für den Erwartungswert und die Varianz gilt

$$\mathbb{E}(y) = \mu(\theta) = b'(\theta) = \exp(\theta) = \lambda,$$

$$\text{Var}(y) = \sigma^2(\theta) = b''(\theta)a(\phi) = \exp(\theta) = \lambda.$$

Beispiel 2.4.2. (*Binomialverteilung*)

Für die Binomialverteilung $B(n, \pi)$ gilt

$$\begin{aligned} f(y) &= \binom{n}{y} \pi^y (1 - \pi)^{n-y} = \exp\{y \log(\pi) + (n - y) \log(1 - \pi) + \log \binom{n}{y}\} \\ &= \exp\left\{y \log\left(\frac{\pi}{1 - \pi}\right) + n \log(1 - \pi) + \log \binom{n}{y}\right\} \end{aligned}$$

mit

$$\theta = \log\left(\frac{\pi}{1-\pi}\right), \quad \pi = \frac{\exp(\theta)}{1+\exp(\theta)},$$

$$b(\theta) = -n \log(1-\pi) = n \log(1+\exp(\theta)),$$

$$\phi = 1 \text{ sowie } a(\phi) = 1 \text{ und } c(\mathbf{y}, \phi) = \log\binom{n}{\mathbf{y}}.$$

Auch die Binomialverteilung gehört somit zur exponentiellen Familie. Für den Erwartungswert und die Varianz gilt

$$\mathbb{E}(y) = \mu(\theta) = b'(\theta) = n \frac{\exp(\theta)}{1+\exp(\theta)} = n\pi,$$

$$\text{Var}(y) = \sigma^2(\theta) = b''(\theta)a(\phi) = n \frac{\exp(\theta)}{(1+\exp(\theta))^2} = n\pi(1-\pi).$$

Das multinomiale Logit-Modell kann als Spezialfall mehrkategorialer generalisierter linearer Modelle aufgefasst werden.

Beispiel 2.4.3. (Multinomialverteilung)

Für die Multinomialverteilung $M(n, \boldsymbol{\pi})$ gilt

$$\begin{aligned} f(y_1, \dots, y_k) &= \frac{n!}{y_1! \cdots y_k! (n - y_1 - \cdots - y_k)!} \pi_1^{y_1} \cdots \pi_k^{y_k} (1 - \pi_1 - \cdots - \pi_k)^{n - y_1 - \cdots - y_k} \\ &= \exp \left\{ \log \left(\frac{n!}{y_1! \cdots y_k! (n - y_1 - \cdots - y_k)!} \right) + y_1 \log \left(\frac{\pi_1}{1 - \pi_1 - \cdots - \pi_k} \right) \right. \\ &\quad \left. + \cdots + y_k \log \left(\frac{\pi_k}{1 - \pi_1 - \cdots - \pi_k} \right) + n \log(1 - \pi_1 - \cdots - \pi_k) \right\} \end{aligned}$$

mit

$$\boldsymbol{\theta} = \left[\log \left(\frac{\pi_1}{1 - \pi_1 - \cdots - \pi_k} \right), \dots, \log \left(\frac{\pi_k}{1 - \pi_1 - \cdots - \pi_k} \right) \right]^T,$$

$$\boldsymbol{\pi} = \left(\frac{\exp(\theta_1)}{1 + \sum_{r=1}^k \exp(\theta_r)}, \dots, \frac{\exp(\theta_k)}{1 + \sum_{r=1}^k \exp(\theta_r)} \right)^T$$

$$b(\boldsymbol{\theta}) = -n \log(1 - \pi_1 - \cdots - \pi_k) = n \log(1 + \exp(\theta_1) + \cdots + \exp(\theta_k)),$$

$$\phi = 1, \quad a(\phi) = 1 \quad \text{und} \quad c(\mathbf{y}, \phi) = \log \left(\frac{n!}{y_1! \cdots y_k! (n - y_1 - \cdots - y_k)!} \right).$$

Für den Erwartungswert von \mathbf{y} gilt

$$\mathbb{E}(\mathbf{y}) = b'(\boldsymbol{\theta}) = n \left(\frac{\exp(\theta_1)}{1 + \sum_{r=1}^k \exp(\theta_r)}, \dots, \frac{\exp(\theta_k)}{1 + \sum_{r=1}^k \exp(\theta_r)} \right)^T = n\boldsymbol{\pi}$$

und für die Varianz

$$\begin{aligned}
\text{Var}(\mathbf{y}) &= b''(\boldsymbol{\theta})a(\phi) \\
&= n \left(\frac{\exp(\theta_1)(1 + \sum_{r=1}^k \exp(\theta_r)) - (\exp(\theta_1))^2}{(1 + \sum_{r=1}^k \exp(\theta_r))^2}, \dots, \right. \\
&\quad \left. \frac{\exp(\theta_k)(1 + \sum_{r=1}^k \exp(\theta_r)) - (\exp(\theta_k))^2}{(1 + \sum_{r=1}^k \exp(\theta_r))^2} \right)^T \\
&= n \left(\frac{\exp(\theta_1)}{1 + \sum_{r=1}^k \exp(\theta_r)} \left(1 - \frac{\exp(\theta_1)}{1 + \sum_{r=1}^k \exp(\theta_r)} \right), \dots, \right. \\
&\quad \left. \frac{\exp(\theta_k)}{1 + \sum_{r=1}^k \exp(\theta_r)} \left(1 - \frac{\exp(\theta_k)}{1 + \sum_{r=1}^k \exp(\theta_r)} \right) \right)^T \\
&= n(\pi_1(1 - \pi_1), \dots, \pi_k(1 - \pi_k))^T = n\boldsymbol{\pi}(1 - \boldsymbol{\pi}).
\end{aligned}$$

2.4.1 Maximum-Likelihood-Schätzung in GLMs

Im klassischen GLM wird die Responsefunktion h als bekannt vorausgesetzt. Die Schätzung des unbekannt Parametervektors $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ erfolgt in der Regel durch das Maximum-Likelihood Prinzip. Bezeichne $l(\boldsymbol{\beta}) = \sum_i l_i(\boldsymbol{\beta})$ die Log-Likelihood der Beobachtungen, so erfüllt ein existierender Maximum-Likelihood-Schätzer $\hat{\boldsymbol{\beta}}$ notwendigerweise die Schätzgleichungen

$$s(\hat{\boldsymbol{\beta}}) = \frac{\partial l(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^k \frac{\partial l_i(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^k s_i(\hat{\boldsymbol{\beta}}) = 0,$$

wobei $s(\boldsymbol{\beta})$ als Scorefunktion bzw. Scorevektor bezeichnet wird. Die individuellen Scorebeiträge der i -ten Beobachtung sind durch

$$s_i(\boldsymbol{\beta}) = \frac{\partial l_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{\partial l_i(\boldsymbol{\beta})}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \boldsymbol{\beta}}$$

gegeben. Da $b'(\theta_i) = \mu_i$ und $b''(\theta_i) = v(\mu_i)$ gilt, dass $\partial \mu_i / \partial \theta_i = b''(\theta_i) = v(\mu_i)$. Wegen $\eta_i = g(\mu_i)$ folgt $\partial \eta_i / \partial \mu_i = g'(\mu_i)$ und aus $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$ erhält man $\partial \eta_i / \partial \boldsymbol{\beta} = \mathbf{x}_i^T$. Damit folgt

$$\begin{aligned}
\frac{\partial l_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \frac{y_i - b'(\theta_i)}{a(\phi)} \cdot \frac{1}{b''(\theta_i)} \cdot \frac{1}{g'(\mu_i)} \cdot \mathbf{x}_i^T \\
&= \frac{y_i - \mu_i}{a(\phi)g'(\mu_i)v(\mu_i)} \mathbf{x}_i^T,
\end{aligned}$$

wobei die Eigenschaft der Matrixinversion $(\partial \eta_i / \partial \mu_i) = (\partial \mu_i / \partial \eta_i)^{-1}$ ausgenutzt wurde. Mit den bekannten Funktionen kann die Scorefunktion daher durch

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{1}{a(\phi)} \sum_{i=1}^k (y_i - \mu_i) \omega_i g'(\mu_i) \mathbf{x}_i^T$$

ausgedrückt werden, wobei $\omega_i = \{v(\mu_i)[g'(\mu_i)]^2\}^{-1}$.

In Matrixnotation ergibt sich folgende Darstellung

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{1}{a(\phi)} \mathbf{X}^T \mathbf{W} \boldsymbol{\Delta} (\mathbf{y} - \boldsymbol{\mu}), \quad (2.11)$$

wobei \mathbf{W} die $k \times k$ -Diagonalmatrix mit den Diagonaleinträgen ω_i ist, die gerade den iterierten Gewichten des GLM entsprechen. $\boldsymbol{\Delta}$ ist die $k \times k$ -Diagonalmatrix mit den Einträgen $g'(\mu_i)$. Die Maximum-Likelihoodgleichungen sind folglich gegeben durch

$$\mathbf{X}^T \mathbf{W} \boldsymbol{\Delta} \mathbf{y} = \mathbf{X}^T \mathbf{W} \boldsymbol{\Delta} \boldsymbol{\mu},$$

wobei \mathbf{W} , $\boldsymbol{\Delta}$ und $\boldsymbol{\mu}$ den unbekannt Parametervektor $\boldsymbol{\beta}$ beschreiben. Zum Lösen der Gleichungen wird häufig auf iterative Verfahren, wie sie auch bei den Multinomialen Modellen angewandt werden, zurückgegriffen.

Unter den Regularitätsbedingungen (vgl. Fahrmeir und Tutz (2001) [15], Seite 46 f.) gelten für den Maximum-Likelihood-Schätzer $\hat{\boldsymbol{\beta}}_n$ von $\boldsymbol{\beta}$ mit Stichprobengröße n folgende Eigenschaften

- Die Wahrscheinlichkeit, dass $\hat{\boldsymbol{\beta}}_n$ existiert und (lokal) eindeutig ist konvergiert gegen 1 für $n \rightarrow \infty$ (asymptotische Existenz und Eindeutigkeit).
- Beschreibt $\boldsymbol{\beta}$ den wahren Wert, so gilt für $n \rightarrow \infty$:
 $\hat{\boldsymbol{\beta}}_n \rightarrow \boldsymbol{\beta}$ in Wahrscheinlichkeit (schwache Konsistenz) oder
 $\hat{\boldsymbol{\beta}}_n \rightarrow \boldsymbol{\beta}$ mit Wahrscheinlichkeit 1 (starke Konsistenz) .
- Für $n \rightarrow \infty$ gilt die Verteilungapproximation $\hat{\boldsymbol{\beta}}_n \stackrel{a}{\sim} N(\boldsymbol{\beta}, \mathbf{F}(\hat{\boldsymbol{\beta}}_n)^{-1})$ mit der Fisher-Matrix $\mathbf{F}(\boldsymbol{\beta}) = \mathbb{E}(-\partial^2 l(\boldsymbol{\beta}) / \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T)$.

Lineare Hypothesen

Für den Parametervektor $\boldsymbol{\beta}$ eines GLM sei die Nullhypothese gegeben durch

$$H_0 : \mathbf{C} \boldsymbol{\beta} = \boldsymbol{\zeta},$$

zur Alternativhypothese

$$H_1 : \mathbf{C} \boldsymbol{\beta} \neq \boldsymbol{\zeta},$$

wobei \mathbf{C} vollen Rang besitzt.

$\tilde{\boldsymbol{\beta}}$ bezeichnet den Maximum-Likelihood-Schätzer unter der Restriktion $\mathbf{C} \boldsymbol{\beta} = \boldsymbol{\zeta}$ und $\hat{\boldsymbol{\beta}}$ den Maximum-Likelihood-Schätzer ohne diese Einschränkung. Als Teststatistik für H_0 verwendet man die Likelihood-Quotienten-Statistik

$$\lambda = -2\{l(\tilde{\boldsymbol{\beta}}) - l(\hat{\boldsymbol{\beta}})\}$$

oder die Wald-Statistik

$$\omega = (\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\zeta})^T [\mathbf{C}\mathbf{F}^{-1}(\hat{\boldsymbol{\beta}})\mathbf{C}^T]^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\zeta}).$$

Tritt die Nullhypothese H_0 ein, so ergibt sich asymptotisch eine χ^2 -Verteilung, deren Freiheitsgrade durch den Rang der Matrix \mathbf{C} ($\text{rg}(\mathbf{C})$) bestimmt sind, d.h.

$$\lambda, \omega \sim \chi_{\text{rg}(\mathbf{C})}^2.$$

H_0 wird zum Niveau α verworfen, falls das $(1 - \alpha)$ -Quantil der $\chi_{\text{rg}(\mathbf{C}), 1-\alpha}^2$ überschritten ist. Als Anpassungstests dienen die *Pearson-Statistik*

$$\chi_P^2 = \sum_{i=1}^k \frac{(y_i - \hat{\mu}_i)^2}{v(\hat{\mu}_i)},$$

wobei $v(\hat{\mu}_i)$ die geschätzte Varianzfunktion beschreibt und die *Devianz*

$$D = -2a(\phi) \sum_{i=1}^k [l_i(\hat{\mu}_i) - l_i(y_i)].$$

2.4.2 Maximum-Quasi-Likelihood

Nicht bei allen statistischen Untersuchungen ist die Verteilung der Daten einfach zu modellieren. Ist die Verteilung nicht bekannt, können der Likelihood nicht konstruiert werden und folglich bekannte Methoden, wie der Likelihood-Ratio-Test (vgl. zum Beispiel McCulloch und Searle [22], S. 24 oder S. 147 f.) nicht durchgeführt werden. Daher ist es sinnvoll nach Methoden zu suchen, die wie die Maximum-Likelihood gut funktionieren aber ohne spezifische Verteilungsannahmen auskommen. Die Idee der Quasi-Likelihoodfunktion besteht darin eine ähnliche Größe zum Likelihood zu konstruieren die mit weniger Annahmen auskommt.

Falls $\mathbb{E} \left(\frac{\partial}{\partial \theta_i} \log(f(y_i | \theta_i, \phi)) \right) = 0$ ist folgt

$$\begin{aligned} \text{Var} \left(\frac{\partial}{\partial \theta_i} \log(f(y_i | \theta_i, \phi)) \right) &\stackrel{(2.5)}{=} \mathbb{E} \left[\left(\frac{\partial}{\partial \theta_i} \log(f(y_i | \theta_i, \phi)) \right)^2 \right] \\ &\stackrel{(2.6)}{=} -\mathbb{E} \left(\frac{\partial^2}{\partial \theta_i^2} \log(f(y_i | \theta_i, \phi)) \right). \end{aligned} \quad (2.12)$$

Nach (2.5) und (2.12), wobei nachfolgend nach $\mu = \mathbb{E}(y)$ statt nach θ abgeleitet wird, soll gelten

$$\mathbb{E} \left[\frac{\partial}{\partial \mu_i} \log(f(y_i | \theta_i, \phi)) \right] = \mathbb{E} \left[\frac{\partial}{\partial \theta_i} \log(f(y_i | \theta_i, \phi)) \right] \frac{d\theta_i}{d\mu_i} = 0 \quad (2.13)$$

und

$$\begin{aligned}
v^* &:= \text{Var} \left[\frac{\partial}{\partial \mu_i} \log(f(y_i | \theta_i, \phi)) \right] = \text{Var} \left[\frac{\partial}{\partial \theta_i} \log(f(y_i | \theta_i, \phi)) \cdot \frac{d\theta_i}{d\mu_i} \right] \\
&= \left[\text{Var} \left(\frac{\partial}{\partial \theta_i} \log(f(y_i | \theta_i, \phi)) \right) \right] \left(\frac{d\theta_i}{d\mu_i} \right)^2 \\
&\stackrel{(2.12)}{=} \left[-\mathbb{E} \left(\frac{\partial^2}{\partial \theta_i^2} \log(f(y_i | \theta_i, \phi)) \right) \right] \left(\frac{d\theta_i}{d\mu_i} \right)^2 \\
&\stackrel{(2.8)}{=} \frac{b''(\theta_i)}{a(\phi)} \left(\frac{1}{v(\mu_i)} \right)^2 \stackrel{(2.3)}{=} \frac{1}{a(\phi)v(\mu_i)}. \tag{2.14}
\end{aligned}$$

Im Folgenden wird eine Größe gesucht, die $\partial \log(f(y_i | \theta_i, \phi)) / \partial \mu_i$ ersetzt und die Eigenschaften (2.13) und (2.14) erfüllt. Man kann leicht zeigen, dass

$$q(\mu_i, y_i) = \frac{y_i - \mu_i}{a(\phi)v(\mu_i)} \tag{2.15}$$

beiden Bedingungen genügt, wobei angenommen wird, dass $\text{Var}(y_i) \propto v(\mu_i)$. Für den Erwartungswert von $q(\mu_i, y_i)$ gilt

$$\begin{aligned}
\mathbb{E}(q(\mu_i, y_i)) &= \mathbb{E} \left(\frac{y_i - \mu_i}{a(\phi)v(\mu_i)} \right) = \frac{1}{a(\phi)v(\mu_i)} \mathbb{E}(y_i - \mu_i) = \frac{1}{a(\phi)v(\mu_i)} (\mathbb{E}(y_i) - \mu_i) \\
&= \frac{1}{a(\phi)v(\mu_i)} (\mu_i - \mu_i) = 0
\end{aligned}$$

und für die Varianz

$$\text{Var}(q(\mu_i, y_i)) = \text{Var} \left(\frac{y_i - \mu_i}{a(\phi)v(\mu_i)} \right) = \frac{1}{(a(\phi)v(\mu_i))^2} \text{Var}(y_i) \stackrel{(2.10)}{=} \frac{a(\phi)v(\mu_i)}{(a(\phi)v(\mu_i))^2} = \frac{1}{a(\phi)v(\mu_i)}.$$

Falls der Beitrag von y_i zur Log-Likelihood das Integral bezüglich μ_i von $\partial \log(f(y_i | \theta_i, \phi)) / \partial \mu_i$ ist, dann kann die Log-Quasi-Likelihoodfunktion über y_i wie folgt definiert werden

$$Q(\mu_i, y_i) = \int_{y_i}^{\mu_i} \frac{y_i - t}{a(\phi)v(t)} dt. \tag{2.16}$$

Die Ableitung nach μ_i entspricht nach Definition gerade $q(\mu_i, y_i)$. Die Schätzung des unbekannten Parametervektors β erfolgt durch Lösen der Maximum-Quasi-Likelihoodgleichungen

$$\frac{\partial}{\partial \beta} \sum_{i=1}^k Q(\mu_i, y_i) = \sum_{i=1}^k \frac{\partial Q(\mu_i, y_i)}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \beta} = \mathbf{0}. \tag{2.17}$$

Nach dem Ableiten ergibt sich mit $g(\mu_i) = \eta_i = \mathbf{x}_i^T \beta$

$$\frac{\partial \mu_i}{\partial \beta} = \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta} = \frac{\mathbf{x}_i^T}{g'(\mu_i)}$$

und folglich

$$\sum_{i=1}^k \frac{y_i - \mu_i}{a(\phi)v(\mu_i)} \frac{\partial \mu_i}{\partial \beta} = \sum_{i=1}^k \frac{y_i - \mu_i}{a(\phi)v(\mu_i)} \frac{\mathbf{x}_i^T}{g'(\mu_i)} = \mathbf{0}. \tag{2.18}$$

In Matrixdarstellung erhält man

$$\frac{1}{a(\phi)} \mathbf{X}^T \mathbf{W} \Delta(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{0} \quad (2.19)$$

analog zu (2.11). Um den Vorteil der Maximum-Quasi-Likelihood Methode gegenüber der Maximum-Likelihood Methode deutlich hervorzuheben, soll eine Regression mit Poisson-verteilter Response betrachtet werden. Die Poisson-Verteilung impliziert die Gleichheit von Erwartungswert und Varianz. Nur unter der Annahme, dass diese Gleichheit gegeben den Regressoren erfüllt ist, kann das Poissonmodell die zu untersuchenden Daten korrekt beschreiben. In vielen Anwendungen übersteigt dagegen die bedingte Varianz den bedingten Erwartungswert deutlich. Man spricht in diesem Fall von *Überdispersion*. Überdispersion kann viele Ursachen haben, die hier nicht näher erläutert werden. Ist die Varianz proportional zum Erwartungswert, so ist die Beschreibung des Modells unter der Quasi-Likelihood Methode immer noch richtig, denn die einzige Annahme ist, dass $Var(y) = a(\phi)v(\mu)$, d.h. $Var(y) \propto v(\mu)$ aber nicht notwendigerweise gleich $a(\phi)v(\mu)$. Die Maximum-Quasi-Likelihood ist folglich sehr robust. Einerseits werden keine Verteilungsannahmen benötigt, andererseits müssen nur Beziehungen zwischen dem Erwartungswert und der Varianz bis auf eine Konstante $a(\phi)$, die aus den Daten geschätzt werden kann, spezifiziert werden.

3 Multinomiale Modelle

Das multinomiale Logit-Modell ist eines der meist angewandten Methoden bei der Analyse ungeordneter kategorischer Responsevariablen in sozialwissenschaftlichen Studien. Populär geworden ist es, da

1. es eine Verallgemeinerung des Binomialen Logit-Modells darstellt;
2. es äquivalent zum Loglinearen Modell mit gruppierten Daten ist; und
3. statistische Softwarepakete (S-Plus, SAS, SPSS, und viele andere) zum Schätzen des Modells weit verbreitet sind.

Eine Einführung in die Binären Logit-Modelle und Loglinearen Modelle findet sich in Powers und Xie (2000) [25]. Nachfolgende Abschnitte basieren auf Fahrmeir & Tutz (2001, Kapitel 2 und 3) [15] und Tutz (2000, Kapitel 5) [27].

3.1 Modellbildung

In diesem Abschnitt sollen Modelle für den Fall einer abhängigen Variablen mit k Responsekategorien vorgestellt werden. Die interessierenden Responsevariablen sind häufig kategorial. Ein *kategoriales* Merkmal liegt vor, wenn nur endlich viele Ausprägungen (Kategorien) existieren.

Für die Auswahl eines Modells ist es entscheidend, von welchem Typ die abhängige Variable ist. Im Rahmen dieser Arbeit spielen nominale Zielvariablen eine zentrale Rolle. Im Fall einer *Nominalskala* stellen die Zahlenwerte nur Etiketten dar. Bei der Analyse der Verkehrsmittelwahl mit den Kategorien „zu Fuß“, „per Rad“, „mit öffentlichen Verkehrsmitteln“ und „mit dem Pkw“ handelt sich um ungeordnete Kategorien.

Für Regressionen mit ordinaler abhängiger Variable werden andere Modelle benötigt. Als Beispiel seien das kumulative- und das sequentielle Modell erwähnt. Eine Darstellung dieser Modelle findet sich in Fahrmeier & Tutz (2001) Abschnitt 3.3 [15].

Im Folgenden wird die Multinomialverteilung zu Grunde gelegt, die im nächsten Abschnitt eingeführt wird.

3.2 Die Multinomialverteilung

Die Multinomialverteilung lässt sich aus der Wiederholung eines Experiments mit den möglichen (sich ausschließenden) Ausgängen A_1, \dots, A_k ableiten. Mit X_i werden die Anzahl der Fälle, in denen bei n -facher unabhängiger Wiederholung des Experiments das

Ereignis A_i auftritt bezeichnet. Die Wahrscheinlichkeit für A_i bei jeder Wiederholung ist π_i . Der Zufallsvektor $\mathbf{X}^T = (X_1, \dots, X_k)$ heißt *multinomialverteilt* (in Zeichen $\mathbf{X} \sim M(n, \boldsymbol{\pi})$) mit den Parametern $\boldsymbol{\pi}^T = (\pi_1, \dots, \pi_k)$ mit $\pi \in [0, 1]$, $\sum_i \pi_i = 1$ und dem Stichprobenumfang n , wenn die gemeinsame Wahrscheinlichkeitsfunktion von \mathbf{X} gegeben ist durch

$$f(x_1, \dots, x_k) = \begin{cases} \frac{n!}{x_1! \cdots x_k!} \pi_1^{x_1} \cdots \pi_k^{x_k} & x_i \in \{0, \dots, n\}, \sum_i x_i = n \\ 0 & \text{sonst.} \end{cases}$$

Wegen den Restriktionen $\sum_{i=1}^k \pi_i = 1$ und $\sum_{i=1}^k x_i = n$ genügt es $q = k - 1$ Komponenten zu betrachten. Der Vektor (X_1, \dots, X_q) ist multinomialverteilt mit Wahrscheinlichkeitsvektor $\boldsymbol{\pi}^T = (\pi_1, \dots, \pi_q)$. π_k ist implizit durch $\pi_k = 1 - \sum_{i=1}^q \pi_i$ gegeben.

Für $k = 2$ bzw. $q = 1$ ergibt sich die Binomialverteilung.

Eigenschaften der Multinomialverteilung (vgl. McCullagh und Nelder, (1989) Abschnitt 5.3 [21])

$$\mathbb{E}(\mathbf{X}) = n\boldsymbol{\pi},$$

$$\text{Cov}(\mathbf{X}) = n(\text{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}^T) = n \begin{pmatrix} \pi_1(1 - \pi_1) & -\pi_1\pi_2 & \cdots & -\pi_1\pi_k \\ -\pi_2\pi_1 & \ddots & \ddots & \vdots \\ \vdots & & & -\pi_{k-1}\pi_k \\ \dots & & & \pi_k(1 - \pi_k) \end{pmatrix},$$

d.h.

$$\text{Var}(X_i) = n\pi_i(1 - \pi_i),$$

$$\text{Cov}(X_i, X_j) = -n\pi_i\pi_j, \quad i \neq j.$$

3.3 Das multinomiale Logit-Modell

Die Idee multinomialer Logit-Modelle ist, eine abhängige Variable mit k Responsekategorien, d.h. $Y \in \{1, \dots, k\}$, durch mehrere unabhängige Variablen (Kovariablen, Regressoren) x_1, \dots, x_p zu beschreiben. Die Responsekategorie wird häufig auf Nominalskalenniveau gemessen. Für jede Ausprägungskombination erhält man eine Vorhersage für die Wahrscheinlichkeit des Auftretens eines interessierenden Ereignisses. Das multinomiale Logit-Modell kann als eine Erweiterung des binären Logit-Modells aufgefasst werden. Das binäre Logit-Modell für den Fall $Y \in \{1, 2\}$ lautet

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

mit $\mathbf{x}^T \boldsymbol{\beta} = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$ wobei $\boldsymbol{\beta}^T = (\beta_0, \beta_1, \dots, \beta_p)$ den Vektor der Parameter und $\mathbf{x}^T = (1, x_1, \dots, x_p)$ den Vektor der Einflussgrößen bezeichnet und $\sum_{s=1}^k P(Y = s | \mathbf{x}) = 1$. F beschreibt die *logistische Verteilungsfunktion*. Die Größe

$$\eta(\mathbf{x}) := \mathbf{x}^T \boldsymbol{\beta} = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

wird *linearer Prädiktor* genannt. Die Einflussgrößen x_1, \dots, x_p können metrische Variablen sein, einen Interaktionsterm darstellen oder für Dummy-Variablen stehen, die ein kategoriales Merkmal kodieren.

Äquivalente Darstellungen sind

$$\frac{P(Y = 1 | \mathbf{x})}{P(Y = 2 | \mathbf{x})} = \exp(\mathbf{x}^T \boldsymbol{\beta})$$

bzw.

$$\log \left(\frac{P(Y = 1 | \mathbf{x})}{P(Y = 2 | \mathbf{x})} \right) = \mathbf{x}^T \boldsymbol{\beta}. \quad (3.1)$$

Diese beiden Größen werden auch *Chancen* (engl. „odds“, d.h. das Verhältnis aus Auftretenswahrscheinlichkeit und Gegenwahrscheinlichkeit) bzw. *logarithmierte Chancen* (*Logits*) genannt. Sie unterscheiden sich in der abhängigen Größe. Die Parameter der zweiten Darstellung haben in dieser Form dieselbe Funktion wie in der linearen Regression, nur wird als abhängige Größe nicht der Erwartungswert modelliert, sondern die logarithmierten Chancen. Der Vorteil der Verwendung der Logits besteht in der einfacheren Überprüfbarkeit der Linearität, da bei der logarithmierten Chance der Prädiktor auf der linearen Skala liegt.

Im allgemeineren Fall besitzt die abhängige Variable mehr als zwei Ausprägungen. Dieses Modell lässt sich auf den binären Fall zurückführen. Betrachtet werden für k Kategorien das Verhältnis jeweils zweier Kategorien. Wählt man als Referenzkategorie die letzte ($Y = k$), kann das logarithmierte Chancenverhältnis beschrieben werden durch

$$\log \left(\frac{P(Y = r | \mathbf{x})}{P(Y = k | \mathbf{x})} \right) = \mathbf{x}^T \boldsymbol{\beta}_r. \quad (3.2)$$

$\boldsymbol{\beta}_r$ ist jetzt spezifisch für die betrachtete Kategorie r . Analog zum binären Fall stellt das Logit-Modell mit Referenzkategorie r eine Parametrisierung der Chancen bzw. der logarithmierten Chancen dar. Für $r = 1, \dots, q = k - 1$ ist Gleichung (3.2) die einfachste Form des multinomialen Logit-Modells. Aus dieser Darstellung ergibt sich durch einfaches Umformen

$$P(Y = r | \mathbf{x}) = P(Y = k | \mathbf{x}) \exp(\mathbf{x}^T \boldsymbol{\beta}_r), \quad r = 1, \dots, k - 1. \quad (3.3)$$

Daraus folgt

$$P(Y = 1|\mathbf{x}) + \cdots + P(Y = k-1|\mathbf{x}) = P(Y = k|\mathbf{x}) \sum_{s=1}^{k-1} \exp(\mathbf{x}^T \boldsymbol{\beta}_s)$$

bzw.

$$1 = P(Y = 1|\mathbf{x}) + \cdots + P(Y = k|\mathbf{x}) = P(Y = k|\mathbf{x}) \left\{ 1 + \sum_{s=1}^{k-1} \exp(\mathbf{x}^T \boldsymbol{\beta}_s) \right\}.$$

Auflösen nach $P(Y = k|\mathbf{x})$ ergibt

$$P(Y = k|\mathbf{x}) = \frac{1}{1 + \sum_{s=1}^{k-1} \exp(\mathbf{x}^T \boldsymbol{\beta}_s)}. \quad (3.4)$$

Nach Einsetzen in (3.3) erhält man

$$P(Y = r|\mathbf{x}) = \frac{\exp(\mathbf{x}^T \boldsymbol{\beta}_r)}{1 + \sum_{s=1}^{k-1} \exp(\mathbf{x}^T \boldsymbol{\beta}_s)}, \quad r = 1, \dots, k-1. \quad (3.5)$$

Die einfachste Form des multinomialen Logit-Modells mit k Kategorien kann folglich dargestellt werden durch

$$\log \left(\frac{P(Y = r|\mathbf{x})}{P(Y = k|\mathbf{x})} \right) = \mathbf{x}^T \boldsymbol{\beta}_r, \quad r = 1, \dots, k-1 \quad (3.6)$$

bzw.

$$\begin{aligned} P(Y = r|\mathbf{x}) &= \frac{\exp(\mathbf{x}^T \boldsymbol{\beta}_r)}{1 + \sum_{s=1}^{k-1} \exp(\mathbf{x}^T \boldsymbol{\beta}_s)}, \quad r = 1, \dots, k-1 \\ P(Y = k|\mathbf{x}) &= \frac{1}{1 + \sum_{s=1}^{k-1} \exp(\mathbf{x}^T \boldsymbol{\beta}_s)}. \end{aligned} \quad (3.7)$$

Das multinomiale Logit-Modell, ohne Wahl einer bestimmten Referenzkategorie, ist von der grundsätzlichen Form

$$P(Y = r|\mathbf{x}) = \frac{\exp(\mathbf{x}^T \boldsymbol{\beta}_r)}{\sum_{s=1}^k \exp(\mathbf{x}^T \boldsymbol{\beta}_s)}. \quad (3.8)$$

Die Parameter in (3.8) sind offensichtlich nicht identifizierbar, denn addiert man zu jedem Vektor $\boldsymbol{\beta}_r$ einen Vektor \mathbf{c} von Konstanten, so erhält man

$$\begin{aligned} P(Y = r|\mathbf{x}) &= \frac{\exp(\mathbf{x}^T (\boldsymbol{\beta}_r + \mathbf{c}))}{\sum_{s=1}^k \exp(\mathbf{x}^T (\boldsymbol{\beta}_s + \mathbf{c}))} = \frac{\exp(\mathbf{x}^T \boldsymbol{\beta}_r + \mathbf{x}^T \mathbf{c})}{\sum_{s=1}^k \exp(\mathbf{x}^T \boldsymbol{\beta}_s + \mathbf{x}^T \mathbf{c})} \\ &= \frac{\exp(\mathbf{x}^T \boldsymbol{\beta}_r) \exp(\mathbf{x}^T \mathbf{c})}{\exp(\mathbf{x}^T \mathbf{c}) \sum_{s=1}^k \exp(\mathbf{x}^T \boldsymbol{\beta}_s)} = \frac{\exp(\mathbf{x}^T \boldsymbol{\beta}_r)}{\sum_{s=1}^k \exp(\mathbf{x}^T \boldsymbol{\beta}_s)}, \end{aligned}$$

d.h. die Modellform bleibt erhalten. Eine Transformation $\beta_r \rightarrow \beta_r + \mathbf{c}$ ändert (3.8) nicht. Für dieses Modell sind daher Nebenbedingungen erforderlich.

Wählt man als Nebenbedingung $\beta_k = (0, \dots, 0)^T$, entspricht dies einer Auszeichnung der Kategorie k als Referenzkategorie. In diesem Fall kann wegen $\exp(\mathbf{x}^T \beta_k) = \exp(0) = 1$ das Modell dargestellt werden durch

$$P(Y = r | \mathbf{x}) = \frac{\exp(\mathbf{x}^T \beta_r)}{\sum_{s=1}^k \exp(\mathbf{x}^T \beta_s)} = \frac{\exp(\mathbf{x}^T \beta_r)}{1 + \sum_{s=1}^{k-1} \exp(\mathbf{x}^T \beta_s)}, \quad r = 1, \dots, k.$$

Dies entspricht der Form (3.6) bzw. (3.7). Alternativ kann jede beliebige Kategorie als Referenzkategorie gewählt werden. Eine weitere Möglichkeit ergibt sich indem $\sum_{s=1}^k \beta_s^T = (0, \dots, 0)^T$ gesetzt wird.

Es genügt zu den k Kategorien der abhängigen Variable Y $k - 1$ Chancen $\gamma_{r, r_0}(\mathbf{x})$, $r \neq r_0$ bzw. Logits $\text{Logit}_{r, r_0}(\mathbf{x})$, $r \neq r_0$ zu parametrisieren, da für die Chancen für r_0 gegenüber r_0 trivialerweise $\gamma_{r_0, r_0}(\mathbf{x}) = 1$ gilt. Die Chancen werden durch

$$\gamma_{r, r_0}(\mathbf{x}) := \frac{P(Y = r | \mathbf{x})}{P(Y = r_0 | \mathbf{x})} = \exp(\mathbf{x}^T \beta_r), \quad r \neq r_0 \quad (3.9)$$

und die Logits durch

$$\text{Logit}_{r, r_0}(\mathbf{x}) := \log \left(\frac{P(Y = r | \mathbf{x})}{P(Y = r_0 | \mathbf{x})} \right) = \mathbf{x}^T \beta_r, \quad r \neq r_0 \quad (3.10)$$

spezifiziert, wobei r_0 die fest gewählte Referenzkategorie ist.

Interpretation der Parameter

Analog zum binären Modell spielen die Chancen und relativen Chancen (engl. „odds-ratios“) eine wichtige Rolle. Für die Interpretation der Parameter ist es hinreichend, sich auf ein metrisches Merkmal x_M und ein kategoriales Merkmal K zu beschränken. Das kategoriale Merkmal K besteht aus I Responsekategorien. Es werden $I - 1$ Dummy- (Stellvertreter-) Variablen $x_{K(1)}, \dots, x_{K(I-1)}$ benötigt. Die I -te Kategorie wird implizit durch $x_{K(1)} = \dots = x_{K(I-1)} = 0$ erfasst. Die linearen Prädiktoren $\mathbf{x}^T \beta_1, \dots, \mathbf{x}^T \beta_k$ besitzen dann (ohne Interaktionen) die einfache Form

$$\eta_r(x) = \mathbf{x}^T \beta_r = \beta_{0r} + \beta_{M,r} x_M + \beta_{K(1),r} x_{K(1)} + \dots + \beta_{K(I-1),r} x_{K(I-1)}.$$

$\beta_{M,r}$ bezeichnet den Parameter der metrischen Variable, $\beta_{K(i),r}$ den Parameter der kategorialen Variable. Der zusätzliche Index r wird angegeben, da beide Parameter spezifisch für diese Responsekategorie sind. Für die Darstellung (3.9) erhält man die Chancen

$$\frac{P(Y = r | \mathbf{x})}{P(Y = r_0 | \mathbf{x})} = \exp(\beta_{0r} + \beta_{M,r} x_M + \beta_{K(1),r} x_{K(1)} + \dots + \beta_{K(I-1),r} x_{K(I-1)}) \quad (3.11)$$

und für Darstellung (3.10) die logarithmierten Chancen

$$\log \left(\frac{P(Y = r | \mathbf{x})}{P(Y = r_0 | \mathbf{x})} \right) = \beta_{0r} + \beta_{M,r} x_M + \beta_{K(1),r} x_{K(1)} + \cdots + \beta_{K(I-1),r} x_{K(I-1)}. \quad (3.12)$$

Um die Konstanten β_{0r} interpretieren zu können ist es notwendig bei der kategorialen Variablen K zwischen (0-1)- und Effektkodierung zu unterscheiden. Im Fall der (0-1)-Kodierung ergibt sich β_{0r} als der Logit bei $x_M = 0$, $K = I$. Die kategoriale Variable liegt demnach in der Referenzkategorie. Im Fall einer Effektkodierung stellt β_{0r} die mittlere logarithmierte Chance bei $x_M = 0$, gemittelt über alle Ausprägungen von K dar.

Bei festgehaltenen übrigen Einflussgrößen enthält

- $\beta_{M,r}$ für die metrische Größe x_M die additive Wirkung (aus (3.12)) auf die Logits für r gegenüber r_0 pro Einheit der Einflussgröße x_M und $e^{\beta_{M,r}}$ die multiplikative Wirkung (aus (3.11)) auf die Chancen für r gegenüber r_0 pro Einheit der Einflussgröße x_M .
- $\beta_{K(i),r}$ für die kategoriale Größe $K \in \{1, \dots, I\}$ in (0-1)-Kodierung die additive Wirkung auf die Logits für r gegenüber r_0 beim Übergang von $K = I$ zu $K = i$ und $e^{\beta_{K(i),r}}$ die multiplikative Wirkung auf die Chancen für r gegenüber r_0 beim Übergang von $K = I$ zu $K = i$.
- $\beta_{K(i),r}$ für die kategoriale Größe $K \in \{1, \dots, I\}$ in Effekt-Kodierung die additive Wirkung auf die Logits für r gegenüber r_0 als Abweichung von den über alle Ausprägungen von K gemittelten Logits und $e^{\beta_{K(i),r}}$ die multiplikative Wirkung auf die Chancen für r gegenüber r_0 , wobei das Produkt der Chancen auf 1 normiert ist, also $\prod_{i=1}^I e^{\beta_{K(i),r}} = 1$.

Das Chancenverhältnis bezieht sich immer auf die Chancen zwischen r und r_0 . Die logarithmierten relativen Chancen im Verhältnis zu Kategorie t erhält man für die Logits bei $x_M = 0$ aus $\beta_{K,r} - \beta_{K,t}$.

Wird nicht berücksichtigt, dass Parameter nur unter der Restriktion „bei gegebenen übrigen Einflussgrößen“ zu interpretieren sind, sind schwerwiegende Fehlinterpretationen die Folge.

Beispiel 3.3.1. (Verkehrsmittelwahl)

Betrachtet werden die Daten der Mobilitätsstudie '97 ([29]), allerdings mit nur einer Dummy-Variablen (Besitz einer Netzkarte). Die Responsekategorie besteht aus den Größen 1: Fuß, 2: Rad, 3: öffentlicher Verkehr (ÖV) und 4: Individualverkehr (IV).

	Fuß	Rad	ÖV	IV
	1	2	3	4
Netzkarte	271	94	318	281
keine Netzkarte	411	220	173	1018

Tabelle 2: Daten aus der Mobilitätsstudie '97

Für das Logit-Modell wird der Besitz einer Netzkarte wie folgt kodiert

$$x_N = \begin{cases} 0 & \text{Netzkarte} \\ 1 & \text{keine Netzkarte.} \end{cases}$$

Als Referenzkategorie r_0 wird Kategorie 4 (IV) gewählt. Das Modell besitzt dann die Form

$$\log \left(\frac{P(Y = r | x_N)}{P(Y = 4 | x_N)} \right) = \beta_{0r} + \beta_{N,r} x_N, \quad r = 1, 2, 3$$

bzw.

$$\frac{P(Y = r | 1)}{P(Y = 4 | 1)} = e^{\beta_{0r}} e^{\beta_{N,r}}, \quad \frac{P(Y = r | 0)}{P(Y = 4 | 0)} = e^{\beta_{0r}}.$$

Für die Parameter β_{0r} und $\beta_{N,r}$ ergibt sich

*** Multiple Logistic Model ***

Re-fitting to get Hessian

Call: multinom(formula = REF.KAT.IV ~ NETZKARTE, data = mobilität, na.action = na.omit, Hess = F, trace = F)

Coefficients:

Std. Errors:

	(Intercept) β_{0r}	NETZKARTE $\beta_{N,r}$		(Intercept)	NETZKARTE
Fuß	-0.036	-0.871	Fuß	0.085	0.103
ÖV	0.124	-1.896	ÖV	0.082	0.116
Rad	-1.095	-0.437	Rad	0.119	0.140

Std. Errors: Residual Deviance: 6677.548 AIC: 6689.548

Die Parameter β_{0r} geben die Logits der Referenzpopulation ($x_N = 0$, Netzkarte) an und $e^{\beta_{0r}}$ die entsprechenden Chancen jeweils im Vergleich zur Population $r_0 = 4$ (Individualverkehr). Die Chance ÖV gegenüber IV zu wählen betragen somit 1.132 ($e^{0.124}$), falls eine Netzkarte vorhanden ist. Der Parameter $\beta_{N,r}$ gibt die additive Veränderung der Logits

$$\log \frac{P(Y = r | N = \text{nein})}{P(Y = r_0 | N = \text{nein})} - \log \frac{P(Y = r | N = \text{ja})}{P(Y = r_0 | N = \text{ja})} = (\beta_{0r} + \beta_{N,r}) - \beta_{0r} = \beta_{N,r}$$

an, d.h. die Veränderung der logarithmierten Chancen für r gegenüber $r_0 = 4$ beim Übergang von der Population Netzkarte vorhanden ($N = ja$) zur Population keine Netzkarte vorhanden ($N = nein$). Der Parameter $e^{\beta_{N,r}}$ gibt entsprechend das Chancenverhältnis bzw. die relativen Chancen

$$\frac{P(Y = r | N = nein)/P(Y = r_0 | N = nein)}{P(Y = r | N = ja)/P(Y = r_0 | N = ja)} = e^{\beta_{N,r}}$$

wieder, d.h. die relative Chance von r gegenüber $r_0 = 4$ zwischen Netzkarte vorhanden und keine Netzkarte vorhanden. Der Wert $e^{\beta_{N,\ddot{O}V}} = e^{-1.896} = 0.15$ besagt, dass, falls keine Netzkarte vorhanden ist, die Chance öffentlichen Verkehr zu wählen gegenüber Individualverkehr zu wählen etwa 3/20 der Chance in der Population mit Netzkarte beträgt. Bei negativen Parameterschätzungen ist es häufig sinnvoll das umgekehrte Chancenverhältnis zu betrachten. Damit ergibt sich $e^{1.896} = 6.6$, d.h. die Chance IV gegenüber ÖV zu wählen beträgt bei Nichtbesitz einer Netzkarte in etwa das 6.6-fache der Chance bei Besitz einer Netzkarte.

Das Chancenverhältnis bezieht sich immer auf die Chance zwischen r und $r_0 = 4$. Die logarithmierten relativen Chancen im Verhältnis zu Kategorie t erhält man für die Logits aus $\beta_{N,r} - \beta_{N,t}$. Die folgende Tabelle gibt die logarithmierten relativen Chancen für die in den Zeilen gegebenen Verkehrsbereiche gegenüber den in den Spalten spezifizierten Verkehrsbereichen wieder. Dabei wird die Differenz zwischen Netzkarte und keine Netzkarte betrachtet.

	1	2	3	4
	Fuß	Rad	ÖV	IV
	$\beta_{N,r} - \beta_{N,1}$	$\beta_{N,r} - \beta_{N,2}$	$\beta_{N,r} - \beta_{N,3}$	$\beta_{N,r}$
1 Fuß	0	-0.434	1.025	-0.871
2 Rad	0.434	0	1.46	-0.437
3 ÖV	-1.025	-1.46	0	-1.896
4 IV	0.871	0.437	1.896	0

Tabelle 3: logarithmierte relative Chancen

Der stärkste Besitzeffekt liegt mit -1.896 bei den Chancen für $r = 3$ (ÖV) gegenüber $r = 4$ (IV) bzw. mit 1.896 bei den Chancen für $r = 4$ gegenüber $r = 3$. Die zugehörige relative Chance

$$\frac{P(Y = 3 | N = nein)/P(Y = 4 | N = nein)}{P(Y = 3 | N = ja)/P(Y = 4 | N = ja)} = e^{-1.896} = 0.15$$

besagt, dass die Chancen ÖV zu wählen gegenüber IV zu wählen ohne Netzkarte nur das 0.15-fache der Chance mit Netzkarte betragen.

Im folgenden Beispiel wird die Interaktion zwischen einer metrischen Variable (Alter) und einer kategorialen Variable (Führerscheinbesitz) modelliert.

Beispiel 3.3.2. (Verkehrsmittelwahl in Abhängigkeit von Alter und Führerscheinbesitz mit Interaktionseffekten)

Bei der Erhebung der Mobilitätsstudie ergibt sich folgende nach Alter und Führerscheinbesitz aufgeschlüsselte Übersicht.

Führerschein	Alter	Anzahl der Wege mit Verkehrsmittel			
		Fuß	Rad	ÖV	IV
ja	≤ 18	2	0	5	4
	18 + bis 31	150	52	96	266
	31 + bis 42	151	71	67	393
	42 + bis 55	92	85	71	326
	55 + bis 65	90	11	65	135
	≥ 65	44	9	33	48
nein	≤ 18	34	37	51	51
	18 + bis 31	5	0	14	21
	31 + bis 42	12	22	4	12
	42 + bis 55	25	3	20	7
	55 + bis 65	58	16	31	25
	≥ 65	13	8	22	6

Tabelle 4: Daten aus der Mobilitätsstudie '97

Betrachtet wird das multinomiale Logit-Modell mit den Einflussgrößen Pkw-Führerschein (PFS) und Alter (A), wobei Alter metrisch in die Analyse eingeht. Als Referenzkategorie wird wieder $r_0 = 4$ (IV) gewählt.

Das Modell besitzt dann folgende Form

$$\log \frac{P(Y = r | x_{PFS}, A)}{P(Y = 4 | x_{PFS}, A)} = \beta_{0r} + \beta_{PFS,r} x_{PFS} + \beta_{A,r} A + \beta_{PFS * A,r} x_{PFS} * A,$$

wobei der Führerscheinbesitz wie folgt Dummy kodiert wurde

$$x_{PFS} = \begin{cases} 0 & \text{kein Pkw-Führerschein} \\ 1 & \text{Pkw-Führerschein.} \end{cases}$$

Tabelle 5 gibt die Schätzungen wieder. An den p -Werten wird deutlich, dass zwischen Alter und Führerschein zumindest für eine Kategorie die Interaktion nicht zu vernachlässigen ist. Der p -Wert gibt an, mit welcher Wahrscheinlichkeit die Teststatistik (vgl. Abschnitt 3.7.4) einen noch extremeren Wert annimmt als den tatsächlich eingetretenen wenn die Hypothese wahr ist. Kleine p -Werte (unter 0.05) sprechen damit für eine hohe Signifikanz.

Effekt	Parameter	Schätzwert	Standardfehler	Wald	p-Wert
Intercept	Fuß	-0.969	0.259	14.028	0.000
	ÖV	-0.537	0.244	4.824	0.028
	Rad	-0.556	0.266	4.376	0.036
Alter	Fuß	0.029	0.006	25.342	0.000
	ÖV	0.019	0.006	10.498	0.001
	Rad	0.006	0.007	0.850	0.357
Pkw-Führerschein	Fuß	-0.096	0.309	0.096	0.756
	ÖV	-1.319	0.318	17.205	0.000
	Rad	-1.102	0.353	9.753	0.002
Alter*Führerschein	Fuß	-0.023	0.007	10.857	0.001
	ÖV	-0.0045	0.007	0.383	0.536
	Rad	-0.006	0.009	0.435	0.510

Tabelle 5: Multinomiales Modell für die Verkehrsmittelwahl

Eine separate Analyse des Modells, einmal für den Fall Pkw-Führerschein vorhanden, einmal für den Fall kein Pkw-Führerschein vorhanden, lässt den Einfluss der Kovariablen Alter deutlich werden. Falls $X_{PFS} = 1$ (Pkw-Führerschein) ergibt sich folgendes Modell

$$\begin{aligned} \log \left(\frac{P(Y = r | X_{PFS} = 1, A)}{P(Y = 4 | X_{PFS} = 1, A)} \right) &= \beta_{0r} + \beta_{PFS,r} + \beta_{A,r}A + \beta_{PFS*A,r}A \\ &= \beta_{0r} + \beta_{PFS,r} + (\beta_{A,r} + \beta_{PFS*A,r})A, \end{aligned}$$

für $x_{PFS} = 0$ (kein Pkw-Führerschein) erhält man

$$\log \left(\frac{P(Y = r | X_{PFS} = 0, A)}{P(Y = 4 | X_{PFS} = 0, A)} \right) = \beta_{0r} + \beta_{A,r}A.$$

Aus diesen Gleichungen ist sofort ersichtlich, dass der Steigungsparameter für Alter in den Subpopulationen Pkw-Führerschein / kein Pkw-Führerschein unterschiedlich ist. Die Steigungsparameter $\beta_{A,r} + \beta_{PFS*A,r}$ bzw. $\beta_{A,r}$ lauten

	Steigung	
	Pkw-Führerschein	kein Pkw-Führerschein
r=Fuß	0.006	0.029
r=ÖV	0.015	0.019
r=Rad	0.000	0.006

Tabelle 6: Steigungsparameter für Alter der Population Führerschein

Aus dem Haupteffekt $\beta_{A, \text{Fuß}} = 0.029$ erkennt man, dass die Entscheidung zu Fuß zu gehen ($Y = 1$) gegenüber Individualverkehr zu benutzen ($Y = 4$) mit zunehmendem Alter zunimmt. Aus der letzten Tabelle ergibt sich, dass die Zunahme im Fall kein Pkw-Führerschein vorhanden größer ist als für den Fall Pkw-Führerschein vorhanden.

3.4 Das bedingte Logit-Modell

Die Einflussgrößen \mathbf{x} des einfachen multinomialen Logit-Modells (3.6) bzw. (3.7) sind unabhängig von der Kategorie. Sind in den Responsekategorien Wahlmöglichkeiten enthalten, beispielsweise die Wahl eines Verkehrsmittels, so enthält der Einflussgrößenvektor \mathbf{x} nur Merkmale, die den Entscheidungsträger charakterisieren. Beispiele sind sozio-ökonomischen Variablen wie das Alter oder der Führerscheinbesitz. Die Wahlmöglichkeiten können jedoch häufig selbst durch Merkmale charakterisiert werden. So ist zum Beispiel die Wahl eines Verkehrsmittels abhängig von der Dauer oder den Kosten einer Fahrt. Um kategorienspezifische Variablen einzuführen, ist der im multinomialen Logit-Modell benutzte Einflussgrößenvektor durch einen entsprechenden Vektor der die alternativen-spezifischen Merkmale enthält zu ersetzen. Das Modell könnte folgende Form annehmen:

$$\log \left(\frac{P(Y = r | \mathbf{x}, \{\omega_j\})}{P(Y = k | \mathbf{x}, \{\omega_j\})} \right) = \mathbf{x}^T \boldsymbol{\beta}_r + (\omega_r - \omega_k)^T \boldsymbol{\alpha}, \quad (3.13)$$

wobei $\{\omega_j\}$ die kategorienspezifischen Merkmale $\omega_1, \dots, \omega_k$ beschreibt. Neben dem Term $\mathbf{x}^T \boldsymbol{\beta}_r$ enthält das Modell nun auch den Vergleichsterm $(\omega_r - \omega_k)^T \boldsymbol{\alpha}$, wobei $\boldsymbol{\alpha}$ ein von der Kategorie unabhängiger Parametervektor ist.

Beispiel 3.4.1. *Betrachtet werden die beiden Verkehrsmittel $r = \text{Rad}$ und $k = \text{Pkw}$ (Referenzkategorie). ω_r beschreibt die Zeitdauer der Alternative r . Ein Vergleich beider Alternativen liefert die Differenz $\omega_r - \omega_k$. Der Parameter α gibt das Gewicht an, dass dieser Zeitdifferenz zukommen soll.*

Der Term kann auch Interaktionen enthalten.

Eine allgemeinere Darstellung des Modells findet man in Powers und Xie (2000) [25].

3.5 Das multinomiale Logit-Modell als verallgemeinertes lineares Modell

Der Einfachheit halber setzt man $\pi_r = P(Y = r | \mathbf{x})$. Es genügt die ersten $q = k - 1$ Wahrscheinlichkeiten zu betrachten, da $\sum_{r=1}^k \pi_r = 1$. Die $k - 1$ Gleichungen lassen sich wie folgt in Matrixschreibweise darstellen

$$\begin{pmatrix} \log(\pi_1 / (1 - \pi_1 - \dots - \pi_q)) \\ \vdots \\ \log(\pi_q / (1 - \pi_1 - \dots - \pi_q)) \end{pmatrix} = \underbrace{\begin{pmatrix} \mathbf{x}^T & 0 \\ & \ddots \\ 0 & \mathbf{x}^T \end{pmatrix}}_{=: \mathbf{Z} \in \mathbb{R}^{q \times (p+1)q}} \underbrace{\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_q \end{pmatrix}}_{\boldsymbol{\beta} \in \mathbb{R}^{(p+1)q}} = \mathbf{Z}\boldsymbol{\beta}, \quad (3.14)$$

mit $\mathbf{x}^T = (1, x_1, \dots, x_p)$, $\boldsymbol{\beta}^T = (\beta_1^T, \dots, \beta_q^T)$ und $\beta_r^T = (\beta_{0r}, \beta_{1r}, \dots, \beta_{pr})$, $r = 1, \dots, q$. Formulierung (3.14) entspricht der Darstellung als verallgemeinertes lineares Modell. Die

allgemeine Form des Modells lautet dann

$$g(\boldsymbol{\pi}) = \mathbf{Z}\boldsymbol{\beta}, \quad (3.15)$$

wobei $\boldsymbol{\pi}^T = (\pi_1, \dots, \pi_q)$, $\boldsymbol{\beta}^T = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_q)$ und \mathbf{Z} die in (3.14) beschriebene Designmatrix darstellt. Für die Komponenten in $g = (g_1, \dots, g_q) : \mathbb{R}^q \rightarrow \mathbb{R}^q$ gilt

$$g_r(\pi_1, \dots, \pi_q) = \log \left(\frac{\pi_r}{1 - \pi_1 - \dots - \pi_q} \right), \quad r = 1, \dots, q.$$

Mit der Umkehrfunktion $h = g^{-1}$ kann (3.15) dargestellt werden durch

$$\boldsymbol{\pi} = h(\mathbf{Z}\boldsymbol{\beta}). \quad (3.16)$$

Für die Komponenten in $h = (h_1, \dots, h_q) : \mathbb{R}^q \rightarrow \mathbb{R}^q$ gilt

$$h_r(\eta_1, \dots, \eta_q) = \frac{\exp(\eta_r)}{1 + \sum_{s=1}^q \exp(\eta_s)}, \quad r = 1, \dots, q.$$

3.6 Explorative Datenanalyse

Ziel der explorative Datenanalyse ist es, neben sachtheoretischen Überlegungen geeignete Kovariablen für ein multinomiales Logit-Modell vorzuschlagen.

3.6.1 Explorative Datenanalyse zum Erkennen marginaler Einflüsse einer Kovariable

Um einen ersten Überblick über mögliche Einflussgrößen zu bekommen, ist es sinnvoll die Plots der *marginalen Effekte*, im Folgenden auch *Haupteffekte* genannt, zu betrachten. Dazu wird die Funktion **haupteffekt**, die die empirischen Logits über die Einflussgröße abträgt, verwendet (siehe Anhang B.1). Die Logits werden unmittelbar aus den Beobachtungen berechnet. Zu beachten ist, dass jeweils nur zwei Responsekategorien gegenübergestellt werden, d.h. man vergleicht die Referenzkategorie r_0 mit r , $r = 1, \dots, q$. Y sei eine binomialverteilte Zufallsvariable mit n Parametern und Wahrscheinlichkeit π . Die relative Häufigkeit $p := Y/n$ ist eine Schätzung von π . Durch das Logit-Modell wird ein linearer Zusammenhang zwischen den logarithmierten Chancen und der Einflussgröße postuliert. Folglich ist es sinnvoll zu untersuchen, ob die Daten diesen linearen Zusammenhang wiedergeben. Daher betrachtet man anstelle der relativen Häufigkeiten p die empirischen Logits

$$\text{emp.logit} = \log \left(\frac{p}{1-p} \right).$$

Wegen der problematischen Berechnung für den Fall, dass die Zielvariable gleich Null bzw. gleich n ist, werden die „korrigierten empirischen Logits“

$$\text{emp.logit}_{\text{korrr}} = \log \left(\frac{p + c/n}{1 - p + c/n} \right)$$

betrachtet (Haldane, 1955 [17], Anscombe, 1956 [2]), d.h. es handelt sich um eine Korrektur um c „falsche Beobachtungen“ mit $y = 1$ und c „falsche Beobachtungen“ mit $y = 0$. Für c wird der Wert 0.5 gewählt, so dass insgesamt eine künstliche Beobachtung hinzugefügt wird. Für großes n lässt sich die relative Häufigkeit p durch die Normalverteilung approximieren, d.h.

$$p \stackrel{approx.}{\sim} N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$$

(vgl. Tutz Seite 370, [27]). Das approximative Konfidenzintervall ist daher von der Form

$$\left[p - z_{1-\frac{\alpha}{2}} \sqrt{Var(p)}, p + z_{1-\frac{\alpha}{2}} \sqrt{Var(p)}\right]$$

für die Wahrscheinlichkeit π . $z_{1-\alpha/2}$ bezeichnet das $(1 - \alpha/2)$ -Quantil der Standardnormalverteilung. Das Konfidenzintervall ist symmetrisch um die Wahrscheinlichkeit.

Die Vorgehensweise soll beispielhaft für eine metrische und für eine kategoriale Variable gezeigt werden.

Marginaler Einfluss einer metrischen Kovariable

Beispiel 3.6.1. (Verkehrsmittelwahl in Abhängigkeit vom Alter)

Untersucht wird, ob die metrische Größe Alter als Kovariable in das Modell eingehen soll. Dazu werden die empirischen Logits und die dazugehörigen Zellgrößen betrachtet. Zunächst muss jedoch die metrische Variable gruppiert werden um binomiale Daten zu erhalten. Als Referenzkategorie wird Kategorie 4 (Individualverkehr) verwendet.

Haupteffekte für Alter

=====

Als Referenzkategorie wurde IV gewählt

Referenzkategorie und Verkehrsmittel Fuß

	18+ bis 31	31+ bis 42	42+ bis 55	55+ bis 65	> 65
emp. Logits	-0.58	-0.93	-1.04	-0.15	0.09
Zellengröße n	431	564	450	297	105

Referenzkategorie und Verkehrsmittel ÖV

	18+ bis 31	31+ bis 42	42+ bis 55	55+ bis 65	> 65
emp. Logits	-0.94	-1.74	-1.29	-0.51	-0.09
Zellengröße n	384	476	424	256	105

Referenzkategorie und Verkehrsmittel Rad

	18+ bis 31	31+ bis 42	42+ bis 55	55+ bis 65	> 65
emp. Logits	-1.66	-1.53	-1.33	-1.76	-1.67
Zellengröße n	328	492	421	187	59

Die zugehörigen Plots haben dann folgende Gestalt:

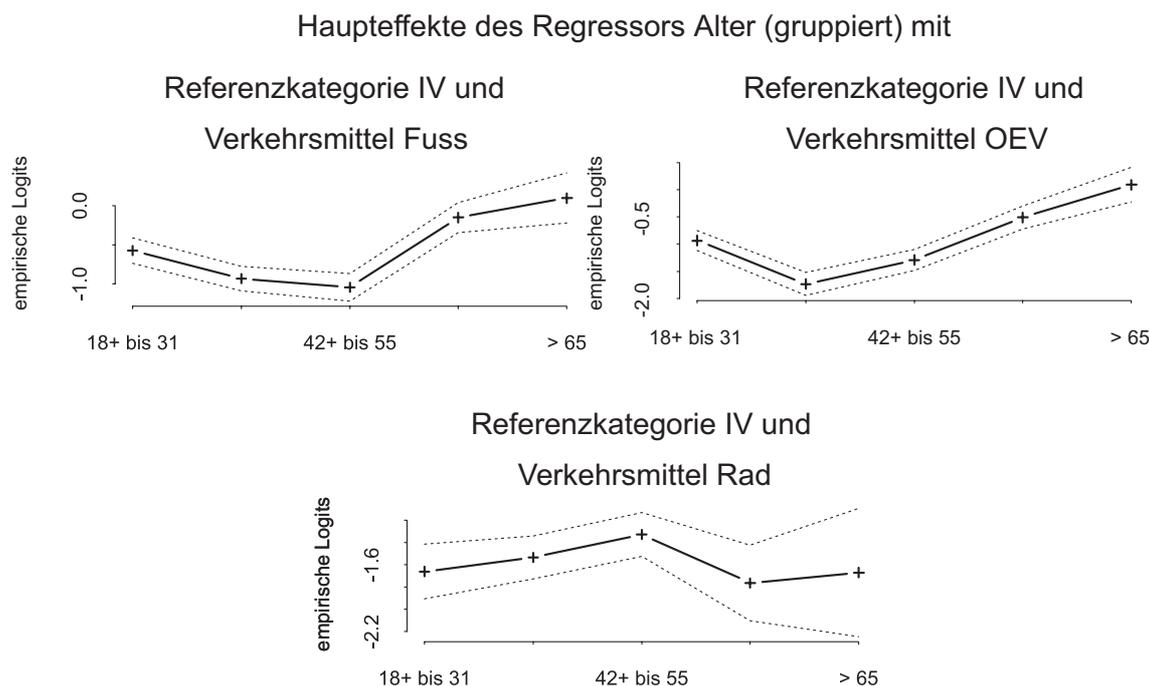


Abbildung 1: Alter als Haupteffekt

Die Grafiken lassen einen nichtlinearen Effekt vermuten. Betrachtet man die empirischen Logits der Referenzkategorie IV gegenüber der Kategorie Fuß erkennt man, dass bis zu einem Alter von 55 Jahren die Wahrscheinlichkeit zu Fuß zu gehen gegenüber der Wahrscheinlichkeit IV zu benutzen abnimmt. Ab einem Alter von 55 Jahren nimmt die Wahrscheinlichkeit zu Fuß zu gehen wieder zu. Die erste Grafik zeigt diesen Verlauf noch einmal sehr deutlich. Es lässt sich zudem ein leichter quadratischer Effekt feststellen. Die zweite Grafik, d.h. die Wahl von öffentlichen Verkehrsmitteln gegenüber Individualverkehr, lässt ebenfalls einen nichtlinearen Effekt erkennen. Alter wird hier quadratisch eingehen. Mit zunehmendem Alter wird das Verkehrsmittel ÖV gegenüber IV verstärkt genutzt. Ein Grund hierfür kann sein, dass insbesondere in den Altersgruppen der über 55-jährigen der Anteil der Führerscheinbesitzer wieder abnimmt. Die dritte Grafik zeigt keinen deutlich signifikanten Einfluss.

Marginaler Einfluss einer kategorialen Kovariable

Beispiel 3.6.2. (Verkehrsmittelwahl in Abhängigkeit vom Besitz einer Netzkarte)

Als Referenzkategorie wird wieder Kategorie 4 (Individualverkehr) verwendet. Untersucht werden soll, ob der Besitz einer Netzkarte einen Einfluss auf die Wahl des Verkehrsmittels hat. Die empirischen Logits und die dazugehörigen Zellgrößen lauten:

Haupteffekte für NETZKARTE

=====

Als Referenzkategorie wurde IV gewählt

Referenzkategorie und Verkehrsmittel Fuß

	ja	nein
emp. Logits	0.00	-0.95
Zellengröße n	501	1357

Referenzkategorie und Verkehrsmittel ÖV

	ja	nein
emp. Logits	0.10	-1.83
Zellengröße n	527	1135

Referenzkategorie und Verkehrsmittel Rad

	ja	nein
emp. Logits	-1.25	-1.63
Zellengröße n	321	1171

Die zugehörigen Plots haben dann folgende Gestalt:

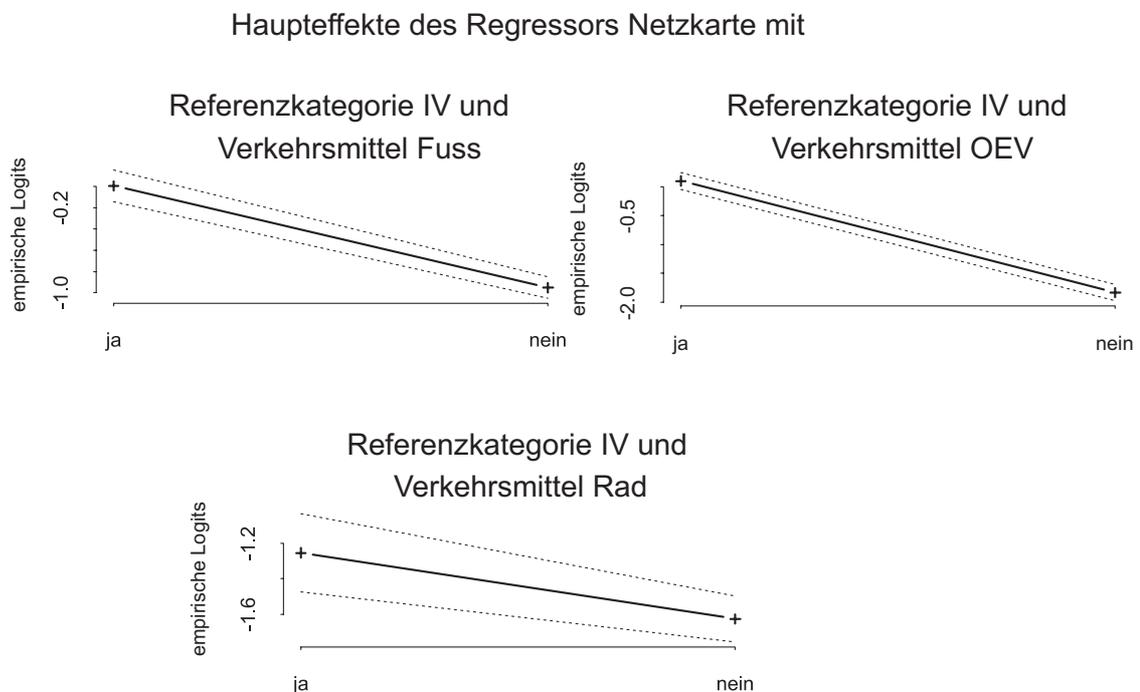


Abbildung 2: Netzkarte als Haupteffekt

Betrachtet man wiederum die empirischen Logits der Referenzkategorie IV gegenüber der Kategorie Fuß wird deutlich, dass Personen die eine Netzkarte besitzen häufiger Wege zu Fuß gehen, als Personen die keine Netzkarte besitzen. Betrachtet man darüberhinaus die

Variable Anbindung an ÖV wird deutlich, dass bei den hier betrachteten Wegen alle Personen eine gute bis sehr gute Anbindung zu öffentlichen Verkehrsmitteln haben. Die gute Anbindung ist allerdings darauf zurückzuführen, dass die Daten auf den Raum München beschränkt sind. Die erste Grafik verdeutlicht diesen Verlauf entsprechend. An der zweiten Grafik lässt sich der Einfluss ebenfalls sehr gut erkennen. Die Chance den ÖV gegenüber IV zu benutzen, bei Besitz einer Netzkarte, ist positiv. Eine mögliche Ursache für die Zunahme der Wege zu Fuß kann sein, dass an ÖV anschließende Wege zu Fuß zurückgelegt werden. Die dritte Grafik lässt keinen großen Einfluss bei Besitz einer Netzkarte vermuten.

Diese beiden Beispiele verdeutlichen, dass die Größen Alter und Netzkarte keinen Einfluss bei der Verkehrsmittelwahl von Rad gegenüber IV aufweisen. Die Verkehrsmittelwahl zu Fuß bzw. ÖV gegenüber IV hängt jedoch sowohl vom Alter, als auch vom Besitz einer Netzkarte ab. Hier ist es hilfreich sogenannte kategorienspezifische Variablen einzuführen (vgl. Abschnitt 3.4). Das hier verwendete Statistikpaket S-Plus verfügt derzeit noch nicht über eine entsprechende Implementierungsmöglichkeit. Daher ist die Programmierung einer entsprechenden Funktion notwendig. Dies soll in einer weiterführenden Diplomarbeit vorgenommen werden.

Häufig können kategoriale Variablen weiter zusammengefasst werden. Eine Hilfestellung bieten dabei die Plots aus der Funktion *haupteffekt*. Sind beispielsweise die empirischen Logits für die Altersgruppe 55+ bis 65 und > 65 nahezu identisch können diese beiden Gruppen zu der Altersgruppe > 55 zusammengefasst werden. Abbildung 1 zeigt, dass dies hier jedoch nicht der Fall ist.

3.6.2 Explorative Datenanalyse zum Erkennen von Interaktionen

Nach der Wahl eines Haupteffekt-Modells können Beziehungen zueinander untersucht werden (*Interaktionen*). In Beispiel 3.3.2 wurde bereits eine Möglichkeit aufgezeigt Interaktionseffekte festzustellen. Mit Hilfe der Funktion *interaktion* (vgl. Anhang B.3) und mit Hilfe des Multinomialen Modells werden die Größen explorativ auf Interaktionen untersucht. Die Funktion stellt analog zur Funktion *haupteffekt* wieder jeweils nur zwei Responsekategorien gegenüber. Wird eine Kategorie als Responsekategorie ausgezeichnet, genügt es, die $k - 1$ korrigierten empirischen Logits zu betrachten. Wie auch bei den Haupteffekten, sind metrische Variablen für die explorative Analyse zu gruppieren. Das folgende Beispiel modelliert eine Interaktion. Das Modell besteht nur aus den Größen Alter und Nutzung.

Beispiel 3.6.3. (Interaktion zwischen Alter und Nutzung)

Als Referenzkategorie Kategorie dient erneut der Verkehrsbereich „Individualverkehr“. Unabhängig davon, welche Kodierung für die Variable Alter gewählt wird, impliziert das Haupteffekt-Modell die Forderung, dass die Steigung der Logits über die Pkw-Nutzung nicht vom Alter abhängt. Die Logits über die Nutzung betrachtet, sollen für die Subpopulation 18+ bis 31, 31+ bis 42, 42+ bis 55, 55+ bis 65 und > 65 parallel zueinander verlaufen. Ist das nicht der Fall, ist von einer Interaktion zwischen der Pkw-Nutzung und dem Alter auszugehen. Die empirischen Logits und die dazugehörigen Zellgrößen lauten dann:

Interaktionen der Regressoren Nutzung und Alter

=====

Als Referenzkategorie wurde IV gewählt

empirische Logits der Referenzkategorie und dem Verkehrsmittel Fuß

	18+ bis 31	31+ bis 42	42+ bis 55	55+ bis 65	> 65
Nutz.Pkw-Hauptn	-0.58	-1.41	-1.46	-1.40	-0.76
Nutz.Pkw-Mitn	-0.54	-0.59	-1.26	1.15	0.93
Nutz.sonst	-0.61	1.22	1.32	0.99	1.27

dazugehörige Zellengrößen

	18+ bis 31	31+ bis 42	42+ bis 55	55+ bis 65	> 65
Nutz.Pkw-Hauptn	273	400	305	153	54
Nutz.Pkw-Mitn	112	110	96	47	36
Nutz.sonst	46	54	49	97	15

empirische Logits der Referenzkategorie und dem Verkehrsmittel ÖV

	18+ bis 31	31+ bis 42	42+ bis 55	55+ bis 65	> 65
Nutz.Pkw-Hauptn	-1.31	-2.50	-2.07	-1.75	-1.48
Nutz.Pkw-Mitn	-0.62	-0.89	-1.65	0.76	-0.85
Nutz.sonst	-0.26	0.28	1.49	0.66	2.52

dazugehörige Zellengrößen

	18+ bis 31	31+ bis 42	42+ bis 55	55+ bis 65	> 65
Nutz.Pkw-Hauptn	222	348	279	144	45
Nutz.Pkw-Mitn	109	100	89	35	14
Nutz.sonst	53	28	56	77	46

empirische Logits der Referenzkategorie und dem Verkehrsmittel Rad

	18+ bis 31	31+ bis 42	42+ bis 55	55+ bis 65	> 65
Nutz.Pkw-Hauptn	-3.03	-2.58	-1.77	-2.80	-2.71
Nutz.Pkw-Mitn	-1.11	-0.89	-0.84	-3.14	-3.04
Nutz.sonst	-0.35	1.02	0.32	-0.26	0.76

dazugehörige Zellengrößen

	18+ bis 31	31+ bis 42	42+ bis 55	55+ bis 65	> 65
Nutz. Pkw-Hauptn	183	346	290	130	39
Nutz. Pkw-Mitn	94	100	107	11	10
Nutz. sonst	51	46	24	46	10

Die zugehörigen Plots haben dann folgende Gestalt:

Interaktionseffekte der Regressoren Nutzung und Alter

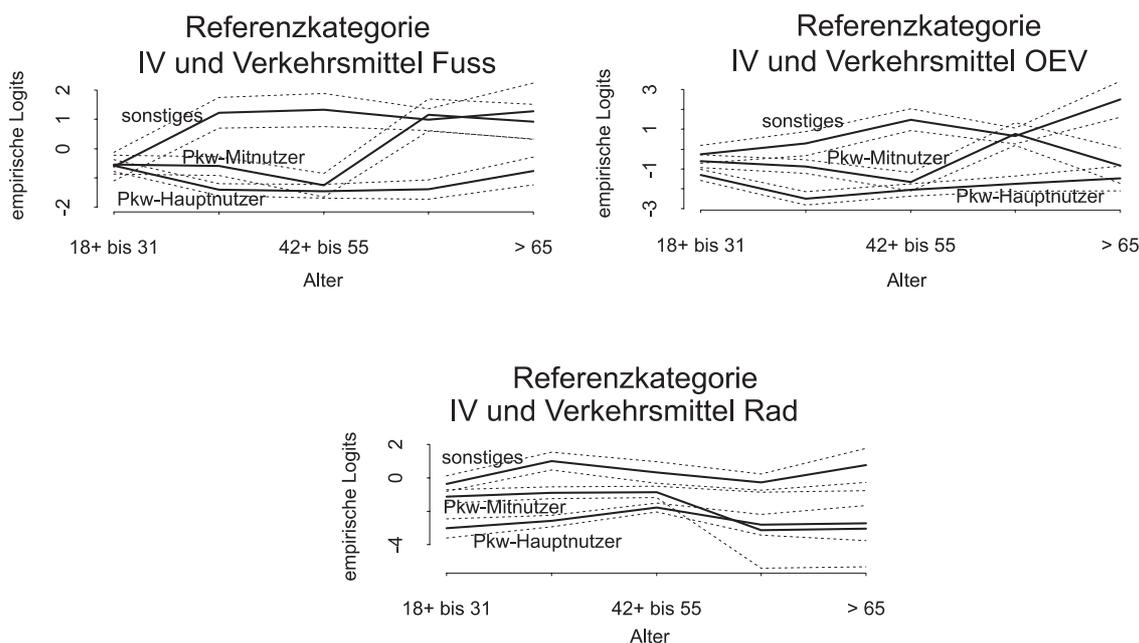


Abbildung 3: Interaktion der Regressoren Nutzung und Alter

In allen drei Grafiken sind Nichtparallelitäten zu erkennen. Vergleicht man beispielsweise in der ersten Grafik den Verlauf der Pkw-Hauptnutzer mit dem der Pkw-Mitnutzer, so lassen sich starke Nichtparallelitäten erkennen. Es werden folglich Interaktionen erwartet. Vergleicht man hingegen diese beiden Verläufe in der zweiten Grafik, stellt man fest, dass die Steigungsstruktur dieselbe ist, d.h. zunächst fallen die empirischen Logits, während sie mit zunehmendem Alter wieder ansteigen. Aufgrund der gleichen Struktur sind hier keine Interaktionen zu erwarten. Vergleicht man den Verlauf der Pkw-Hauptnutzer mit dem der Nichtnutzer (sonstiges), so lassen sich wieder deutliche Nichtparallelitäten erkennen. Hier kann wieder mit Interaktionen gerechnet werden. Analog können nun alle anderen Verläufe miteinander verglichen und logische Schlüsse gezogen werden. An den Grafiken ist ebenfalls sofort ersichtlich, dass Pkw-Hauptnutzer, im Vergleich zu den anderen Bereichen Fuß, öffentlicher Verkehr und Rad, am häufigsten den Verkehrsbereich Individualverkehr wählen.

Bemerkung 3.6.1. Die explorative Datenanalyse dient in erster Linie zur Identifizierung einzelner marginaler Einflüsse. Alle übrigen Einflussgrößen werden hierbei zunächst ver-

nachlässigt. Man bekommt folglich eine erste Hilfestellung zur weiteren Variablenselektion, aber noch kein endgültiges Modell. Ein weiterer Schritt ist es daher, ein Modell zu verifizieren.

3.7 Schätzen und Testen für multinomiale Modelle

Das multinomiale Logit-Modell ist durch die Linkfunktion $g = (g_1, \dots, g_q) : \mathbb{R}^q \rightarrow \mathbb{R}^q$ (bzw. Responsefunktion $h = g^{-1}$) und einen linearen Prädiktor bestimmt, wobei die Linkfunktion in erster Linie die Responsewahrscheinlichkeiten angibt. Welche Variablen einen signifikanten Beitrag leisten, wird durch die linearen Prädiktoren spezifiziert. Die Aussagekraft eines Modells, von der Relevanz bis hin zur Anpassung des Modells lässt sich unter verschiedenen Aspekten betrachten:

- Verträglichkeit des Modells mit den Daten
- Überprüfung der Relevanz einzelner Einflussgrößen
- Prognostische Aussagekraft des Modells

Die Aspekte sind zum Teil sehr eng miteinander verbunden. So ist zum Beispiel die Verträglichkeit des Modells mit den Daten Voraussetzung zur weiteren Untersuchung der Einflussgrößen. Bei der Analyse der Variablen auf relevante Einflüsse, wird versucht die Kovariablen, die keinen signifikanten Beitrag leisten zu identifizieren und im nächsten Schritt zu eliminieren, während beim Erklärungswert der Modelle die Quantifizierung des Erklärungswertes der übrigen Variablen im Vordergrund steht.

Grundlage aller Schätzungen ist die Parameterschätzung, die im folgenden Abschnitt betrachtet wird.

3.7.1 Maximum-Likelihood-Schätzung

Die Maximum-Likelihood-Schätzung wird für den Fall unabhängiger Beobachtungen eingeführt, d.h. Ausgangspunkt sind unabhängige gruppierte Beobachtungen $(\mathbf{y}_i, \mathbf{x}_i)$, $i = 1, \dots, g$, wobei die \mathbf{y}_i multinomialverteilt sind, d.h. $\mathbf{y}_i \sim M(n_i, \boldsymbol{\pi}_i)$. Die Komponente y_{ir} des Beobachtungsvektors $\mathbf{y}_i^T = (y_{i1}, \dots, y_{iq})$ enthält die Anzahl der Beobachtungen in den Kategorien $1, \dots, q$, bei n_i Beobachtungen für die Einflussgrößenkombination \mathbf{x}_i . Dabei bezeichnet n_i den lokalen Stichprobenumfang an der Messstelle \mathbf{x}_i . Durch $n = n_1 + \dots + n_g$ ist der Gesamtstichprobenumfang gegeben. Durch $\mathbf{p}_i^T = (p_{i1}, \dots, p_{iq})$ mit $p_{ir} = y_{ir}/n_i$ ist die *relative Häufigkeit* bestimmt. Im Folgenden wird das Schätzprinzip für das allgemeine Modell $g(\boldsymbol{\pi}) = \mathbf{Z}\boldsymbol{\beta}$ für alle Beobachtungen betrachtet. In Matrixform lässt sich das

Modell wie folgt darstellen

$$g(\boldsymbol{\pi}) = \begin{bmatrix} g(\boldsymbol{\pi}_1) \\ \vdots \\ g(\boldsymbol{\pi}_g) \end{bmatrix} = \begin{bmatrix} \mathbf{Z}^{(1)}\boldsymbol{\beta} \\ \vdots \\ \mathbf{Z}^{(g)}\boldsymbol{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1^T & 0 \\ & \ddots \\ 0 & \mathbf{x}_1^T \\ \mathbf{x}_2^T & 0 \\ & \ddots \\ 0 & \mathbf{x}_2^T \\ & \vdots \\ \mathbf{x}_g^T & 0 \\ & \ddots \\ 0 & \mathbf{x}_g^T \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_q \end{bmatrix}$$

wobei

$\boldsymbol{\pi}^T = (\boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_g) = (\pi_{11}, \dots, \pi_{1q}, \dots, \pi_{g1}, \dots, \pi_{gq}) \in \mathbb{R}^{gq}$ mit $\pi_{ir} = P(Y = r \mid \mathbf{x}_i)$ den Vektor der Responsewahrscheinlichkeiten darstellt,

$\mathbf{x}_i^T = (1, x_{i1}, \dots, x_{ip}) \in \mathbb{R}^{p+1}$,

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}^{(1)} \\ \vdots \\ \mathbf{Z}^{(g)} \end{bmatrix} \in \mathbb{R}^{gq \times (p+1)q},$$

$\mathbf{Z}^{(i)} \in \mathbb{R}^{q \times (p+1)q}$ die aus den Einflussgrößen \mathbf{x}_i konstruierte Designmatrix ist,

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_q \end{bmatrix} \in \mathbb{R}^{(p+1)q} \text{ mit } \boldsymbol{\beta}_r = \begin{bmatrix} \beta_{0r} \\ \vdots \\ \beta_{pr} \end{bmatrix} \in \mathbb{R}^{(p+1)}$$
 den zu schätzenden Parameter-

vektor repräsentiert,

$i = 1, \dots, g$ die Anzahl der Gruppen,

$r = 1, \dots, q = k - 1$ die Anzahl der Ausprägungen und

$0, \dots, p$ die Anzahl der Kovariablen beschreibt.

Im Folgenden wird nur die i -te Gruppe betrachtet, d.h. das Modell

$$g(\boldsymbol{\pi}_i) = \mathbf{Z}^{(i)}\boldsymbol{\beta} = \underbrace{\begin{pmatrix} \mathbf{x}_i^T & 0 \\ & \ddots \\ 0 & \mathbf{x}_i^T \end{pmatrix}}_{\mathbf{Z}^{(i)} \in \mathbb{R}^{q \times (p+1)q}} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_q \end{pmatrix} = \underbrace{\begin{pmatrix} \beta_{01} + \beta_{11}x_{i1} + \dots + \beta_{p1}x_{ip} + 0 \\ \vdots \\ 0 + \beta_{0q} + \beta_{1q}x_{i1} + \dots + \beta_{pq}x_{ip} \end{pmatrix}}_{\in \mathbb{R}^q}.$$

Weitere Notationen:

$$\mathbf{Z}^{(i)} := \underbrace{\begin{pmatrix} \mathbf{z}_{11}^{(i)} & \cdots & \mathbf{z}_{1q}^{(i)} \\ \vdots & \ddots & \vdots \\ \mathbf{z}_{q1}^{(i)} & \cdots & \mathbf{z}_{qq}^{(i)} \end{pmatrix}}_{\in \mathbb{R}^{q \times (p+1)q}}, \text{ wobei } \mathbf{z}_{mj}^{(i)} = \begin{cases} \mathbf{x}_i^T & m = j \\ \mathbf{0} & \text{sonst.} \end{cases} \quad (3.17)$$

$$\mathbf{Z}_{\cdot j}^{(i)} := \begin{pmatrix} \mathbf{z}_{1j}^{(i)} \\ \vdots \\ \mathbf{z}_{qj}^{(i)} \end{pmatrix} \text{ beschreibt die } j\text{-te Spalte von } \mathbf{Z}^{(i)}. \quad (3.18)$$

$\mathbf{Z}_{j\cdot}^{(i)} := (\mathbf{z}_{j1}^{(i)}, \dots, \mathbf{z}_{jq}^{(i)})$ beschreibt die j -te Zeile von $\mathbf{Z}^{(i)}$.

$$g(\boldsymbol{\pi}_i) = \mathbf{Z}^{(i)} \boldsymbol{\beta} := \boldsymbol{\eta}^{(i)} = \begin{pmatrix} \eta_1^{(i)} \\ \vdots \\ \eta_q^{(i)} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_i^T \boldsymbol{\beta}_1 + 0 \\ \vdots \\ 0 + \mathbf{x}_i^T \boldsymbol{\beta}_q \end{pmatrix} \quad (3.19)$$

Mit $h = g^{-1}$ kann das Modell auch wie folgt beschrieben werden

$$\boldsymbol{\pi}_i = h(\mathbf{Z}^{(i)} \boldsymbol{\beta}) = (h_1(\mathbf{Z}^{(i)} \boldsymbol{\beta}), \dots, h_q(\mathbf{Z}^{(i)} \boldsymbol{\beta})) \stackrel{(3.19)}{=} (h_1(\boldsymbol{\eta}^{(i)}), \dots, h_q(\boldsymbol{\eta}^{(i)})),$$

wobei $h : \mathbb{R}^q \rightarrow \mathbb{R}^q$ eine hinreichend oft differenzierbare Responsefunktion ist.

Aus der Dichte der Multinomialverteilung

$$f(y_{i1}, \dots, y_{iq}) = \frac{n_i!}{y_{i1}! \cdots y_{iq}! (n_i - y_{i1} - \cdots - y_{iq})!} \pi_{i1}^{y_{i1}} \cdots \\ \cdot \pi_{iq}^{y_{iq}} (1 - \pi_{i1} - \cdots - \pi_{iq})^{n_i - y_{i1} - \cdots - y_{iq}},$$

erhält man für die Beobachtungen y_1, \dots, y_g bei gegebenen Regressoren die Likelihoodfunktion

$$L(\boldsymbol{\beta}) = \prod_{i=1}^g f(\mathbf{y}_i^T) \\ = \prod_{i=1}^g c_i \pi_{i1}^{y_{i1}} \cdots \pi_{iq}^{y_{iq}} (1 - \pi_{i1} - \cdots - \pi_{iq})^{n_i - y_{i1} - \cdots - y_{iq}} \\ = \prod_{i=1}^g c_i \left(\frac{\pi_{i1}}{1 - \pi_{i1} - \cdots - \pi_{iq}} \right)^{y_{i1}} \cdots \left(\frac{\pi_{iq}}{1 - \pi_{i1} - \cdots - \pi_{iq}} \right)^{y_{iq}} \\ \cdot (1 - \pi_{i1} - \cdots - \pi_{iq})^{n_i},$$

wobei $c_i = n_i! / (y_{i1}! \cdot \dots \cdot y_{iq}! (n_i - y_{i1} - \dots - y_{iq})!)$. Einfacher ist es die logarithmierte Form zu betrachten. Mit $\pi_{ik} = 1 - \pi_{i1} - \dots - \pi_{iq}$ und $p_{ir} = y_{ir}/n_i$ erhält man die Log-Likelihoodfunktion

$$\begin{aligned}
l(\boldsymbol{\beta}) &= \log(L(\boldsymbol{\beta})) \\
&= \sum_{i=1}^g \left\{ \log(c_i) + y_{i1} \log \left(\frac{\pi_{i1}}{1 - \pi_{i1} - \dots - \pi_{iq}} \right) + \dots \right. \\
&\quad \left. + y_{iq} \log \left(\frac{\pi_{iq}}{1 - \pi_{i1} - \dots - \pi_{iq}} \right) + n_i \log(1 - \pi_{i1} - \dots - \pi_{iq}) \right\} \\
&= \sum_{i=1}^g \left\{ y_{i1} \log \left(\frac{\pi_{i1}}{\pi_{ik}} \right) + \dots + y_{iq} \log \left(\frac{\pi_{iq}}{\pi_{ik}} \right) + n_i \log \pi_{ik} + \log(c_i) \right\} \\
&= \sum_{i=1}^g \left\{ \sum_{r=1}^q y_{ir} \log \left(\frac{\pi_{ir}}{\pi_{ik}} \right) + n_i \log \pi_{ik} + \log(c_i) \right\} \\
&= \sum_{i=1}^g \left\{ \sum_{r=1}^q n_i \left[p_{ir} \log \left(\frac{\pi_{ir}}{\pi_{ik}} \right) + \log \pi_{ik} \right] + \log(c_i) \right\}.
\end{aligned}$$

Der Log-Likelihood lässt sich alternativ mit den Vektoren der Auftretenswahrscheinlichkeit $\boldsymbol{\pi}_i^T = (\pi_{i1}, \dots, \pi_{iq})$ darstellen durch

$$l(\boldsymbol{\beta}) = \sum_{i=1}^g l_i(\boldsymbol{\pi}_i),$$

wobei

$$l_i(\boldsymbol{\pi}_i) = n_i \left\{ \sum_{r=1}^q p_{ir} \log \left(\frac{\pi_{ir}}{\pi_{ik}} \right) + \log \pi_{ik} \right\} + \log(c_i). \quad (3.20)$$

Der Maximum-Likelihood-Schätzer ist definiert als derjenige Parametervektor $\hat{\boldsymbol{\beta}}$ für den die Funktion $L(\boldsymbol{\beta})$ bzw. $l(\boldsymbol{\beta})$ maximal wird. Betrachtet wird folglich die Ableitung von $l(\boldsymbol{\beta})$ nach jedem zu schätzenden Parametervektor $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_q$, die den Vektor $\boldsymbol{\beta}^T = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_q)$ bestimmen. Fasst man diese Ableitungen zu einem Vektor zusammen, erhält man die sogenannte *Scorefunktion*

$$\mathbf{s}^{(i)}(\boldsymbol{\beta}) = \begin{pmatrix} \mathbf{s}_1^{(i)}(\boldsymbol{\beta}) \\ \vdots \\ \mathbf{s}_q^{(i)}(\boldsymbol{\beta}) \end{pmatrix} = \begin{pmatrix} \frac{\partial l_i(\boldsymbol{\pi}_i)}{\partial \boldsymbol{\beta}_1} \\ \vdots \\ \frac{\partial l_i(\boldsymbol{\pi}_i)}{\partial \boldsymbol{\beta}_q} \end{pmatrix} \in \mathbb{R}^{(p+1)q}. \quad (3.21)$$

Der Maximum-Likelihood-Schätzer lässt sich wie folgt berechnen. Grundlage ist das allgemeine Modell $\pi_{ir} = h_r(\mathbf{Z}^{(i)} \boldsymbol{\beta})$. Da die Konstanten c_i nicht von dem zu schätzenden

Parametervektor $\boldsymbol{\beta}$ abhängen, können sie vernachlässigt werden. Damit ergibt sich

$$\begin{aligned} l(\boldsymbol{\beta}) &= \sum_{i=1}^g \left\{ \sum_{r=1}^q n_i \left[p_{ir} \log \left(\frac{h_r(\mathbf{Z}^{(i)} \boldsymbol{\beta})}{1 - \sum_{r=1}^q h_r(\mathbf{Z}^{(i)} \boldsymbol{\beta})} \right) + \log \left(1 - \sum_{r=1}^q h_r(\mathbf{Z}^{(i)} \boldsymbol{\beta}) \right) \right] \right\} \\ &= \sum_{i=1}^g \left\{ \sum_{r=1}^q n_i p_{ir} \log(h_r(\mathbf{Z}^{(i)} \boldsymbol{\beta})) + \left(n_i - \sum_{r=1}^q n_i p_{ir} \right) \log \left(1 - \sum_{r=1}^q h_r(\mathbf{Z}^{(i)} \boldsymbol{\beta}) \right) \right\} \end{aligned}$$

bzw.

$$l_i(\boldsymbol{\pi}_i) = \sum_{r=1}^q n_i p_{ir} \log(h_r(\mathbf{Z}^{(i)} \boldsymbol{\beta})) + \left(n_i - \sum_{r=1}^q n_i p_{ir} \right) \log \left(1 - \sum_{r=1}^q h_r(\mathbf{Z}^{(i)} \boldsymbol{\beta}) \right).$$

Elementares Differenzieren mit $h_r(\mathbf{Z}^{(i)} \boldsymbol{\beta}) = h_r(\boldsymbol{\eta}^{(i)})$ ergibt

$$\mathbf{s}_j^{(i)}(\boldsymbol{\beta}) = \frac{\partial l_i(\boldsymbol{\pi}_i)}{\partial \boldsymbol{\beta}_j} = \sum_{r=1}^q \frac{\partial l_i(\boldsymbol{\pi}_i)}{\partial h_r(\boldsymbol{\eta}^{(i)})} \cdot \frac{\partial h_r(\boldsymbol{\eta}^{(i)})}{\partial \boldsymbol{\beta}_j}, \quad j = 1, \dots, q. \quad (3.22)$$

Zunächst wird die Matrix der Ableitungen an der Stelle $\boldsymbol{\eta}^{(i)} = \mathbf{Z}^{(i)} \boldsymbol{\beta}$ definiert:

$$\mathbf{D}^{(i)}(\boldsymbol{\beta}) := \left. \frac{\partial h(\boldsymbol{\eta}^{(i)})}{\partial \boldsymbol{\eta}} \right|_{\boldsymbol{\eta}=\boldsymbol{\eta}^{(i)}} = \begin{pmatrix} \frac{\partial h_1(\boldsymbol{\eta}^{(i)})}{\partial \eta_1^{(i)}} & \dots & \frac{\partial h_q(\boldsymbol{\eta}^{(i)})}{\partial \eta_1^{(i)}} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_1(\boldsymbol{\eta}^{(i)})}{\partial \eta_q^{(i)}} & \dots & \frac{\partial h_q(\boldsymbol{\eta}^{(i)})}{\partial \eta_q^{(i)}} \end{pmatrix} := \begin{pmatrix} d_{11}^{(i)} & \dots & d_{1q}^{(i)} \\ \vdots & & \vdots \\ d_{q1}^{(i)} & \dots & d_{qq}^{(i)} \end{pmatrix} \quad (3.23)$$

$\mathbf{D}^{(i)}$ ist eine $q \times q$ -Matrix. Die r -te Zeile wird definiert als

$$\mathbf{D}_r^{(i)}(\boldsymbol{\beta}) := (d_{r1}^{(i)}, \dots, d_{rq}^{(i)}), \quad (3.24)$$

die r -te Spalte als

$$\mathbf{D}_{\cdot r}^{(i)}(\boldsymbol{\beta}) := \begin{pmatrix} d_{1r}^{(i)} \\ \vdots \\ d_{qr}^{(i)} \end{pmatrix}. \quad (3.25)$$

Die Ableitung von h_r nach dem Vektor $\boldsymbol{\beta}_j$ wird definiert durch

$$\begin{aligned} \frac{\partial h_r(\boldsymbol{\eta}^{(i)})}{\partial \boldsymbol{\beta}_j} &= \frac{\partial h_r(\eta_1^{(i)}, \dots, \eta_q^{(i)})}{\partial \boldsymbol{\beta}_j} = \sum_{m=1}^q \frac{\partial h_r(\boldsymbol{\eta}^{(i)})}{\partial \eta_m^{(i)}} \cdot \frac{\partial \eta_m^{(i)}}{\partial \boldsymbol{\beta}_j} \\ &\stackrel{(3.19), (3.23)}{=} \begin{cases} \mathbf{x}_i^T \sum_{m=1}^q d_{mr}^{(i)} & \text{falls } j = m \\ \mathbf{0} & \text{sonst} \end{cases} \stackrel{(3.17)}{=} \begin{cases} \sum_{m=1}^q \mathbf{z}_{mj}^{(i)} d_{mr}^{(i)} & \text{falls } j = m \\ \mathbf{0} & \text{sonst} \end{cases} \\ &= \underbrace{\mathbf{z}_{1j}^{(i)} d_{1r}^{(i)} + \dots + \mathbf{z}_{j-1,j}^{(i)} d_{j-1,r}^{(i)}}_{=0} + \mathbf{z}_{jj}^{(i)} d_{jr}^{(i)} + \underbrace{\mathbf{z}_{j+1,j}^{(i)} d_{j+1,r}^{(i)} + \dots + \mathbf{z}_{qj}^{(i)} d_{qr}^{(i)}}_{=0} \\ &= (\mathbf{z}_{1j}^{(i)}, \dots, \mathbf{z}_{qj}^{(i)}) \begin{pmatrix} d_{1r}^{(i)} \\ \vdots \\ d_{qr}^{(i)} \end{pmatrix} \stackrel{(3.18), (3.24)}{=} (\mathbf{Z}_{\cdot j}^{(i)})^T \cdot \mathbf{D}_{\cdot r}^{(i)}(\boldsymbol{\beta}). \quad (3.26) \end{aligned}$$

Dabei wurde die Kettenregel (vgl. Satz 6.9, Seite 570 und Folgerung 6.5, Seite 571 f. in Burg/Haf/Wille [7]) angewandt. Damit folgt für (3.22)

$$\begin{aligned}
\mathbf{s}_j^{(i)}(\boldsymbol{\beta}) &\stackrel{(3.21)}{=} \frac{\partial l_i(\boldsymbol{\pi}_i)}{\partial \beta_j} \\
&= \sum_{r=1}^q n_i p_{ir} \frac{\partial}{\partial \beta_j} (\log(h_r(\boldsymbol{\eta}^{(i)}))) + \left(n_i - \sum_{r=1}^q n_i p_{ir} \right) \frac{\partial}{\partial \beta_j} \left[\log \left(1 - \sum_{r=1}^q h_r(\boldsymbol{\eta}^{(i)}) \right) \right] \\
&= \sum_{r=1}^q \frac{n_i p_{ir}}{h_r(\boldsymbol{\eta}^{(i)})} \frac{\partial h_r(\boldsymbol{\eta}^{(i)})}{\partial \beta_j} + \frac{n_i - \sum_{r=1}^q n_i p_{ir}}{1 - \sum_{r=1}^q h_r(\boldsymbol{\eta}^{(i)})} \frac{\partial}{\partial \beta_j} \left(1 - \sum_{r=1}^q h_r(\boldsymbol{\eta}^{(i)}) \right) \\
&= \sum_{r=1}^q \frac{n_i p_{ir}}{h_r(\boldsymbol{\eta}^{(i)})} \frac{\partial h_r(\boldsymbol{\eta}^{(i)})}{\partial \beta_j} - \frac{n_i - \sum_{r=1}^q n_i p_{ir}}{1 - \sum_{r=1}^q h_r(\boldsymbol{\eta}^{(i)})} \sum_{r=1}^q \frac{\partial h_r(\boldsymbol{\eta}^{(i)})}{\partial \beta_j} \\
&\stackrel{(3.26)}{=} \sum_{r=1}^q \frac{n_i p_{ir}}{h_r(\boldsymbol{\eta}^{(i)})} (\mathbf{Z}_{\cdot j}^{(i)})^T \cdot \mathbf{D}_{\cdot r}^{(i)}(\boldsymbol{\beta}) - \frac{n_i - \sum_{r=1}^q n_i p_{ir}}{1 - \sum_{r=1}^q h_r(\boldsymbol{\eta}^{(i)})} \sum_{r=1}^q (\mathbf{Z}_{\cdot j}^{(i)})^T \cdot \mathbf{D}_{\cdot r}^{(i)}(\boldsymbol{\beta}) \\
&= \sum_{r=1}^q (\mathbf{Z}_{\cdot j}^{(i)})^T \cdot \mathbf{D}_{\cdot r}^{(i)}(\boldsymbol{\beta}) \left\{ \frac{n_i p_{ir}}{h_r(\boldsymbol{\eta}^{(i)})} - \frac{n_i - \sum_{r=1}^q n_i p_{ir}}{1 - \sum_{r=1}^q h_r(\boldsymbol{\eta}^{(i)})} \right\} \\
&= (\mathbf{Z}_{\cdot j}^{(i)})^T \sum_{r=1}^q \mathbf{D}_{\cdot r}^{(i)}(\boldsymbol{\beta}) \left\{ \frac{n_i p_{ir}}{h_r(\boldsymbol{\eta}^{(i)})} - \frac{n_i - \sum_{r=1}^q n_i p_{ir}}{1 - \sum_{r=1}^q h_r(\boldsymbol{\eta}^{(i)})} \right\}, \quad j = 1, \dots, q \quad (3.27)
\end{aligned}$$

d.h. die Scorefunktion (3.21) lautet

$$\begin{aligned}
\mathbf{s}^{(i)}(\boldsymbol{\beta}) &= \begin{pmatrix} \mathbf{s}_1^{(i)}(\boldsymbol{\beta}) \\ \vdots \\ \mathbf{s}_q^{(i)}(\boldsymbol{\beta}) \end{pmatrix} \\
&\stackrel{(3.27)}{=} \begin{pmatrix} (\mathbf{Z}_{\cdot 1}^{(i)})^T \sum_{r=1}^q \mathbf{D}_{\cdot r}^{(i)}(\boldsymbol{\beta}) \left\{ \frac{n_i p_{ir}}{h_r(\boldsymbol{\eta}^{(i)})} - \frac{n_i - \sum_{r=1}^q n_i p_{ir}}{1 - \sum_{r=1}^q h_r(\boldsymbol{\eta}^{(i)})} \right\} \\ \vdots \\ (\mathbf{Z}_{\cdot q}^{(i)})^T \sum_{r=1}^q \mathbf{D}_{\cdot r}^{(i)}(\boldsymbol{\beta}) \left\{ \frac{n_i p_{ir}}{h_r(\boldsymbol{\eta}^{(i)})} - \frac{n_i - \sum_{r=1}^q n_i p_{ir}}{1 - \sum_{r=1}^q h_r(\boldsymbol{\eta}^{(i)})} \right\} \end{pmatrix}. \quad (3.28)
\end{aligned}$$

Mit $\pi_{ir} = h_r(\boldsymbol{\eta}^{(i)})$ ergibt sich aus (3.28)

$$\begin{aligned}
\mathbf{s}^{(i)}(\boldsymbol{\beta}) &= \begin{bmatrix} (\mathbf{Z}_{\cdot 1}^{(i)})^T \mathbf{D}_{\cdot 1}^{(i)}(\boldsymbol{\beta}) \cdots (\mathbf{Z}_{\cdot 1}^{(i)})^T \mathbf{D}_{\cdot q}^{(i)}(\boldsymbol{\beta}) \\ (\mathbf{Z}_{\cdot 2}^{(i)})^T \mathbf{D}_{\cdot 1}^{(i)}(\boldsymbol{\beta}) \cdots (\mathbf{Z}_{\cdot 2}^{(i)})^T \mathbf{D}_{\cdot q}^{(i)}(\boldsymbol{\beta}) \\ \vdots \\ (\mathbf{Z}_{\cdot q}^{(i)})^T \mathbf{D}_{\cdot 1}^{(i)}(\boldsymbol{\beta}) \cdots (\mathbf{Z}_{\cdot q}^{(i)})^T \mathbf{D}_{\cdot q}^{(i)}(\boldsymbol{\beta}) \end{bmatrix} \underbrace{\begin{bmatrix} \frac{n_i p_{i1}}{\pi_{i1}} - \frac{n_i - \sum_{r=1}^q n_i p_{ir}}{1 - \sum_{r=1}^q \pi_{ir}} \\ \frac{n_i p_{i2}}{\pi_{i2}} - \frac{n_i - \sum_{r=1}^q n_i p_{ir}}{1 - \sum_{r=1}^q \pi_{ir}} \\ \vdots \\ \frac{n_i p_{iq}}{\pi_{iq}} - \frac{n_i - \sum_{r=1}^q n_i p_{ir}}{1 - \sum_{r=1}^q \pi_{ir}} \end{bmatrix}}_{=\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\beta})(\mathbf{p}_i - \boldsymbol{\pi}_i)} \\
&= \begin{pmatrix} (\mathbf{Z}_{\cdot 1}^{(i)})^T \\ (\mathbf{Z}_{\cdot 2}^{(i)})^T \\ \vdots \\ (\mathbf{Z}_{\cdot q}^{(i)})^T \end{pmatrix} (\mathbf{D}_{\cdot 1}^{(i)}(\boldsymbol{\beta}), \dots, \mathbf{D}_{\cdot q}^{(i)}(\boldsymbol{\beta})) \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\beta})(\mathbf{p}_i - \boldsymbol{\pi}_i) \\
&\stackrel{(3.18), (3.25)}{=} \begin{pmatrix} \mathbf{z}_{11}^{(i)} & \cdots & \mathbf{z}_{q1}^{(i)} \\ \vdots & & \vdots \\ \mathbf{z}_{1q}^{(i)} & \cdots & \mathbf{z}_{qq}^{(i)} \end{pmatrix} \begin{pmatrix} d_{11}^{(i)} & \cdots & d_{1q}^{(i)} \\ \vdots & & \vdots \\ d_{q1}^{(i)} & \cdots & d_{qq}^{(i)} \end{pmatrix} \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\beta})(\mathbf{p}_i - \boldsymbol{\pi}_i) \\
&\stackrel{(3.17), (3.23)}{=} (\mathbf{Z}^{(i)})^T \mathbf{D}^{(i)}(\boldsymbol{\beta}) \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\beta})(\mathbf{p}_i - \boldsymbol{\pi}_i),
\end{aligned}$$

falls

$$\begin{aligned}
&\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\beta})(\mathbf{p}_i - \boldsymbol{\pi}_i) \\
&= \begin{pmatrix} \frac{n_i p_{i1}}{\pi_{i1}} - (n_i - \sum_{r=1}^q n_i p_{ir}) \frac{1}{1 - \sum_{r=1}^q \pi_{ir}} \\ \vdots \\ \frac{n_i p_{iq}}{\pi_{iq}} - (n_i - \sum_{r=1}^q n_i p_{ir}) \frac{1}{1 - \sum_{r=1}^q \pi_{ir}} \end{pmatrix} = n_i \begin{pmatrix} \frac{p_{i1}}{\pi_{i1}} - \frac{1 - \sum_{r=1}^q p_{ir}}{1 - \sum_{r=1}^q \pi_{ir}} \\ \vdots \\ \frac{p_{iq}}{\pi_{iq}} - \frac{1 - \sum_{r=1}^q p_{ir}}{1 - \sum_{r=1}^q \pi_{ir}} \end{pmatrix} \quad (3.29)
\end{aligned}$$

gilt, wobei

$$\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\beta}) = n_i \begin{pmatrix} \frac{1}{\pi_{i1}} + \frac{1}{\alpha} & \frac{1}{\alpha} & \frac{1}{\alpha} & \cdots & \frac{1}{\alpha} \\ \frac{1}{\alpha} & \frac{1}{\pi_{i2}} + \frac{1}{\alpha} & \frac{1}{\alpha} & \cdots & \frac{1}{\alpha} \\ \frac{1}{\alpha} & \frac{1}{\alpha} & \frac{1}{\pi_{i3}} + \frac{1}{\alpha} & \cdots & \frac{1}{\alpha} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\alpha} & \frac{1}{\alpha} & \frac{1}{\alpha} & \cdots & \frac{1}{\pi_{iq}} + \frac{1}{\alpha} \end{pmatrix} \quad (3.30)$$

mit $1/\alpha = 1/(1 - \sum_{r=1}^q \pi_{ir})$ die Inverse der Kovarianzmatrix

$$\Sigma_i(\boldsymbol{\beta}) = \frac{1}{n_i} \begin{pmatrix} \pi_{i1}(1 - \pi_{i1}) & -\pi_{i1}\pi_{i2} & \dots & -\pi_{i1}\pi_{iq} \\ -\pi_{i2}\pi_{i1} & \pi_{i2}(1 - \pi_{i2}) & & \vdots \\ \vdots & & \ddots & \\ -\pi_{iq}\pi_{i1} & \dots & & \pi_{iq}(1 - \pi_{iq}) \end{pmatrix} \quad (3.31)$$

von \mathbf{p}_i ist (vgl. Abschnitt 8.3 Tutz (1990) [26]).

Beweis. Da \mathbf{y}_i multinomialverteilt ist und nach Modellannahme $\mathbf{y}_i \sim M(n_i, \boldsymbol{\pi}_i)$ gilt, erhält man

$$\text{Var}(\mathbf{y}_i) = n_i(\text{diag}(\boldsymbol{\pi}_i) - \boldsymbol{\pi}_i\boldsymbol{\pi}_i^T) \quad (3.32)$$

bzw.

$$\text{Var}(\mathbf{p}_i) = \text{Var}\left(\frac{\mathbf{y}_i}{n_i}\right) = \frac{1}{n_i}(\text{diag}(\boldsymbol{\pi}_i) - \boldsymbol{\pi}_i\boldsymbol{\pi}_i^T).$$

Zunächst wird gezeigt, dass $\Sigma_i(\boldsymbol{\beta}) \cdot \Sigma_i^{-1}(\boldsymbol{\beta}) = \mathbf{I}_q$ die $q \times q$ -dimensionale Einheitsmatrix beschreibt.

$$\begin{aligned}
\Sigma_i(\beta) \cdot \Sigma_i^{-1}(\beta) &= \begin{pmatrix} \pi_{i1}(1 - \pi_{i1}) & -\pi_{i1}\pi_{i2} & \dots & -\pi_{i1}\pi_{iq} \\ -\pi_{i2}\pi_{i1} & \pi_{i2}(1 - \pi_{i2}) & & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ -\pi_{iq}\pi_{i1} & \dots & & \pi_{iq}(1 - \pi_{iq}) \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{\pi_{i1}} + \frac{1}{1 - \sum_{r=1}^q \pi_{ir}} & \dots & \dots & \frac{1}{1 - \sum_{r=1}^q \pi_{ir}} \\ \frac{1}{1 - \sum_{r=1}^q \pi_{ir}} & \frac{1}{1 - \sum_{r=1}^q \pi_{ir}} & \dots & \frac{1}{1 - \sum_{r=1}^q \pi_{ir}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{1 - \sum_{r=1}^q \pi_{ir}} & \frac{1}{1 - \sum_{r=1}^q \pi_{ir}} & \dots & \frac{1}{1 - \sum_{r=1}^q \pi_{ir}} + \frac{1}{\pi_{iq}} \end{pmatrix} \\
&= \begin{pmatrix} \frac{\pi_{i1}(1 - \pi_{i1})}{\pi_{i1}} + \frac{\pi_{i1}(1 - \pi_{i1}) - \pi_{i1}\pi_{i2} - \dots - \pi_{i1}\pi_{iq}}{1 - \sum_{r=1}^q \pi_{ir}} & \dots, & \dots, & \frac{\pi_{i1}\pi_{iq}}{\pi_{iq}} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\pi_{iq}\pi_{i1}}{\pi_{i1}} + \frac{\pi_{iq}(1 - \pi_{iq}) - \pi_{iq}\pi_{i1} - \pi_{iq}\pi_{i2} - \dots - \pi_{iq}\pi_{i,q-1}}{1 - \sum_{r=1}^q \pi_{ir}} & \dots, & \dots, & \frac{\pi_{iq}(1 - \pi_{iq})}{\pi_{iq}} \end{pmatrix} \\
&= \begin{pmatrix} 1 - \pi_{i1} + \frac{\pi_{i1} - \pi_{i1} \sum_{r=1}^q \pi_{ir}}{1 - \sum_{r=1}^q \pi_{ir}} & \dots, & \dots, & \frac{1 - \sum_{r=1}^q \pi_{ir}}{1 - \sum_{r=1}^q \pi_{ir}} - \pi_{i1} \\ \vdots & \vdots & \ddots & \vdots \\ -\pi_{iq} + \frac{\pi_{iq} - \pi_{iq} \sum_{r=1}^q \pi_{ir}}{1 - \sum_{r=1}^q \pi_{ir}} & \dots, & \dots, & \frac{1 - \sum_{r=1}^q \pi_{ir}}{1 - \sum_{r=1}^q \pi_{ir}} - \pi_{iq} \end{pmatrix} = \begin{pmatrix} 1 - \pi_{i1} + \pi_{i1} \frac{1 - \sum_{r=1}^q \pi_{ir}}{1 - \sum_{r=1}^q \pi_{ir}} & \dots, & \dots, & \frac{1 - \sum_{r=1}^q \pi_{ir}}{1 - \sum_{r=1}^q \pi_{ir}} - \pi_{i1} \\ \vdots & \vdots & \ddots & \vdots \\ -\pi_{iq} + \pi_{iq} \frac{1 - \sum_{r=1}^q \pi_{ir}}{1 - \sum_{r=1}^q \pi_{ir}} & \dots, & \dots, & 1 - \pi_{iq} + \frac{1 - \sum_{r=1}^q \pi_{ir}}{1 - \sum_{r=1}^q \pi_{ir}} \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{pmatrix} = \mathbf{I}_q
\end{aligned}$$

Im Folgenden soll Gleichung (3.29) bewiesen werden.

$$\begin{aligned} \Sigma_i^{-1}(\boldsymbol{\beta})(\mathbf{p}_i - \boldsymbol{\pi}_i) &= n_i \begin{pmatrix} \frac{1}{\pi_{i1}}(p_{i1} - \pi_{i1}) + \frac{1}{1 - \sum_{r=1}^q \pi_{ir}} \sum_{r=1}^q (p_{ir} - \pi_{ir}) \\ \vdots \\ \frac{1}{\pi_{iq}}(p_{iq} - \pi_{iq}) + \frac{1}{1 - \sum_{r=1}^q \pi_{ir}} \sum_{r=1}^q (p_{ir} - \pi_{ir}) \end{pmatrix} \\ &= n_i \begin{pmatrix} \frac{p_{i1}}{\pi_{i1}} - 1 + \frac{\sum_{r=1}^q (p_{ir} - \pi_{ir})}{1 - \sum_{r=1}^q \pi_{ir}} \\ \vdots \\ \frac{p_{iq}}{\pi_{iq}} - 1 + \frac{\sum_{r=1}^q (p_{ir} - \pi_{ir})}{1 - \sum_{r=1}^q \pi_{ir}} \end{pmatrix} = n_i \begin{pmatrix} \frac{p_{i1}}{\pi_{i1}} - \frac{1 - \sum_{r=1}^q p_{ir}}{1 - \sum_{r=1}^q \pi_{ir}} \\ \vdots \\ \frac{p_{iq}}{\pi_{iq}} - \frac{1 - \sum_{r=1}^q p_{ir}}{1 - \sum_{r=1}^q \pi_{ir}} \end{pmatrix} \end{aligned}$$

□

Damit hat die Scorefunktion die Form

$$\mathbf{s}^{(i)}(\boldsymbol{\beta}) = (\mathbf{Z}^{(i)})^T \mathbf{D}^{(i)}(\boldsymbol{\beta}) \Sigma_i^{-1}(\boldsymbol{\beta})(\mathbf{p}_i - \boldsymbol{\pi}_i).$$

Wegen (3.32) gilt auch $\Sigma_i(\boldsymbol{\beta}) = (1/n_i^2) \text{Var}(\mathbf{y}_i)$.

Da $\mathbf{s}(\boldsymbol{\beta}) = \sum_{i=1}^g \mathbf{s}^{(i)}(\boldsymbol{\beta})$ hat die Scorefunktion in geschlossener Schreibweise die Form

$$\mathbf{s}(\boldsymbol{\beta}) = \sum_{i=1}^g (\mathbf{Z}^{(i)})^T \mathbf{D}^{(i)}(\boldsymbol{\beta}) \Sigma_i^{-1}(\boldsymbol{\beta})(\mathbf{p}_i - \boldsymbol{\pi}_i).$$

Zur Bestimmung der Maximum-Likelihood-Schätzer ist das Gleichungssystem

$$\mathbf{s}(\boldsymbol{\beta}) = \mathbf{0}$$

zu lösen. Da keine geschlossene Form der Lösung existiert, wird auf iterative Verfahren zurückgegriffen (vgl. Anhang A). Wichtig für die numerische Maximierung von $l(\boldsymbol{\beta})$ ist die zweite Ableitung der Log-Likelihoodfunktion. Sie liefert die beobachtete Informationsmatrix. Eine einfachere Form ergibt sich, wenn man den Erwartungswert bildet. Damit erhält man die Informations- oder Fishermatrix. Für das asymptotische Verhalten lässt sich die einfache Asymptotik des wachsenden Stichprobenumfangs zu Grunde legen ($n \rightarrow \infty$). Man erhält

$$F(\boldsymbol{\beta}) = \sum_{i=1}^g \mathbf{F}_i(\boldsymbol{\beta}) = \mathbb{E} \left(-\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right) = \sum_{i=1}^g (\mathbf{Z}^{(i)})^T \mathbf{D}^{(i)}(\boldsymbol{\beta}) \Sigma_i^{-1}(\boldsymbol{\beta}) \mathbf{D}^{(i)}(\boldsymbol{\beta})^T \mathbf{Z}^{(i)}.$$

Beweis. Der Erwartungswert von $\mathbf{s}(\boldsymbol{\beta})$ lässt sich wie folgt berechnen:

$$\begin{aligned} \mathbb{E}(\mathbf{s}(\boldsymbol{\beta})) &= \mathbb{E} \left(\sum_{i=1}^g (\mathbf{Z}^{(i)})^T \mathbf{D}^{(i)}(\boldsymbol{\beta}) \Sigma_i^{-1}(\boldsymbol{\beta})(\mathbf{p}_i - \boldsymbol{\pi}_i) \right) \\ &= \sum_{i=1}^g (\mathbf{Z}^{(i)})^T \mathbf{D}^{(i)}(\boldsymbol{\beta}) \Sigma_i^{-1}(\boldsymbol{\beta}) \underbrace{\mathbb{E}(\mathbf{p}_i - \boldsymbol{\pi}_i)}_{=0} = 0, \end{aligned}$$

da

$$\mathbb{E}(\mathbf{p}_i) = \mathbb{E}\left(\frac{\mathbf{y}_i}{n_i}\right) = \frac{n_i \boldsymbol{\pi}_i}{n_i} = \boldsymbol{\pi}_i. \quad (3.33)$$

Ohne Einschränkung der Allgemeinheit sei hier $l(\boldsymbol{\beta}) = \log f(\mathbf{y})$ angenommen. Unter den Regularitätsbedingungen (vgl. Abschnitt 4.3 in Bickel und Doksum [5]) gilt

$$\begin{aligned} \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \log f(\mathbf{y}) &= \frac{\partial}{\partial \boldsymbol{\beta}^T} \frac{\frac{\partial}{\partial \boldsymbol{\beta}} f(\mathbf{y})}{f(\mathbf{y})} = \frac{f(\mathbf{y}) \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} f(\mathbf{y}) - \left(\frac{\partial}{\partial \boldsymbol{\beta}} f(\mathbf{y}) \frac{\partial}{\partial \boldsymbol{\beta}^T} f(\mathbf{y}) \right)}{[f(\mathbf{y})]^2} \\ &= \frac{1}{f(\mathbf{y})} \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} f(\mathbf{y}) - \frac{\partial}{\partial \boldsymbol{\beta}} \log f(\mathbf{y}) \frac{\partial}{\partial \boldsymbol{\beta}^T} \log f(\mathbf{y}). \end{aligned}$$

Damit ergibt sich

$$\begin{aligned} \mathbb{E}\left(-\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}\right) &= -\int \cdots \int \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \log f(\mathbf{y}) \cdot f(\mathbf{y}) d\mathbf{y} \\ &= -\frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \underbrace{\int f(\mathbf{y}) d\mathbf{y}}_{=1} + \int \underbrace{\frac{\partial}{\partial \boldsymbol{\beta}} \log f(\mathbf{y})}_{= \mathbf{s}(\boldsymbol{\beta})} \underbrace{\frac{\partial}{\partial \boldsymbol{\beta}^T} \log f(\mathbf{y})}_{= (\mathbf{s}(\boldsymbol{\beta}))^T} \cdot f(\mathbf{y}) d\mathbf{y} \\ &= \mathbb{E}(\mathbf{s}(\boldsymbol{\beta})(\mathbf{s}(\boldsymbol{\beta}))^T) \stackrel{\mathbb{E}(\mathbf{s}(\boldsymbol{\beta}))=0}{=} \text{Cov}(\mathbf{s}(\boldsymbol{\beta})) = F(\boldsymbol{\beta}). \end{aligned}$$

Damit kann $\mathbf{F}_i(\boldsymbol{\beta})$ bestimmt werden.

$$\begin{aligned} \mathbf{F}_i(\boldsymbol{\beta}) &= \text{Cov}(\mathbf{s}^{(i)}(\boldsymbol{\beta})) = \mathbb{E}(\mathbf{s}^{(i)}(\boldsymbol{\beta})(\mathbf{s}^{(i)}(\boldsymbol{\beta}))^T) \\ &= \mathbb{E}\left[\left(\mathbf{Z}^{(i)}\right)^T \mathbf{D}^{(i)}(\boldsymbol{\beta}) \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\beta}) (\mathbf{p}_i - \boldsymbol{\pi}_i) (\mathbf{p}_i - \boldsymbol{\pi}_i)^T (\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\beta}))^T (\mathbf{D}^{(i)}(\boldsymbol{\beta}))^T \mathbf{Z}^{(i)}\right] \\ &= \left(\mathbf{Z}^{(i)}\right)^T \mathbf{D}^{(i)}(\boldsymbol{\beta}) \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\beta}) \mathbb{E}[(\mathbf{p}_i - \boldsymbol{\pi}_i) (\mathbf{p}_i - \boldsymbol{\pi}_i)^T] \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\beta}) (\mathbf{D}^{(i)}(\boldsymbol{\beta}))^T \mathbf{Z}^{(i)} \\ &\stackrel{(3.33)}{=} \left(\mathbf{Z}^{(i)}\right)^T \mathbf{D}^{(i)}(\boldsymbol{\beta}) \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\beta}) \text{Cov}(\mathbf{p}_i) \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\beta}) (\mathbf{D}^{(i)}(\boldsymbol{\beta}))^T \mathbf{Z}^{(i)} \\ &\stackrel{(3.31)}{=} \left(\mathbf{Z}^{(i)}\right)^T \mathbf{D}^{(i)}(\boldsymbol{\beta}) \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\beta}) \boldsymbol{\Sigma}_i(\boldsymbol{\beta}) \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\beta}) (\mathbf{D}^{(i)}(\boldsymbol{\beta}))^T \mathbf{Z}^{(i)} \\ &= \left(\mathbf{Z}^{(i)}\right)^T \mathbf{D}^{(i)}(\boldsymbol{\beta}) \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\beta}) (\mathbf{D}^{(i)}(\boldsymbol{\beta}))^T \mathbf{Z}^{(i)} \end{aligned}$$

Dabei wurde die Symmetrie von $\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\beta})$ ausgenutzt. Folglich ergibt sich

$$\begin{aligned} \mathbf{F}(\boldsymbol{\beta}) &= \sum_{i=1}^g \mathbf{F}_i(\boldsymbol{\beta}) = \mathbb{E}\left(-\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}\right) = \text{Cov}(\mathbf{s}(\boldsymbol{\beta})) = \sum_{i=1}^g \text{Cov}(\mathbf{s}^{(i)}(\boldsymbol{\beta})) = \mathbb{E}(\mathbf{s}(\boldsymbol{\beta}) \mathbf{s}(\boldsymbol{\beta})^T) \\ &= \sum_{i=1}^g \mathbb{E}(\mathbf{s}^{(i)}(\boldsymbol{\beta}) \mathbf{s}^{(i)}(\boldsymbol{\beta})^T) = \sum_{i=1}^g \left(\mathbf{Z}^{(i)}\right)^T \mathbf{D}^{(i)}(\boldsymbol{\beta}) \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{\beta}) \mathbf{D}^{(i)}(\boldsymbol{\beta})^T \mathbf{Z}^{(i)}. \end{aligned}$$

□

Unter schwachen Bedingungen (vgl. Fahrmeir und Kaufmann (1985) Seite 348 f. [12]) gilt, dass der Maximum-Likelihood-Schätzer für wachsenden Stichprobenumfang asymptotisch

existiert, konsistent ist, d.h. $\hat{\boldsymbol{\beta}} \rightarrow \boldsymbol{\beta}$ in Wahrscheinlichkeit und gegen die Normalverteilung konvergiert. Damit ergibt sich für $n \rightarrow \infty$ folgende Approximation

$$\begin{array}{c} \hat{\boldsymbol{\beta}} \stackrel{a}{\sim} N(\boldsymbol{\beta}, \mathbf{F}(\hat{\boldsymbol{\beta}})^{-1}) \\ \text{mit} \\ \mathbf{F}(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^g (\mathbf{Z}^{(i)})^T \mathbf{D}^{(i)}(\hat{\boldsymbol{\beta}}) \boldsymbol{\Sigma}_i^{-1}(\hat{\boldsymbol{\beta}}) \mathbf{D}^{(i)}(\hat{\boldsymbol{\beta}})^T \mathbf{Z}^{(i)}, \end{array} \quad (3.34)$$

d.h. $\hat{\boldsymbol{\beta}}$ ist asymptotisch normalverteilt mit asymptotischer Kovarianzmatrix $\text{Cov}(\hat{\boldsymbol{\beta}}) \stackrel{a}{=} \mathbf{F}(\hat{\boldsymbol{\beta}})^{-1}$, wobei $\mathbf{F}(\hat{\boldsymbol{\beta}})^{-1}$ die Inverse der Fishermatrix beschreibt und $\mathbf{F}(\hat{\boldsymbol{\beta}}) \rightarrow \mathbf{F}(\boldsymbol{\beta})$ in Wahrscheinlichkeit.

Bemerkung 3.7.1. Es liegt nahe, das multinomiale Logit-Modell durch $k - 1$ separate logistische Regressionsmodelle anzupassen. Nach (3.6) kann das multinomiale Modell mit k Kategorien durch

$$\log \left(\frac{P(Y = r | \mathbf{x})}{P(Y = k | \mathbf{x})} \right) = \mathbf{x}^T \boldsymbol{\beta}_r, \quad r = 1, \dots, k - 1$$

beschrieben werden, wobei k als Referenzkategorie gewählt wurde. Im Folgenden werden die Fälle betrachtet, für die $Y = k$ oder $Y = r$ ist, d.h. das Problem reduziert sich auf den binären Fall

$$\begin{aligned} & \log \left(\frac{P(Y = r | \mathbf{x}, Y \in (k, r))}{1 - P(Y = r | \mathbf{x}, Y \in (k, r))} \right) = \log \left(\frac{P(Y = r | \mathbf{x}, Y \in (k, r))}{P(Y = k | \mathbf{x}, Y \in (k, r))} \right) \\ & = \log \left(\frac{P(Y = r | \mathbf{x})}{P(Y = k | \mathbf{x})} \right) = \mathbf{x}^T \boldsymbol{\beta}_r, \quad r = 1, \dots, k - 1. \end{aligned}$$

Der letzte Ausdruck entspricht genau der Annahme für das multinomiale Logit-Modell. Im Weiteren können für $r = 1$ die Koeffizienten der logarithmierten Chance

$$\text{Logit}_{1,k} = \log \left(\frac{\pi_1}{\pi_k} \right)$$

geschätzt werden, indem $Y = 1$ und $Y = k$ durch Erfolg und Misserfolg ausgedrückt werden, während die übrigen Responsekategorien $r = 2, \dots, k - 1$ ignoriert werden. Dieses Vorgehen würde für die verbleibenden logarithmierten Chancen

$$\text{Logit}_{r,k-1} = \log \left(\frac{\pi_r}{\pi_k} \right), \quad r = 2, \dots, k - 1$$

analog durchgeführt werden. Ein solches Vorgehen ist nicht möglich. Die Modelle sind im statistischen Sinn nicht unabhängig. Die Parameterschätzungen für $k - 1$ unabhängige Logit-Modelle wären in dem Fall nicht korrekt.

Bemerkung 3.7.2. Außerdem liegt nahe, dass die Likelihoodfunktion $L(\boldsymbol{\beta})$ des multinomialen Modells gleich dem Produkt der Likelihoodfunktionen der $k - 1 = g$ Binomialmodelle ist. Das dies nicht der Fall ist, wird im Folgenden kurz skizziert.

$$\begin{aligned}
\prod_{r=1}^g L(\boldsymbol{\beta}_r) &= \prod_{r=1}^g \prod_{i=1}^g \binom{n_{ir} + n_{ik}}{n_{ir}} \pi_{ir}^{n_{ir}} (1 - \pi_{ir})^{n_{ir} + n_{ik} - n_{ir}} \\
&= \prod_{i=1}^g \prod_{r=1}^g \binom{n_{ir} + n_{ik}}{n_{ir}} \pi_{ir}^{n_{ir}} (1 - \pi_{ir})^{n_{ik}} \\
&= \prod_{i=1}^g \binom{n_{i1} + n_{ik}}{n_{i1}} \pi_{i1}^{n_{i1}} (1 - \pi_{i1})^{n_{ik}} \cdots \binom{n_{iq} + n_{ik}}{n_{iq}} \pi_{iq}^{n_{iq}} (1 - \pi_{iq})^{n_{ik}} \\
&= \prod_{i=1}^g \frac{(n_{i1} + n_{ik})!}{n_{i1}! n_{ik}!} \cdots \frac{(n_{iq} + n_{ik})!}{n_{iq}! n_{ik}!} \pi_{i1}^{n_{i1}} \cdots \pi_{iq}^{n_{iq}} [(1 - \pi_{i1}) \cdots (1 - \pi_{iq})]^{n_{ik}} \\
&= \prod_{i=1}^g \frac{(n_{i1} + n_{ik})! \cdots (n_{iq} + n_{ik})!}{n_{i1}! \cdots n_{iq}! (n_{ik}!)^g} \pi_{i1}^{n_{i1}} \cdots \pi_{iq}^{n_{iq}} [(1 - \pi_{i1}) \cdots (1 - \pi_{iq})]^{n_i - n_{i1} - \cdots - n_{iq}} \\
L(\boldsymbol{\beta}) &= \prod_{i=1}^g \frac{n_i!}{(n_{i1}! \cdots n_{iq}!) (n_i - n_{i1} - \cdots - n_{iq})!} \pi_{i1}^{n_{i1}} \cdots \pi_{iq}^{n_{iq}} \\
&\quad \cdot (1 - \pi_{i1} - \cdots - \pi_{iq})^{n_i - n_{i1} - \cdots - n_{iq}}
\end{aligned}$$

Damit ergibt sich, dass $L(\boldsymbol{\beta}) \neq \prod_{r=1}^{k-1} L(\boldsymbol{\beta}_r)$.

Fazit: Eine Ursache für die unterschiedlichen Parameterschätzungen liegt darin, dass im Fall der $k - 1$ Binomialmodelle Unabhängigkeit vorausgesetzt wird, die im multinomialen Modell nicht gegeben ist. Eine weitere Ursache liegt in der unterschiedlichen Berechnung der Maximum-Likelihoodfunktion für das binomiale und multinomiale Modell. Außerdem konnte gezeigt werden, dass der Likelihood des multinomialen Modells nicht dem Produkt der Likelihoods der $k - 1$ Binomialmodelle entspricht.

Beispiel 3.7.1. (*Theoretische Bestimmung der Maximum-Likelihood-Schätzer*)

Betrachtet werden erneut die Daten der Mobilitätsstudie '97. Die abhängige Variable Y besteht aus den vier Responsekategorien Fuß ($Y = 1$), Rad ($Y = 2$), öffentlicher Verkehr ($Y = 3$) und Individualverkehr ($Y = 4$). Als Einflussgröße dient die metrische Variable Alter (A). Als Referenzkategorie wird $k = 4$ (Individualverkehr) gewählt. Die Daten werden in sog. Kontingenztafeln dargestellt. Die allgemeine ($g \times 4$)-Kontingenztafel mit der Einflussgröße Alter hat die Form

		Y				
		1	2	3	4	
A lter	1	n_{11}	n_{12}	n_{13}	n_{14}	n_1
	2	n_{21}	\ddots		\vdots	\vdots
	\vdots	\vdots		\ddots	\vdots	
	g	n_{g1}	\dots		n_{g4}	n_g

wobei $n_i = \sum_{r=1}^k n_{ir}$, für $i = 1, \dots, g$ die jeweilige Zeilensumme bezeichnet. Zunächst sollen die Parameter des multinomialen Modells berechnet werden. y_i sei multinomialverteilt, d.h.

$$y_i \sim M(n_i, \boldsymbol{\pi}_i).$$

Die Log-Likelihoodfunktion $l_M(\boldsymbol{\beta})$ im multinomialen Modell mit $q = k - 1 = 3$ lautet dann

$$l_M(\boldsymbol{\beta}) = \sum_{i=1}^g \left\{ \sum_{r=1}^3 n_{ir} \log \left(\frac{\pi_{ir}}{\pi_{ik}} \right) + n_i \log \pi_{ik} + \log(c_i) \right\}. \quad (3.35)$$

Der Term $\log(c_i)$ kann ignoriert werden. Folglich ergibt sich mit $k = 4$ als Referenzkategorie

$$l_M(\boldsymbol{\beta}) = \sum_{i=1}^g \left\{ \sum_{r=1}^3 n_{ir} \log \left(\frac{P(Y = r | A_i)}{P(Y = 4 | A_i)} \right) \right\} + \sum_{i=1}^g n_i \log(P(Y = 4 | A_i)).$$

Setzt man in diese Gleichung (3.4) und (3.5) ein, ergibt sich

$$l_M(\boldsymbol{\beta}) = \sum_{i=1}^g \left\{ \sum_{r=1}^3 n_{ir} (\beta_{0r} + \beta_{A,r} A_i) \right\} - \sum_{i=1}^g n_i \log \left(1 + \sum_{r=1}^3 \exp(\beta_{0r} + \beta_{A,r} A_i) \right).$$

Die Maximum-Likelihood-Schätzer erfüllen folgende Gleichungen:

$$\frac{\partial l_M(\boldsymbol{\beta})}{\partial \beta_{0r}} = \sum_{i=1}^g n_{ir} - \sum_{i=1}^g n_i \frac{\exp(\beta_{0r} + \beta_{A,r} A_i)}{1 + \sum_{r=1}^3 \exp(\beta_{0r} + \beta_{A,r} A_i)} = 0, \quad r = 1, \dots, 3$$

bzw.

$$\frac{\partial l_M(\boldsymbol{\beta})}{\partial \beta_{A,r}} = \sum_{i=1}^g A_i n_{ir} - \sum_{i=1}^g n_i \frac{\exp(\beta_{0r} + \beta_{A,r} A_i) A_i}{1 + \sum_{r=1}^3 \exp(\beta_{0r} + \beta_{A,r} A_i)} = 0, \quad r = 1, \dots, 3.$$

Sei nun y_i binomialverteilt, d.h.

$$y_i \sim B(n_{ir} + n_{ik}, \pi(\mathbf{x}_i)).$$

Das multinomiale Modell soll durch drei Binomialmodelle angepasst werden. Die Log-Likelihoodfunktion eines Binomialmodells lässt sich dann wie folgt darstellen

$$\begin{aligned} l_B(\boldsymbol{\beta}) &= \sum_{i=1}^g \left\{ n_{ir} \log \left(\frac{\pi_{ir}}{1 - \pi_{ir}} \right) + (n_{ir} + n_{ik}) \log(1 - \pi_{ir}) + \log \binom{n_{ir} + n_{ik}}{n_{ir}} \right\} \\ &= \sum_{i=1}^g \left\{ n_{ir} \log \left(\frac{P(Y = r | A_i)}{P(Y = 4 | A_i)} \right) + (n_{ir} + n_{ik}) \log(P(Y = 4 | A_i)) + \log \binom{n_{ir} + n_{ik}}{n_{ir}} \right\}, \end{aligned}$$

wobei r eine der drei Responsekategorien Fuß ($Y = 1$), Rad ($Y = 2$) oder öffentlicher Verkehr ($Y = 3$) beschreibt. Der Term $\log \binom{n_{ir} + n_{ik}}{n_{ir}}$ kann ignoriert werden. Setzt man in die letzte Gleichung wieder (3.4) und (3.5) ein, ergibt sich

$$l_B(\boldsymbol{\beta}) = \sum_{i=1}^g n_{ir} (\beta_{0r} + \beta_{A,r} A_i) - \sum_{i=1}^g (n_{ir} + n_{ik}) \log(1 + \exp(\beta_{0r} + \beta_{A,r} A_i)).$$

Die Maximum-Likelihood-Schätzer lassen sich durch Lösen der Gleichungen

$$\frac{\partial l_B(\boldsymbol{\beta})}{\partial \beta_{0r}} = \sum_{i=1}^g n_{ir} - \sum_{i=1}^g (n_{ir} + n_{ik}) \frac{\exp(\beta_{0r} + \beta_{A,r} A_i)}{1 + \exp(\beta_{0r} + \beta_{A,r} A_i)} = 0, \quad r = 1, \dots, 3$$

und

$$\frac{\partial l_B(\boldsymbol{\beta})}{\partial \beta_{A,r}} = \sum_{i=1}^g A_i n_{ir} - \sum_{i=1}^g (n_{ir} + n_{ik}) \frac{\exp(\beta_{0r} + \beta_{A,r} A_i) A_i}{1 + \exp(\beta_{0r} + \beta_{A,r} A_i)} = 0, \quad r = 1, \dots, 3$$

finden.

Bereits ein Vergleich der Ableitungen der Log-Likelihoodfunktion des multinomialen Modells mit den Log-Likelihoodfunktionen der drei Binomialmodelle zeigt deutlich, dass die Schätzungen nicht exakt dieselben Werte liefern können:

$$\begin{aligned} \frac{\partial l_M(\boldsymbol{\beta})}{\partial \beta_{0r}} &= \sum_{i=1}^g n_{ir} - \sum_{i=1}^g n_i \frac{\exp(\beta_{0r} + \beta_{A,r} A_i)}{1 + \sum_{r=1}^3 \exp(\beta_{0r} + \beta_{A,r} A_i)} \neq \\ &\sum_{i=1}^g n_{ir} - \sum_{i=1}^g (n_{ir} + n_{ik}) \frac{\exp(\beta_{0r} + \beta_{A,r} A_i)}{1 + \exp(\beta_{0r} + \beta_{A,r} A_i)} = \frac{\partial l_B(\boldsymbol{\beta})}{\partial \beta_{0r}} \end{aligned}$$

bzw.

$$\begin{aligned} \frac{\partial l_M(\boldsymbol{\beta})}{\partial \beta_{A,r}} &= \sum_{i=1}^g A_i n_{ir} - \sum_{i=1}^g n_i \frac{\exp(\beta_{0r} + \beta_{A,r} A_i) A_i}{1 + \sum_{r=1}^3 \exp(\beta_{0r} + \beta_{A,r} A_i)} \neq \\ &\sum_{i=1}^g A_i n_{ir} - \sum_{i=1}^g (n_{ir} + n_{ik}) \frac{\exp(\beta_{0r} + \beta_{A,r} A_i) A_i}{1 + \exp(\beta_{0r} + \beta_{A,r} A_i)} = \frac{\partial l_B(\boldsymbol{\beta})}{\partial \beta_{A,r}} \end{aligned}$$

Die Parameterschätzungen beider Modelle lauten:

	Multinomiales Modell		Binomialmodelle	
	Intercept	Alter	Intercept	Alter
Fuß	-1.327	0.015	-1.366	0.016
Std. Errors	0.160	0.003	0.165	0.003
ÖV	-2.219	0.026	-2.234	0.026
Std. Errors	0.189	0.004	0.191	0.004
Rad	1.675	0.003	-1.700	0.004
Std. Errors	0.217	0.005	0.233	0.005

Tabelle 7: Parameterschätzer für Alter

Die Berechnung der Maximum-Likelihood-Schätzer des multinomialen Logit-Modells und der $k - 1$ Binomialmodelle, mit Hilfe des Statistikprogrammes S-Plus, bestätigt nochmals die These, dass die Parameter Schätzungen beider Modelle nicht exakt dieselben Werte liefern.

3.7.2 Anpassungstests und Residuen

In diesem Abschnitt wird untersucht inwieweit ein Modell als adäquate Beschreibung der Daten gelten kann. Da das Logit-Modell einen linearen Zusammenhang zwischen den Logits und der Einflussgröße \mathbf{x} postuliert, liegt es nahe zu überprüfen, ob die Daten diesen linearen Zusammenhang wiedergeben. Speziell bei kategorialen Kovariablen können die empirischen Logits mit den geschätzten Logits graphisch verglichen werden (vgl. Anhang B.4). Für kategoriale Kovariablen existieren außerdem formale Anpassungstests, die im Folgenden beschrieben werden.

Pearson-Statistik

Die Idee der Residualanalyse besteht darin zu zeigen, ob die Beobachtungen durch das gewählte Modell gut oder schlecht erklärt werden. Bei einem gut angepassten Modell sollten die Beobachtungen nicht zu weit von dem durch das Modell vorgeschagten Wert entfernt sein. Die Analyse der Residuen, d.h. die Differenz zwischen beobachteten und erwarteten Response eignet sich um Beobachtungen zu identifizieren, die durch das Modell schlecht angepasst sind. Alle Beobachtungen tragen bekanntlich zur Anpassung bei. Daher ist es möglich, dass eine schlechte Anpassung des gesamten Modells auf nur wenige Beobachtungen bzw. Messstellen zurückzuführen ist. Zu untersuchen ist, warum diese Beobachtungen starke Abweichungen aufweisen. Gründe können Übertragungsfehler sein, beispielsweise bei fehlerhafter Eingabe der Daten.

Bei der Betrachtung der Residuen ist darauf zu achten, dass es sich bei den Größen um vektorwertige Größen handelt. Die einfachste Definition eines Residuums für die Meßstelle \mathbf{x}_i ist die Differenz zwischen der relativen Häufigkeit p_i und dem angepassten Wert π_i . Naheliegend ist es die Güte der Anpassung an der Differenz zwischen dem Vektor der relativen Häufigkeit $\mathbf{p}_i = (p_{i1}, \dots, p_{iq})^T$ und den Schätzungen $\hat{\boldsymbol{\pi}}_i = (\hat{\pi}_{i1}, \dots, \hat{\pi}_{iq})^T$ festzumachen, d.h. an $(\mathbf{p}_i - \hat{\boldsymbol{\pi}}_i)$, die noch durch die geschätzte Wahrscheinlichkeit normiert wird. Mit $\hat{\pi}_{ik} = 1 - \hat{\pi}_{i1} - \dots - \hat{\pi}_{iq}$ und $p_{ik} = 1 - p_{i1} - \dots - p_{iq}$ erhält man das *quadrierte Pearson-Residuum* zur i -ten Beobachtung:

$$\begin{aligned} \chi_p^2(\mathbf{p}_i, \hat{\boldsymbol{\pi}}_i) &= \sum_{r=1}^k n_i \frac{(\text{relative Häufigkeit} - \text{geschätzte Wahrscheinlichkeit})^2}{\text{geschätzte Wahrscheinlichkeit}} \\ &= n_i \sum_{r=1}^k \frac{(p_{ir} - \hat{\pi}_{ir})^2}{\hat{\pi}_{ir}}. \end{aligned}$$

Die Summe über alle Messpunkte \mathbf{x}_i , $i = 1, \dots, g$ ergibt die *Pearson-Anpassungsstatistik*

$$\chi_p^2 = \sum_{i=1}^g \chi_p^2(\mathbf{p}_i, \hat{\boldsymbol{\pi}}_i) = \sum_{i=1}^g n_i \sum_{r=1}^k \frac{(p_{ir} - \hat{\pi}_{ir})^2}{\hat{\pi}_{ir}} = \sum_{i=1}^g (\mathbf{p}_i - \hat{\boldsymbol{\pi}}_i)^T \boldsymbol{\Sigma}_i^{-1}(\hat{\boldsymbol{\beta}}) (\mathbf{p}_i - \hat{\boldsymbol{\pi}}_i).$$

Die letzte Gleichung soll im Folgenden kurz bewiesen werden.

Beweis.

$$n_i \sum_{r=1}^k \frac{(p_{ir} - \hat{\pi}_{ir})^2}{\hat{\pi}_{ir}} = n_i \left(\sum_{r=1}^q \frac{(p_{ir} - \hat{\pi}_{ir})^2}{\hat{\pi}_{ir}} + \frac{(p_{ik} - \hat{\pi}_{ik})^2}{\hat{\pi}_{ik}} \right) \quad (3.36)$$

$$\begin{aligned} (p_{ik} - \hat{\pi}_{ik})^2 &= (1 - p_{i1} - \cdots - p_{iq} - (1 - \hat{\pi}_{i1} - \cdots - \hat{\pi}_{iq}))^2 \\ &= (p_{i1} + \cdots + p_{iq} - \hat{\pi}_{i1} - \cdots - \hat{\pi}_{iq})^2 \\ &= ((p_{i1} - \hat{\pi}_{i1}) + \cdots + (p_{iq} - \hat{\pi}_{iq}))^2 = \left(\sum_{r=1}^q (p_{ir} - \hat{\pi}_{ir}) \right)^2 \end{aligned}$$

Einsetzen dieser Gleichung in (3.36) ergibt

$$\begin{aligned} &= n_i \left(\sum_{r=1}^q \frac{(p_{ir} - \hat{\pi}_{ir})^2}{\hat{\pi}_{ir}} + \frac{1}{\hat{\pi}_{ik}} \left(\sum_{r=1}^q (p_{ir} - \hat{\pi}_{ir}) \right)^2 \right) \\ &= n_i \left(\sum_{r=1}^q \frac{(p_{ir} - \hat{\pi}_{ir})^2}{\hat{\pi}_{ir}} + \frac{1}{\hat{\pi}_{ik}} \left[\sum_{r=1}^q (p_{ir} - \hat{\pi}_{ir}) \right] [(p_{i1} - \hat{\pi}_{i1}) + \cdots + (p_{iq} - \hat{\pi}_{iq})] \right) \\ &= n_i \left(\frac{(p_{i1} - \hat{\pi}_{i1})^2}{\hat{\pi}_{i1}} + \frac{1}{1 - \sum_{r=1}^q \hat{\pi}_{ir}} \sum_{r=1}^q (p_{ir} - \hat{\pi}_{ir})(p_{i1} - \hat{\pi}_{i1}) + \cdots + \frac{(p_{iq} - \hat{\pi}_{iq})^2}{\hat{\pi}_{iq}} \right. \\ &\quad \left. + \frac{1}{1 - \sum_{r=1}^q \hat{\pi}_{ir}} \sum_{r=1}^q (p_{ir} - \hat{\pi}_{ir})(p_{iq} - \hat{\pi}_{iq}) \right) \\ &= n_i \begin{pmatrix} \frac{(p_{i1} - \hat{\pi}_{i1})}{\hat{\pi}_{i1}} + \frac{(p_{i1} - \hat{\pi}_{i1})}{1 - \sum_{r=1}^q \hat{\pi}_{ir}} + \cdots + \frac{(p_{iq} - \hat{\pi}_{iq})}{1 - \sum_{r=1}^q \hat{\pi}_{ir}} \\ \vdots \\ \frac{(p_{i1} - \hat{\pi}_{i1})}{1 - \sum_{r=1}^q \hat{\pi}_{ir}} + \cdots + \frac{(p_{iq} - \hat{\pi}_{iq})}{1 - \sum_{r=1}^q \hat{\pi}_{ir}} + \frac{(p_{iq} - \hat{\pi}_{iq})}{\hat{\pi}_{iq}} \end{pmatrix}^T \cdot \begin{pmatrix} p_{i1} - \hat{\pi}_{i1} \\ \vdots \\ p_{iq} - \hat{\pi}_{iq} \end{pmatrix} \\ &= n_i (p_{i1} - \hat{\pi}_{i1}, \dots, p_{iq} - \hat{\pi}_{iq}) \\ &\quad \cdot \begin{pmatrix} \frac{1}{\hat{\pi}_{i1}} + \frac{1}{1 - \sum_{r=1}^q \hat{\pi}_{ir}} & \frac{1}{1 - \sum_{r=1}^q \hat{\pi}_{ir}} & \cdots & \frac{1}{1 - \sum_{r=1}^q \hat{\pi}_{ir}} \\ \frac{1}{1 - \sum_{r=1}^q \hat{\pi}_{ir}} & \frac{1}{\hat{\pi}_{i2}} + \frac{1}{1 - \sum_{r=1}^q \hat{\pi}_{ir}} & & \vdots \\ \vdots & & \ddots & \\ \frac{1}{1 - \sum_{r=1}^q \hat{\pi}_{ir}} & \cdots & & \frac{1}{\hat{\pi}_{iq}} + \frac{1}{1 - \sum_{r=1}^q \hat{\pi}_{ir}} \end{pmatrix} \cdot \begin{pmatrix} p_{i1} - \hat{\pi}_{i1} \\ \vdots \\ p_{iq} - \hat{\pi}_{iq} \end{pmatrix} \\ &\stackrel{(3.30)}{=} (p_i - \hat{\pi}_i)^T \Sigma_i^{-1}(\hat{\beta})(p_i - \hat{\pi}_i). \end{aligned}$$

□

Devianz oder Likelihood-Quotienten-Statistik

Eine weitere Anpassungsstatistik ist die Devianz- oder Likelihood-Quotientenstatistik. Diese lässt sich aus dem Likelihood-Quotienten

$$LQ = \frac{\sup_{\beta} L(\beta)}{\sup_{\pi_1, \dots, \pi_g} L(\pi_1, \dots, \pi_g)}$$

ableiten. Der Nenner enthält die supremale Likelihoodfunktion, wenn nur die Multinomialverteilung, kein darüber hinausgehendes Modell, als Annahme eingeht. Der Zähler enthält die supremale Likelihoodfunktion bei Gültigkeit des Modells. Als Teststatistik betrachtet man die transformierte Form $D = -2 \log(LQ)$. Daraus ergibt sich die Devianz

$$D = -2 \log(LQ) = -2 \sum_{i=1}^g (l_i(\hat{\pi}_i) - l_i(\mathbf{p}_i)), \quad (3.37)$$

wobei $l_i(\hat{\pi}_i)$, $l_i(\mathbf{p}_i)$ jeweils die Likelihoodbeiträge der einzelnen Beobachtungen bezeichnen. $l_i(\hat{\pi}_i)$ entspricht dem Beitrag nach Maximierung unter der Modellannahme und $l_i(\mathbf{p}_i)$ dem Beitrag nach Maximierung ohne Modellannahme nur unter Annahme der Multinomialverteilung. Alternativ lässt sich (3.37) durch

$$D = 2n_i \sum_{r=1}^k p_{ir} \log \left(\frac{p_{ir}}{\hat{\pi}_{ir}} \right)$$

darstellen.

Beweis. Maximierung von $L(\beta)$ mit $\hat{\pi}_{ik} = 1 - \hat{\pi}_{i1} - \dots - \hat{\pi}_{iq}$ und $p_{ik} = 1 - p_{i1} - \dots - p_{iq}$ liefert

$$\begin{aligned} l_i(\hat{\pi}_i) &\stackrel{(3.20)}{=} n_i \left\{ \sum_{r=1}^q p_{ir} \log \left(\frac{\hat{\pi}_{ir}}{\hat{\pi}_{ik}} \right) + \log(\hat{\pi}_{ik}) \right\} + \log(c_i) \\ &= n_i \left\{ \sum_{r=1}^q p_{ir} \log(\hat{\pi}_{ir}) + p_{ik} \log(\hat{\pi}_{ik}) \right\} + \log(c_i). \end{aligned}$$

Maximierung von $L(\pi_1, \dots, \pi_g)$ führt zu

$$l_i(\mathbf{p}_i) = n_i \left\{ \sum_{r=1}^q p_{ir} \log(p_{ir}) + p_{ik} \log(p_{ik}) \right\} + \log(c_i). \quad (3.38)$$

Der Einfachheit halber wird $l(\pi_1, \dots, \pi_g) = \log(L(\pi_1, \dots, \pi_g))$ maximiert.

$$\begin{aligned} l(\pi_1, \dots, \pi_g) &= \sum_{i=1}^g \left\{ \sum_{r=1}^q n_i \left[p_{ir} \log \left(\frac{\pi_{ir}}{\pi_{ik}} \right) + \log(\pi_{ik}) \right] + \log(c_i) \right\} \\ &= \sum_{i=1}^g \left\{ \sum_{r=1}^q n_i p_{ir} \log(\pi_{ir}) + \left(n_i - \sum_{r=1}^q n_i p_{ir} \right) \log(\pi_{ik}) + \log(c_i) \right\} \\ &= \sum_{i=1}^g \left\{ \sum_{r=1}^q n_i p_{ir} \log(\pi_{ir}) + \left(n_i - \sum_{r=1}^q n_i p_{ir} \right) \log \left(1 - \sum_{r=1}^q \pi_{ir} \right) + \log(c_i) \right\} \end{aligned}$$

$$\frac{\partial l(\boldsymbol{\pi})}{\partial \pi_{ir}} = \sum_{i=1}^g \left\{ \frac{n_i p_{ir}}{\pi_{ir}} - \frac{n_i - \sum_{r=1}^q n_i p_{ir}}{1 - \sum_{r=1}^q \pi_{ir}} \right\} = 0$$

Die Ableitung von l nach π_{ir} ist genau dann Null, wenn

$$\frac{n_i p_{ir}}{\pi_{ir}} - \frac{n_i - \sum_{r=1}^q n_i p_{ir}}{1 - \sum_{r=1}^q \pi_{ir}} = 0 \iff \frac{p_{ir}}{\pi_{ir}} = \frac{1 - \sum_{r=1}^q p_{ir}}{1 - \sum_{r=1}^q \pi_{ir}} \iff p_{ir} = \pi_{ir}.$$

Für die zweite Ableitung gilt

$$\frac{\partial^2 l(\boldsymbol{\pi})}{\partial \pi_{ir} \partial \pi_{is}} = - \sum_{i=1}^g n_i \frac{1 - \sum_{r=1}^q p_{ir}}{\left(1 - \sum_{r=1}^q \pi_{ir}\right)^2}.$$

Ersetzt man nun p_{ir} durch π_{ir} so folgt $\frac{\partial^2 l(\boldsymbol{\pi})}{\partial \pi_{ir} \partial \pi_{is}} < 0$, d.h. falls $p_{ir} = \pi_{ir}$ ist $l(\boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_g)$ und folglich $L(\boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_g)$ maximal.

Mit (3.38) ergibt sich folglich

$$\begin{aligned} D &= 2 \sum_{i=1}^g (l_i(\mathbf{p}_i) - l_i(\hat{\boldsymbol{\pi}}_i)) = 2 \sum_{i=1}^g n_i \left\{ \sum_{r=1}^q p_{ir} \log \left(\frac{p_{ir}}{\hat{\pi}_{ir}} \right) + (1 - p_{ik}) \log \left(\frac{p_{ik}}{\hat{\pi}_{ik}} \right) \right\} \\ &= 2 \sum_{i=1}^g n_i \left\{ \sum_{r=1}^k p_{ir} \log \left(\frac{p_{ir}}{\hat{\pi}_{ir}} \right) \right\}. \end{aligned}$$

□

Das entsprechende quadrierte *Devianz-Residuum* ist gegeben durch

$$\chi_D^2(\mathbf{p}_i, \hat{\boldsymbol{\pi}}_i) = 2n_i \sum_{r=1}^k p_{ir} \log \left(\frac{p_{ir}}{\hat{\pi}_{ir}} \right).$$

Die Pearson-Statistik und die Devianz sind unter der Voraussetzung der Modellannahmen asymptotisch χ^2 -verteilt mit $g(k-1) - p$ Parametern, wobei $g(k-1)$ für die Anzahl an Beobachtungen steht und p für die Anzahl der zu schätzenden Parameter. Die Modellanpassung ist ungenügend genau dann wenn

$$D > \chi_{g(k-1)-p, 1-\alpha}^2,$$

d.h. bei Überschreiten des $(1-\alpha)$ -Quantils der $\chi_{g(k-1)-p}^2$ Verteilung.

Bemerkung 3.7.3. Da für metrische Einflussgrößen der lokale Stichprobenumfang häufig $n_i = 1$ beträgt, sind die Pearson-Statistik und Devianz nicht zur Beurteilung der Anpassungsgüte über Teststatistiken geeignet. Eine Alternative, eine Teststatistik zu gewinnen, besteht darin, die metrische Einflussgröße in Intervalle einzuteilen, die Beobachtungen in den Intervallen zusammenzufassen und als gruppierte Beobachtungen zu betrachten (vgl. Hosmer und Lemeshow (1989) [19]). In Tabelle 1 beispielsweise wurde die metrische Variable Alter in fünf Gruppen aufgeteilt. Die Teststatistiken sind daher nur für gruppierte Daten geeignet.

Devianzanalyse

Nach der Anpassung des Modells ist zu untersuchen, inwiefern einzelne Variablen oder Interaktionseffekte vernachlässigt werden können, ohne die Gültigkeit des Modells einzuschränken. Die Idee der Devianzanalyse besteht darin zu untersuchen inwieweit sich das Modell ändert, wenn bestimmte Modellkomponenten als irrelevant betrachtet werden. Dazu wird das Modell mit den in Frage gestellten Komponenten (M_v), mit dem reduzierten Modell ohne diese Komponenten (M_r) verglichen.

Die Devianzanalyse beruht auf konditionalen Teststatistiken. Ausgangspunkt sind die beiden Modelle M_v und M_r , wobei M_r Untermodell von M_v ist. Getestet wird die Relevanz von Termen, die in M_v , nicht aber in M_r enthalten sind. Ausgehend von diesen zwei Modellen kann der Test durch eine Transformation des Likelihood-Quotienten

$$LQ(M_r | M_v) = \frac{\sup_{M_r} L(\boldsymbol{\beta})}{\sup_{M_v} L(\boldsymbol{\beta})}$$

gewonnen werden. Der Maximierung im Zähler wird das speziellere Modell M_r (in dem einige Parameter Null sind), der im Nenner das allgemeinere Modell M_v ($\boldsymbol{\pi}_i = h(\mathbf{Z}^{(i)}\boldsymbol{\beta})$) zu Grunde gelegt, d.h. M_v muss größer als M_r sein. Durch Erweiterung der Likelihoodfunktion mit dem Supremum ergibt sich die Erweiterung als Anpassungsstatistik. Dabei wird die Multinomialverteilung zu Grunde gelegt. Mit $D(M_r)$ bzw. $D(M_v)$ wird im Folgenden die Devianz des Modells M_r bzw. M_v zu identisch gruppierten Daten bezeichnet. Als Teststatistik erhält man folglich

$$\begin{aligned} D(M_r | M_v) &= -2 \log(LQ(M_r | M_v)) = -2 \log \left(\frac{\sup_{M_r} L(\boldsymbol{\beta})/L(\boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_g)}{\sup_{M_v} L(\boldsymbol{\beta})/L(\boldsymbol{\pi}_1, \dots, \boldsymbol{\pi}_g)} \right) \\ &= D(M_r) - D(M_v). \end{aligned}$$

Die Teststatistik $D(M_r | M_v)$ ist asymptotisch χ^2 -verteilt. Die dazugehörigen Freiheitsgrade ergeben sich aus der Differenz der Freiheitsgrade von $D(M_r)$ und $D(M_v)$. Beispiele finden sich in Abschnitt 3.7.6. Im Folgenden wird dieser Test auch häufig als *Likelihood-Ratio-Test* bezeichnet.

3.7.3 Akaike-Kriterium

Der Likelihood lässt sich für Modelle mit gleicher Anzahl von Parametern als vergleichendes Maß der Anpassung verwenden. Die Log-Likelihoodfunktion besitzt mit den geschätzten Wahrscheinlichkeiten $\hat{\pi}_i = h(\mathbf{Z}^{(i)}\hat{\boldsymbol{\beta}})$ die Form (3.20). Ein größerer Likelihood spricht für eine bessere Anpassung. Zu beachten ist, dass der $\log(L)$ negativ ist.

Um zwei Modelle miteinander vergleichen zu können, müssen zwei Fälle betrachtet werden. Der erste Fall ist gegeben, wenn eines der Modelle ein Untermodell des anderen ist, beispielsweise wenn zwei Logit-Modelle, einmal mit einer bestimmten Einflussgröße und einmal ohne diese, miteinander verglichen werden sollen. In diesem Fall lässt sich die Differenz der beiden Devianzen als Teststatistik benutzen (vgl. Devianzanalyse, Abschnitt 3.7.2). Besitzen die Modelle nicht die gleiche Anzahl von gefitteten Parametern und sind nicht vom selben Typ, d.h. sind nicht hierarchisch verschachtelt müssen andere Vergleichsmaße gewählt werden. Ein Maß, das einen expliziten Strafterm für die Komplexität des Modells einbezieht ist das *Akaike-Kriterium* (Akaike, 1974 [1]). Der Faktor mit dem die Anzahl der gefitteten Parameter multipliziert wird ist 2, d.h.

$$AIC = -2\log(L) + 2 \cdot (\text{Anzahl der gefitteten Parameter}).$$

Die Log-Likelihoodfunktion kann mit zunehmender Parameterzahl tendenziell auch größere Werte annehmen. Der Log-Likelihood wird um die Anzahl der Parameter reduziert, d.h. je komplexer das Modell ist, desto mehr wird der Log-Likelihood bestraft. Das Akaike-Kriterium sollte möglichst klein sein.

Beispiel 3.7.2. *Bei der Analyse der Verkehrsmittelwahl stehen als Variablen Alter, Geschlecht, persönliches Einkommen, Haushaltstyp, Kinder im Haushalt, Anzahl der Personen im Haushalt, Hauptnutzer, Mitnutzer, Netzkarte, Fahrgemeinschaft, Einstellung zum Individualverkehr (IV), Einstellung zum öffentlichen Verkehr (ÖV), Anbindung an den schienengebundenen ÖV, Zufriedenheit mit ÖV, Dauer, Tageszeit, Anzahl der Pkw im Haushalt, zur Zeit in Ausbildung, Temperatur, Entfernung, Wege mit mehreren Personen oder alleine und viele weitere zur Verfügung. Angenommen bei der Variable Alter ist unklar, ob sie besser als metrische oder in kategorisierter Form in die Analyse eingehen soll. Da Modelle mit metrischen bzw. kategorialen Alter nicht hierarchisch verschachtelt sind, lässt sich die Devianzanalyse nicht als Entscheidungskriterium heranziehen. Das Akaike-Kriterium kann hier angewandt werden. Die folgende Tabelle gibt neben dem Modell mit metrischen und kategorisierten Alter zusätzlich noch das Modell mit polynomialen und logarithmierten Alter wieder. Das AIC-Kriterium verweist hier auf das Modell mit metrischen Alter. Analog können die anderen metrischen Variablen Temperatur, Entfernung, usw. untersucht werden.*

Modell	AIC
Modell mit Alter	2191.675
Modell mit Alter ²	2197.992
Modell mit Alter ³	2201.168
Modell mit Alter ⁴	2337.638
Modell mit log(Alter)	2193.959
Modell mit Alter (kategorial)	2205.152

Tabelle 8: AIC für Alter

3.7.4 Hypothesentest und Wald-Statistik

Nach der Schätzung des Parametervektors $\hat{\beta}$ stellt sich die Frage, wie genau diese Schätzung ist. Ein numerisch großes $\hat{\beta}$ ist nur dann sinnvoll, wenn die Schätzung genügend zuverlässig ist. Ein erster Schritt ist die Betrachtung der Kovarianzmatrix $F(\hat{\beta})^{-1}$, deren Diagonalelemente die geschätzten asymptotischen Standardabweichungen von $\hat{\beta}$ enthalten. In einem zweiten Schritt kann nun die Signifikanz der Komponenten durch Tests überprüft werden. Solche Typen von Hypothesen können durch Anwendung der allgemeinen linearen Hypothese

$$H_0 : \mathbf{C}\beta = \zeta \quad H_1 : \mathbf{C}\beta \neq \zeta$$

getestet werden, wobei \mathbf{C} eine fest vorgegebene Matrix und ζ ein fest vorgegebener Vektor ist.

Der *Likelihood-Quotient*

$$LQ = \frac{\sup_{\beta : \mathbf{C}\beta = \zeta} L(\beta)}{\sup_{\beta} L(\beta)},$$

liefert eine mögliche Teststatistik, den *Likelihood-Quotienten-Test* (LQ-Test). Dabei wird im Zähler die supremale Likelihood $L(\beta)$ unter der Restriktion $\mathbf{C}\beta = \zeta$ gebildet, im Nenner die supremale Likelihood ohne diese Einschränkung. Es gilt $LQ \leq 1$, da das Supremum ohne diese Restriktion immer mindestens so groß ist wie mit dieser Restriktion. Als Teststatistik verwendet man

$$\lambda = -2\log(LQ) = -2\{l(\tilde{\beta}) - l(\hat{\beta})\},$$

wobei $\tilde{\beta}$ den Schätzer unter der Einschränkung $\mathbf{C}\beta = \zeta$ und $\hat{\beta}$ den Schätzer ohne diese Einschränkung bezeichnet. Folglich ist $l(\tilde{\beta})$ die supremale Log-Likelihood unter dieser Restriktion und $l(\hat{\beta})$ die supremale Log-Likelihood ohne diese Restriktion.

Falls die Nullhypothese nicht gültig ist, stellt die Nebenbedingung $\mathbf{C}\boldsymbol{\beta} = \boldsymbol{\zeta}$ eine nicht zu vernachlässigende Einschränkung bei der Maximierung der Likelihood dar. Der Likelihood-Quotient wird relativ klein. Ist die Nullhypothese gültig, sollte sich die Nebenbedingung bei der Maximierung wenig auswirken. Tritt die Nullhypothese ein, erhält man asymptotisch eine χ^2 -Verteilung, deren Freiheitsgrade durch den Rang der Matrix \mathbf{C} bestimmt sind. Ein wichtiger Spezialfall ist die einfache Hypothese

$$H_0 : \beta_i = 0 \quad \text{gegen} \quad H_1 : \beta_i \neq 0,$$

für $i \in A$, wobei A eine beliebige Indexmenge ist. Es genügt einmal das gesamte Modell und einmal das reduzierte Modell unter Weglassen der durch H_0 spezifizierten Variablen zu schätzen. Die Log-Likelihood des Gesamtmodells liefert $l(\hat{\boldsymbol{\beta}})$, die Log-Likelihood des reduzierten Modells liefert $l(\tilde{\boldsymbol{\beta}})$. Es ist darauf zu achten, dass für das reduzierte Modell nicht erneut gruppiert wird, d.h. das Design der Einflussgrößen zwischen dem vollen und dem reduzierten Modell unterscheidet sich nur durch Streichung von Spalten. Ist $l(\hat{\boldsymbol{\beta}})$ signifikant größer als $l(\tilde{\boldsymbol{\beta}})$ wird λ groß, d.h. H_0 wird zugunsten H_1 verworfen.

Ein alternatives Testverfahren ist der *Wald-Test* (vgl. Abschnitt 2.2.2 Fahrmeier und Tutz [15])

$$\omega = (\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\zeta})^T [\mathbf{C}\mathbf{F}^{-1}(\hat{\boldsymbol{\beta}})\mathbf{C}^T]^{-1} (\mathbf{C}\hat{\boldsymbol{\beta}} - \boldsymbol{\zeta}),$$

der auf der gewichteten Differenz zwischen der uneingeschränkten Schätzung $\mathbf{C}\hat{\boldsymbol{\beta}}$ und dem unter H_0 (asymptotisch) zu erwartenden Wert $\boldsymbol{\zeta}$ beruht. Die Differenz wird mit dem Inversen der asymptotischen Kovarianzmatrix $\mathbf{C}\mathbf{F}^{-1}(\hat{\boldsymbol{\beta}})\mathbf{C}^T$ von $\mathbf{C}\hat{\boldsymbol{\beta}}$ gewichtet. Ein großer Vorteil der Wald-Statistik liegt darin, dass sie unmittelbar aus dem Maximum-Likelihood-Schätzer $\hat{\boldsymbol{\beta}}$ berechnet werden kann, d.h. eine Neuschätzung unter Einschränkungen wie beim LQ-Test ist nicht notwendig.

Bei Eintreten der Nullhypothese erhält man asymptotisch wie für λ eine χ^2 -Verteilung, wobei die Freiheitsgrade durch den Rang der Matrix \mathbf{C} ($\text{rg}(\mathbf{C})$) bestimmt sind, d.h.

$$\lambda, \omega \sim \chi_{\text{rg}(\mathbf{C})}^2.$$

Wenn das $(1 - \alpha)$ -Quantil der $\chi_{\text{rg}(\mathbf{C}), 1-\alpha}^2$ -Verteilung überschritten ist, wird die Nullhypothese verworfen.

Beispiel 3.7.3. *In Beispiel 3.3.2 wurden bereits die Einflussgrößen Alter und Führerscheinbesitz betrachtet. In Tabelle 5 ist der Output wiedergeben. Neben den Maximum-Likelihood Parameterschätzern sind die Standardfehler wiedergegeben. In den nächsten Spalten finden sich die Ergebnisse des Wald-Tests ((Schätzwert/Standardfehler)²) jeweils*

zum Hypothesenpaar $H_0 : \beta_i = 0$, $H_1 : \beta_i \neq 0$ und die entsprechenden p -Werte. Es genügt die Wald-Statistik zu betrachten. Die Freiheitsgrade sind in diesem Beispiel für jede Kategorie gleich 1. Bei einem zu Grunde liegenden Signifikanzniveau von 5% wird die Nullhypothese verworfen, falls $\omega < \chi_{1,0.95}^2 = 3.8415$. Andernfalls wird sie abgenommen. Damit wird noch einmal deutlich, dass bei einem Vergleich von IV und Fuß die Interaktion *Alter*Führerschein* nicht zu vernachlässigen ist.

Zu überprüfen bleibt ob alle festgestellten Interaktionen tatsächlich zur Erklärung des Modells beitragen und die Anpassung verbessern.

3.7.5 Modellverifikation

Ziel ist es, die Beobachtungen der Zielvariablen durch möglichst wenige unabhängige Merkmale zu erklären. Das Modell soll die wesentlichen Datenstrukturen berücksichtigen, ohne wesentliche Details zu übersehen. Wie bei jeder statistischen Analyse geht es also um eine angemessene Informationsreduktion bei hinreichendem Detaillierungsgrad. Ein geeignetes Modell liegt zwischen dem Null- und dem vollen Modell (auch *saturiertes Modell* genannt). Wird als Erklärungsmodell der Erwartungswert der Zielvariablen als beste Prognose verwendet, so spricht man von einem Nullmodell. Es besteht aus nur einem Parameter. Viele Details der Daten bleiben unberücksichtigt. Die Modellanpassung ist daher minimal. Das volle Modell hingegen berücksichtigt jede konkrete Beobachtung. Man verwendet praktisch alle Informationen, ohne wesentlichen Strukturen aufzudecken. Dieses Modell hat genauso viele Parameter wie Beobachtungen. Es werden alle Details des Datenmaterials berücksichtigt, die Modellanpassung ist daher perfekt.

Beispiel 3.7.4. *Beispielsweise ist man daran interessiert, welches Verkehrsmittel Personen zwischen 65 und 80 Jahren benutzen. Die beiden Modelle könnten wie folgt aussehen:*

1. *Die Personen fahren mit öffentlichen Verkehrsmitteln.*
2. *Die Personen werden durch die Zahlen 1, 2, 3, ... gekennzeichnet. Person 1 fährt mit dem Rad, Person 2 geht zu Fuß, Person 3 fährt mit der U-Bahn, ...*

Die erste Beschreibung ist sehr pauschal, während die zweite so differenziert ist, dass genauso gut die gesamte Datenerhebung betrachtet werden kann. Ein optimales Modell mit hoher Erklärungskraft und wenigen Parametern wird daher zwischen diesen beiden Extremmodellen liegen.

Eine Möglichkeit die Güte der Modellanpassung zu beurteilen besteht darin die Differenzen empirischen Logits mit den geschätzten Parametern $\beta_{0r}, \beta_{1r}, \dots, \beta_{pr}$ zu vergleichen. Dazu wird die Funktion **vergleich** (Anhang B.4) verwendet. Die Vorgehensweise soll an einem einfachen Modell näher erläutert werden.

Beispiel 3.7.5. *Zu untersuchen ist die Anpassung des Modells*

$$\log \left(\frac{P(Y = r | \text{Nutzung}, \text{Geschlecht})}{P(Y = 4 | \text{Nutzung}, \text{Geschlecht})} \right) = \beta_{0r} + \beta_{G,r}x_G + \beta_{N,r}x_N, \quad r = 1, 2, 3$$

an die Daten, mit den unabhängigen Kovariablen Pkw-Nutzung (N) und Geschlecht (G).
Die Funktion liefert folgende Ergebnisse:

empirische Logits für Pkw-Nutzung

=====

Als Referenzkategorie wurde IV gewählt

Referenzkategorie und Verkehrsmittel Fuß

	Pkw-Hauptnutzer	Pkw-Mitnutzer	sonstiges
emp. Logits	-1.18	-0.38	0.82
Zellengröße n	1189.00	401.00	268.00

Referenzkategorie und Verkehrsmittel ÖV

	Pkw-Hauptnutzer	Pkw-Mitnutzer	sonstiges
emp. Logits	-1.89	-0.77	0.82
Zellengröße n	1046.00	348.00	268.00

Referenzkategorie und Verkehrsmittel Rad

	Pkw-Hauptnutzer	Pkw-Mitnutzer	sonstiges
emp. Logits	-2.39	-1.04	0.16
Zellengröße n	992.00	322.00	178.00

Differenzen der empirischen Logits:

Referenzkategorie IV und Verkehrsmittel Fuß [1] 0.8 2.0

Referenzkategorie IV und Verkehrsmittel ÖV [1] 1.12 2.71

Referenzkategorie IV und Verkehrsmittel Rad [1] 1.35 2.55

geschätzte Parameter:

	(Intercept) $\hat{\beta}_{0r}$	as.factor(NUTZUNG) Pkw-Mitnutzer $\hat{\beta}_{1r}$	as.factor(NUTZUNG) sonstiges $\hat{\beta}_{2r}$	GESCHLECHT $\hat{\beta}_{3r}$
Fuß	-1.11	0.78	1.97	-0.12
ÖV	-1.86	1.11	2.70	-0.05
Rad	-2.73	1.46	2.66	0.53

Die Differenzen der empirischen Logits entsprechen gerade den zu schätzenden Parametern, da Folgendes gilt:

$$\text{logit}(Pkw\text{-Mitnutzer}) - \text{logit}(Pkw\text{-Hauptnutzer}) = \beta_{0r} + \beta_{1r} x_{Pkw\text{-Mitnutzer}} - \beta_{0r} = \beta_{1r}$$

$$\text{logit}(sonstige) - \text{logit}(Pkw\text{-Hauptnutzer}) = \beta_{0r} + \beta_{2r} x_{sonstiges} - \beta_{0r} = \beta_{2r}$$

Daher können diese miteinander verglichen werden. Zu beachten ist, dass Pkw-Hauptnutzer als Referenzkategorie der erklärenden Variable Nutzung gewählt wurde.

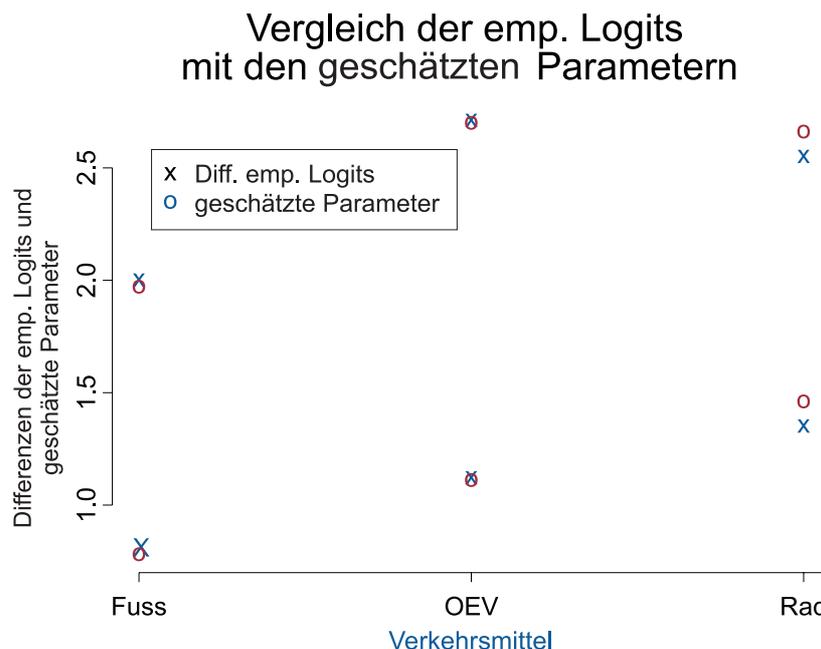


Abbildung 4: Vergleich empirischer Logits mit geschätzten Parametern $\hat{\beta}_{ir}$

Der Plot enthält die Werte der Differenzen der empirischen Logits und die geschätzten Parameter der Einflussgröße Nutzung (Pkw-Mitnutzer) bzw. Nutzung (sonstiges). Die Grafik zeigt bei diesem einfachen Modell eine gute Anpassung der Daten sehr deutlich. Die geschätzten Parametern liegen dicht an den Differenzen der empirischen Logits. Die Anpassung dieses Modells kann folglich als adäquate Beschreibung der Daten gelten.

Je mehr Variablen das Modell erklären, desto größer ist die Wahrscheinlichkeit, dass Variablen nicht für alle Responsekategorien Einfluss haben, d.h. ohne die Einführung kategorienspezifischer Variablen ist der Vergleichstest bei zunehmender Modellgröße nicht mehr geeignet.

Ein weitere Möglichkeit die Güte der Anpassung zu beurteilen kann mit der Devianz durchgeführt werden, da sie unter der Gültigkeit des Modells $\chi^2_{g(k-1)-p}$ verteilt ist. Die Devianz ist genau dann Null, wenn das saturierte Modell vorliegt. Sie ist am größten, wenn das Nullmodell betrachtet wird. Die Modellsuche kann mit dem Modell begonnen werden, dass alle Informationen des Datensatzes berücksichtigt. Die Komplexität wird schrittweise reduziert, d.h. einige Größen werden für unwichtig gehalten und folglich eliminiert. Dabei

nimmt die Devianz des reduzierten Modells gegenüber dem vollen Modell zu. Ab einem bestimmten kritischen Wert sollen keine weiteren Abweichungen zugelassen werden. Das letzte Modell wird als passendes Modell akzeptiert. Es enthält nur wenige Parameter. Die Abweichung von den Daten ist gering. Es wurden alle wesentlichen Einflussgrößen erfasst. Die Modellsuche wird folglich abgebrochen, wenn die Devianz einen bestimmten Wert ($D > \chi_{g(k-1)-p, 1-\alpha}^2$, vgl. Abschnitt 3.7.2, Devianz oder Likelihood-Quotienten-Statistik) übersteigt. Bei den Signifikanztests wird ein Niveau von 0.05 unterstellt. Entspricht die Devianz ungefähr der Anzahl an Freiheitsgraden, kann das Modell als gut angepasst akzeptiert werden. Wenn ein gut angepasstes Modell gefunden wurde, kann einerseits untersucht werden, ob die Hinzunahme weiterer Variablen oder Interaktionseffekte (erweitertes Modell) einen signifikanten besseren Modellfit ergibt, andererseits kann getestet werden, ob das Eliminieren einiger Variablen (reduziertes Modell) einen signifikant schlechteren Modellfit zur Folge hat ohne dass das Modell an Gültigkeit verliert. Dazu wird das bereits angepasste Modell mit dem erweiterten bzw. reduzierten Modell verglichen (vgl. ebenfalls Abschnitt 3.7.2, Devianzanalyse). Das Signifikanzniveau hängt von dem Ziel des jeweiligen Tests ab. Angenommen es wurde ein passendes Modell gefunden. Es wird erwartet, dass das Weglassen jedes Effektes eine signifikante Modellverschlechterung zur Folge hat. Eine Überprüfung folgt mit Hilfe der partiellen Devianzanalyse. Wenn die partielle Devianz des reduzierten Modells gegenüber dem Ausgangsmodell einen bestimmten kritischen Wert überschreitet, ist davon auszugehen, dass der jeweilige Parameter nicht eliminiert werden sollte. Umgekehrt wird erwartet, dass das Hinzufügen weiterer Effekte keine wesentliche Verbesserung ergibt. Diese Hypothese kann mit den klassischen Anpassungstest überprüft werden. Wenn die partielle Devianz des Ausgangsmodells gegenüber dem erweiterten Modell das Signifikanzniveau nicht überschreitet hat sich gezeigt, dass der in Frage stehende Parameter das Modell nicht verbessert. Im Allgemeinen ist darauf zu achten, dass die Stichprobe pro Zelle genügend groß ist (jede Zelle sollte mindestens 5 Beobachtungen haben). Nur dann greift die Asymptotik.

3.7.6 Modellwahlstrategie bei Daten mit vielen Kovariablen

Eine Untersuchung aller Submodelle ist hier unmöglich. Aus diesem Grund wird ein sequentielles Variablenselektionsverfahren verwendet. Die *Vorwärtselektion* beginnt mit dem kleinsten Modell und fügt sukzessive weitere Variablen hinzu. Bei allen statistischen Tests wird ein Signifikanzniveau von 5% zu Grunde gelegt.

Beispiel 3.7.6. *Es werden erneut die Daten der Mobilitätsstudie betrachtet. Der Verkehrsbereich „Individualverkehr“ wird als Referenzkategorie gewählt. Angenommen Modell 1 hat die Form:*

$$\log \left(\frac{P(Y = r | \mathbf{x})}{P(Y = IV | \mathbf{x})} \right) = \beta_{0r} + \beta_{\text{Alter}, r} * \text{Alter} + \beta_{\text{pers.Eink.}, r} * x_{\text{pers.Eink.}} \\ + \beta_{\text{Pkw-Nutzung}, r} * x_{\text{Pkw-Nutzung}},$$

wobei als bekannt vorausgesetzt wird, dass Alter, persönliches Einkommen und die Pkw-Nutzung Einfluss auf die Verkehrsmittelwahl haben. r steht für die Verkehrsbereiche Fuß, öffentlicher Verkehr und Rad. Im Folgenden soll untersucht werden, ob der Besitz einer Netzkarte ebenfalls einen Einfluss bei der Verkehrsmittelwahl hat. Eine rechnerische Überprüfung erfolgt mittels Likelihood-Ratio-Test (vgl. Devianzanalyse in Abschnitt 3.7.4). Die Nullhypothese $\beta_{\text{Netzkarte}, r} = 0$ wird genau dann verworfen, wenn die Devianz von Modell 2 (entspricht Modell 1 einschließlich der Netzkarte) abzüglich der Devianz von Modell 1 größer als $\chi_{3, 0.95}^2 = 7.815$ ist. Der Likelihood-Ratio-Test für multinomiale Modelle ergibt

Likelihood ratio tests of Multinomial Models

Response: REF.KAT.IV

					Model
1		ALTER + as.factor(PERS.EINK) + as.factor(NUTZUNG)			
2		NETZKARTE + ALTER + as.factor(PERS.EINK) + as.factor(NUTZUNG)			
	Resid. df	Resid. Dev	Test	Df	LR stat. Pr(Chi)
1	558	5351.439			
2	555	5209.936	1 vs 2	3	141.503 0

Die Nullhypothese wird demzufolge verworfen, d.h. die Verfügung über eine Netzkarte hat einen signifikanten Einfluss bei der Verkehrsmittelwahl. Gruppieren der Daten liefert 198 nicht leere Zeilen, d.h. in jeder Zeile existiert mindestens eine Beobachtung. Die Parameter der *Residual Deviance* von Modell 1 ergeben sich aus $g(k - 1)$ abzüglich der zu schätzenden Parameter p , d.h. $g(k - 1) - p = (196 \text{ (nicht leere Zeilen)}) * 3 - (1 \text{ (Freiheitsgrad für Alter)} + 6 \text{ (Freiheitsgrade für persönliches Einkommen)} + 2 \text{ (Freiheitsgrade für Pkw-Nutzung)} + 1 \text{ (Intercept)}) * 3 = 558$. Modell 2 hat drei weitere zu schätzende Parameter, d.h. die *Resid. df* sind hier gleich 555.

Diese beiden Modelle mit nur drei bzw. vier erklärenden Variablen haben bereits 558 bzw. 555 Freiheitsgrade. Die Ursache liegt darin, dass die metrische Variable Alter aus insgesamt 60 verschiedenen Alterswerten besteht. Aus diesem Grund ist es häufig sinnvoll metrischen Variablen zu *scoren*, d.h. die metrischen Größen werden zunächst in Gruppen eingeteilt und diesen schließlich der Mittelwert des Gruppenintervalls zugeteilt. Für die Altersgruppe der 18+ bis 31-jährigen ergibt sich somit ein gescortes Alter von 25 Jahren. Außerdem wird die gruppierte Größe persönliches Einkommen neu generiert. Sie

besteht nun nur noch aus den Kategorien niedriges persönliches Einkommen, mittleres persönliches Einkommen und hohes persönliches Einkommen. Betrachtet man erneut den Likelihood-Ratio-Test so ergibt sich:

Likelihood ratio tests of Multinomial Models

Response: REF.KAT.IV

					Model
1	S.ALTER + as.factor(PERS.EINK.NEU) + as.factor(NUTZUNG)				
2	NETZKARTE + S.ALTER + as.factor(PERS.EINK.NEU) + as.factor(NUTZUNG)				
	Resid. df	Resid. Dev	Test	Df	LR stat. Pr(Chi)
1	99	5390.443			
2	96	5253.064	1 vs 2	3	137.38 0

Die Resid. df sind deutlich geringer.

3.7.7 Effektivitätsmaße vom Typ R^2

Auch wenn ein Modell als adäquate Beschreibung der Daten gilt und alle Variablen einen signifikanten Einfluss besitzen, ist noch nicht geklärt, inwieweit sich die Variation der abhängigen Variable durch die Einflussgrößen erklären lässt. Das vordergründige Interesse besteht an der Effektivität des gewählten Modells. Da es zur Überprüfung der Anpassung kein echtes Analogon zum R^2 des linearen Regressionsmodells gibt, werden häufig sogenannte *Pseudo R^2* verwendet.

Ein weit verbreitetes Maß ist das sogenannte *Pseudo R^2 nach McFadden* (1974) [23]

$$R_{\text{MF}}^2 = 1 - \frac{\log(L(\hat{\beta}))}{\log(L_0)},$$

wobei $\log(L(\hat{\beta}))$ die Log-Likelihoodfunktion des vollen Modells (Modell mit allen erklärenden Variablen) und $\log(L_0)$ die Log-Likelihoodfunktion des Nullmodells (Modell nur mit einer Konstanten) bezeichnet.

Eigenschaften von R_{MF}^2

- (1) Ebenso wie das klassische R^2 des linearen Regressionsmodells liegt dieses Maß zwischen 0 und 1.
- (2) $R_{\text{MF}}^2 = 0$, falls $\log(L(\hat{\beta})) = \log(L_0)$, d.h. wenn alle nach dem ML-Prinzip geschätzten Koeffizienten (außer den Konstanten) verschwinden.
- (3) $R_{\text{MF}}^2 = 1$, falls $L(\hat{\beta}) = 1$ ist.

Urban (1993) [28] spricht bereits bei Werten von 0.2 bis 0.4 von einer guten Anpassung. Häufig benutzt man dieses Maß auch um Aussagen über den Informationsgehalt, die in den erklärenden Variablen steckt, zu treffen. Powers und Xie [25] weisen darauf hin, dieses Maß mit Vorsicht zu benutzen.

Durch die Hinzunahme weiterer Einflussgrößen kann der R_{MF}^2 nicht fallen. Soll der Zuwachs an Freiheitsgraden berücksichtigt werden, ist es sinnvoll das korrigierte Pseudo R^2

$$R_{\text{MF, korrr}}^2 = 1 - \frac{\log(L(\hat{\beta})) - A}{\log(L_0)},$$

zu betrachten, wobei A die Anzahl der Parameter im vollen Modell ist. Dieses Maß wird jedoch nur dann wachsen, wenn der Beitrag der Variablen den Verlust an Freiheitsgraden ausgleicht.

3.8 Alternativen und Vorteile des Logit-Modells

Neben dem hier vorgestellten Logit-Modell gibt es zahlreiche alternative Responsefunktionen. Die Festlegung der Funktion F bestimmt den Typ des Modells. Wählt man zum Beispiel statt der logistischen Verteilung die Verteilungsfunktion der Standardnormalverteilung, erhält man das Probit-Modell. Aus der Gleichverteilung über dem Intervall $(0, 1)$ erhält man das lineare Modell. Weitere Modelle sind die Extremwertverteilungsmodelle, das komplementäre Log-Modell oder das Exponentialmodell. Eine ausführliche Darstellung dieser Modelle findet sich in Tutz (2000) [27], Kapitel 4. Das Logit-Modell zeichnet sich jedoch durch einige Besonderheiten gegenüber diesen Alternativen aus, die das Modell als besonders geeignet erscheinen lässt.

1. Parameterschätzungen sind leichter interpretierbar bezüglich der Chancen bzw. logarithmierten Chancen.
2. Die Elimination von Parametern lässt sich im Fall kategorialer Einflussgrößen als bedingte stochastische Unabhängigkeit zwischen Responsegröße und Einflussgröße interpretieren (siehe Abschnitt 7.4.4 in Tutz (2000) [27]).
3. Man erhält einfachere Schätzgleichungen zur Maximum-Likelihood-Schätzung.
4. Unter schwachen Bedingungen ist die Existenz des Maximum-Likelihood-Schätzers gesichert (vgl. Fahrmeir & Kaufmann (1985) [12]).
5. Angenommen \mathbf{x} ist multivariat normalverteilt in zwei Populationen mit unterschiedlichem Erwartungswert aber identischer Kovarianzmatrix. Dann erhält man explizit

das logistische Modell

$$P(Y = 1 | \mathbf{x}) = \frac{\exp(\beta_0 + \mathbf{x}^T \boldsymbol{\beta})}{1 + \exp(\beta_0 + \mathbf{x}^T \boldsymbol{\beta})}$$

mit

$$\begin{aligned} \boldsymbol{\beta} &= \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \\ \beta_0 &= -\frac{1}{2}\boldsymbol{\mu}_1^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 + \frac{1}{2}\boldsymbol{\mu}_2^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2 + \log \left(\frac{P(Y = 1)}{P(Y = 2)} \right). \end{aligned}$$

Beweis. Die a posteriori Wahrscheinlichkeit lässt sich mit dem Satz von Bayes wie folgt darstellen

$$P(Y = 1 | \mathbf{x}) = \frac{P(\mathbf{x} | 1)p(1)}{P(\mathbf{x} | 1)p(1) + P(\mathbf{x} | 2)p(2)} = \frac{1}{1 + \frac{P(\mathbf{x} | 2)p(2)}{P(\mathbf{x} | 1)p(1)}}, \quad (3.39)$$

wobei $P(Y = 1) := p(1)$ und $P(Y = 2) := p(2)$ definiert wurden. Einsetzen der Dichten der Normalverteilung

$$P(\mathbf{x} | r) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \exp \left\{ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_r)^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_r) \right\}$$

in (3.39) ergibt

$$\begin{aligned} & \frac{1}{1 + \frac{\exp\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_2)\}p(2)}{\exp\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_1)\}p(1)}} \\ &= \frac{1}{1 + \exp \left\{ -\frac{1}{2} [2\mathbf{x}^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) + \boldsymbol{\mu}_2^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2 - \boldsymbol{\mu}_1^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1] - \log \left(\frac{p(1)}{p(2)} \right) \right\}} \\ &= \frac{1}{1 + \exp \left\{ - \left[\frac{1}{2} \boldsymbol{\mu}_2^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2 - \boldsymbol{\mu}_1^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 + \log \left(\frac{p(1)}{p(2)} \right) + \mathbf{x}^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \right] \right\}} \\ &= \frac{1}{1 + \exp(-(\beta_0 + \mathbf{x}^T \boldsymbol{\beta}))} = \frac{\exp(\beta_0 + \mathbf{x}^T \boldsymbol{\beta})}{1 + \exp(\beta_0 + \mathbf{x}^T \boldsymbol{\beta})} \end{aligned}$$

mit

$$\begin{aligned} \boldsymbol{\beta} &= \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \\ \beta_0 &= -\frac{1}{2}\boldsymbol{\mu}_1^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 + \frac{1}{2}\boldsymbol{\mu}_2^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2 + \log \left(\frac{P(Y = 1)}{P(Y = 2)} \right). \end{aligned}$$

□

Insbesondere erkennt man, dass nur die Konstante β_0 von den a priori Wahrscheinlichkeiten $P(Y = 1)$ und $P(Y = 2)$ abhängt, nicht aber der in $\boldsymbol{\beta}$ enthaltene Einfluss der Kovariablen.

4 Zufällige Effekte

Um die vielen Daten zu Grunde liegende zufällige Struktur adäquat zu berücksichtigen, ist es notwendig zufällige Effekte in das Modell aufzunehmen. In diesem Kapitel sollen Modelle mit zufälligen Effekten vorgestellt werden.

Parameter gewöhnlicher linearer Modelle werden als feste Effekte bezeichnet. Sie beziehen sich auf alle Einflüsse (d.h. Geschlecht, Beruf, usw.). Zufällige Effekte hingegen beziehen sich auf eine Stichprobe aller möglichen Gruppen. Eine Erweiterung des linearen Modells mit festen Effekten ist das *gemischte lineare Modell*, bei dem zusätzlich zufällige Effekte berücksichtigt werden. Das Modell besitzt dann folgende Form:

$$y = \mathbf{x}^T \boldsymbol{\beta} + \mathbf{z}^T \mathbf{u} + \varepsilon,$$

wobei $\mathbf{z}^T = (z_1, \dots, z_n)$ bekannt ist und $\mathbf{u}^T = (u_1, \dots, u_n)$ die zufälligen Einflüsse beschreiben. Der Vektor \mathbf{u} ist hier also eine nicht (direkt) beobachtbare Zufallsvariable, von der angenommen wird, dass sie normalverteilt ist mit Erwartungswert $\mathbf{0}$ und Kovarianzmatrix $\sigma^2 \mathbf{A}$, wobei die bekannte Matrix \mathbf{A} häufig die Einheitsmatrix beschreibt. Eine weitere wichtige Erweiterung des linearen Modells, bildet das generalisierte lineare Modell. Dieses wurde im Abschnitt 2.4 vorgestellt. Darauf aufbauend wird in den folgenden Abschnitten auf zufällige Effekte in generalisierten linearen Modellen eingegangen. Nachfolgender Abschnitte basieren auf McCulloch und Searle (2001) [22].

4.1 Zufällige Effekte in generalisierten linearen Modellen

Bei Modellen mit zufälligen Effekten geht man davon aus, dass neben den interessierenden Merkmalen auch zumindest einige der Einflussgrößen zufällig sind. Bei qualitativen, zufälligen Einflussgrößen, wie sie in dieser Arbeit betrachtet werden, bedeutet dies, dass es sich bei den in einem Experiment betrachteten, endlich vielen Faktorstufen lediglich um eine Auswahl aus einer Anzahl von interessierenden Faktorstufen handelt. Die Aussagen über die Effekte des Faktors bezüglich der interessierenden Merkmale beziehen sich auf alle Stufen des Faktors. Eine Klasse der generalisierten linearen Mischungsmodelle (GLMM) kann durch Hinzufügen *zufälliger Effekte* zum linearen Prädiktor eines generalisierten linearen Modells (GLM) erzeugt werden.

Beispiel 4.1.1. (Einführendes Beispiel)

Untersucht werden soll die Verkehrsmittelwahl von 167 Haushalten in München. Die Wahl eines Haushaltes erfolgt zufällig aus einer Population von Haushalten in München. Falls im Haushalt i T_i Wege für diesen Versuch zur Verfügung stehen und die Verkehrsmittel-

wahl von Weg t in Haushalt i y_{it} ist, kann das Modell durch

$$\mathbb{E}(y_{it} | \mathbf{x}_{it}, \mathbf{u}_i) = \mathbf{x}_{it}^T \boldsymbol{\beta} + u_i, \quad i = 1, \dots, n, \quad t = 1, \dots, T_i \quad (4.1)$$

beschrieben werden. u_i in Gleichung (4.1) bezeichnet den zufälligen Einfluss, da Haushalt i nur zufällig aus der Gesamtpopulation der Haushalte ausgewählt wurde. Der Haushalt ist nur von Interesse als ein Haushalt der zufällig aus allen Haushalten der Stichprobe ausgewählt wurde. Damit können beispielsweise Rückschlüsse für die gesamte Stichprobe gezogen werden.

Das Modell wird auch als ein *Mischungsmodell* bezeichnet, da sich das Verhalten des Responses aus einer Mischung verschiedener Populationen ergibt, die durch den unbeobachtbaren Effekt \mathbf{u}_i charakterisiert sind. Die Komponente y_{it} beschreibt die t -te Beobachtung in Cluster i , $i = 1, \dots, n$, $t = 1, \dots, T_i$. Der Spaltenvektor \mathbf{x}_{it} beschreibt die Einflussgrößen für die t -te Beobachtung i -ten Clusters und dient im Modell als Kovariable der festen Effekte. \mathbf{z}_{it} ist der entsprechende Kovariablenvektor der zufälligen Effekte, die in einem Vektor \mathbf{u} zusammengefasst sind, dabei beschreibt \mathbf{u}_i den Vektor der zufälligen Einflüsse in Cluster i . Bedingt auf \mathbf{u} ähnelt ein GLMM einem gewöhnlichen GLM. Zur Darstellung des Modells wird die bedingte Verteilung von \mathbf{y} gegeben \mathbf{u} benötigt. Es wird angenommen, dass der Responsevektor \mathbf{y} aus bedingt unabhängigen Elementen besteht, jedes mit einer Verteilung mit Dichte der Exponentialfamilie wie sie für GLMs schon eingeführt wurde. Es wird angenommen, dass die bedingte Dichte $f(y_{it} | \mathbf{u}_i, \theta_i, \phi)$ vom Typ der uni- oder multivariaten exponentiellen Familie

$$f(y_{it} | \mathbf{u}_i, \theta_i, \phi) = \exp \left\{ \frac{y_{it} \theta_i - b(\theta_i)}{a(\phi)} + c(y_{it}, \theta_i) \right\} \quad (4.2)$$

ist mit bedingtem Erwartungswert

$$\mu_{it} = \mathbb{E}(y_{it} | \mathbf{u}_i) = h(\eta_{it}),$$

mit

$$g(\mu_{it}) = \eta_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta} + \mathbf{z}_{it}^T \mathbf{u}_i,$$

wobei h die Responsefunktion, g die Linkfunktion und η_{it} den linearen Prädiktor beschreibt. Die Verteilung des zufälligen Effektes \mathbf{u}_i wird durch multivariate Normalverteilung

$$\mathbf{u}_i \sim N(\mathbf{0}, \mathbf{D}), \quad i = 1, \dots, n$$

mit unbekannter Kovarianzmatrix \mathbf{D} spezifiziert. Die bedingte Varianz der abhängigen Variable steht über

$$\text{Var}(y_i | \mathbf{u}) = a(\phi)v(\mu_i)$$

mit dem Dispersionsparameter ϕ in Zusammenhang (siehe Gleichung (2.10)). Die Funktion $a(\phi)$ hat häufig die Form ϕ/ω_i , wobei ω_i ein bekanntes Gewicht und $v(\cdot)$ die Varianzfunktion darstellt. Zu beobachten ist eine Abhängigkeit vom bedingten Erwartungswert μ_i .

In der Betrachtung der Tatsache, dass die bedingte Verteilung von \mathbf{y} gegeben \mathbf{u} die Erweiterung eines GLMs ist, gelten viele Beziehungen die dort abgeleitet wurden auch für GLMMs.

Es gibt verschiedene Ansätze, Modelle mit zufälligen Effekten zu schätzen. Eine Schätzung nach der Maximum-Likelihood Methode ist prinzipiell möglich (Jansen, 1990 [20]). Maximum-Likelihood-Schätzung für normalverteilte zufällige Effekte erfordert jedoch häufig numerische Integrationen, wobei mit Einschränkungen auf die Anzahl und Struktur der hinzugefügten zufälligen Effekte zu rechnen ist. Dieser Ansatz bereitet insbesondere bei größeren Datensätzen rechentechnische Probleme. Aus diesem Grund sind alternative Verfahren vorgeschlagen worden, die auf einer Approximation der Likelihood beruhen. Engel und Buist (1993) [9] haben eine Methode vorgestellt, die ohne diese Einschränkung auskommt. Die Methode beruht darauf, dass die Quasi-Likelihood-Schätzung approximativ einer gemischten linearen Modellstruktur folgt. Die Parameter werden durch eine Kombination von Quasi-Likelihood und iterierten Kleinste-Quadrate-Verfahren geschätzt, wobei nicht alle Verteilungsannahmen benötigt werden. Engel und Keen (1993) [10] haben ebenfalls eine Methode vorgeschlagen, die ohne diese Einschränkungen auskommt. Die Schätzmethode ist ähnlich zur penalisierten Quasi-Likelihood Methode, die im folgenden Abschnitt betrachtet wird. Als Alternative zum Maximum-Likelihood wurde der *Gibbs Sampler* vorgeschlagen (vgl. Zeger und Karim (1991) [30]). Der Gibbs Sampler ist eine Monte Carlo Methode (basierend auf dem Bayes Ansatz) um Beobachtungen aus der gemeinsamen posteriori Verteilung zu gewinnen. Dabei sind die a priori Verteilung für die festen Effekte und Varianzkomponenten zu wählen. Jonathan Myles und David Clayton haben den Gibbs Sampler in S-Plus (library *glm-01*) implementiert. In dieser Arbeit wird nicht näher auf die Gibbs Sampler Methode eingegangen.

4.2 Penalisierte Quasi-Likelihood Methode

In diesem Abschnitt soll eine Methode zum gemeinsamen Schätzen der Parameter β und \mathbf{u} vorgestellt werden. Insbesondere die Robustheit der Quasi-Likelihoodfunktion führte zu der Suche nach allgemeinen Anwendungen für GLMMs. Zentral dazu ist die Anwendung einer Laplace Approximation zur Berechnung höher dimensionaler Integrale der Likelihoodfunktion. Grundlage ist die Methode von Breslow und Clayton (1993) [6].

Die marginale Likelihoodfunktion ist gegeben durch

$$L(\boldsymbol{\beta}) = \int f(\mathbf{y} | \mathbf{u}, \theta, \phi) f(\mathbf{u} | \mathbf{D}) d\mathbf{u}$$

und die marginale Log-Likelihoodfunktion der GLMMs durch

$$\begin{aligned} l(\boldsymbol{\beta}) &= \log \int f(\mathbf{y} | \mathbf{u}, \theta, \phi) f(\mathbf{u} | \mathbf{D}) d\mathbf{u} = \log \int \exp[\log(f(\mathbf{y} | \mathbf{u}, \theta, \phi)) + \log(f(\mathbf{u} | \mathbf{D}))] d\mathbf{u} \\ &= \log \int \exp(h(\mathbf{u})) d\mathbf{u}, \end{aligned} \quad (4.3)$$

wobei $h(\mathbf{u}) := \log(f(\mathbf{y} | \mathbf{u}, \theta, \phi)) + \log(f(\mathbf{u} | \mathbf{D}))$. Da die Integration im Allgemeinen nicht analytisch durchführbar ist, wird $l(\boldsymbol{\beta})$ approximiert. Das Grundgerüst der Laplace Approximation beruht auf der Taylorreihenentwicklung zweiter Ordnung (vgl. Gleichung (3.17) auf Seite 60 in Barndorff-Nielsen und Cox (1989) [4]) und hat die Form

$$\log \int \exp(h(\mathbf{u})) d\mathbf{u} \approx h(\mathbf{u}_0) + \frac{q}{2} \log(2\pi) - \frac{1}{2} \log \left| -\frac{\partial^2 h(\mathbf{u})}{\partial \mathbf{u} \partial \mathbf{u}^T} \Big|_{\mathbf{u}=\mathbf{u}_0} \right|, \quad (4.4)$$

wobei \mathbf{u}_0 die Lösung von

$$\frac{\partial h(\mathbf{u})}{\partial \mathbf{u}} \Big|_{\mathbf{u}=\mathbf{u}_0} = \mathbf{0} \quad (4.5)$$

ist. Falls $\mathbf{u} \sim N(\mathbf{0}, \mathbf{D})$, $\mathbf{u} \in \mathbb{R}^n$ ergibt sich aus der Dichte der multivariaten Normalverteilung

$$f(\mathbf{u} | \mathbf{D}) = \frac{\exp(-\frac{1}{2} \mathbf{u}^T \mathbf{D}^{-1} \mathbf{u})}{(2\pi)^{q/2} |\mathbf{D}|^{1/2}},$$

dass

$$\log f(\mathbf{u} | \mathbf{D}) = -\frac{1}{2} \mathbf{u}^T \mathbf{D}^{-1} \mathbf{u} - \frac{q}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{D}| \quad (4.6)$$

und daraus folgt

$$\begin{aligned} h(\mathbf{u}) &= \log(f(\mathbf{y} | \mathbf{u}, \theta, \phi)) + \log(f(\mathbf{u} | \mathbf{D})) \\ &= \log(f(\mathbf{y} | \mathbf{u}, \theta, \phi)) - \frac{1}{2} \mathbf{u}^T \mathbf{D}^{-1} \mathbf{u} - \frac{q}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{D}|. \end{aligned} \quad (4.7)$$

Differenzieren dieser Gleichung nach \mathbf{u} ergibt

$$\frac{\partial h(\mathbf{u})}{\partial \mathbf{u}} = \frac{\partial}{\partial \mathbf{u}} \log(f(\mathbf{y} | \mathbf{u}, \theta, \phi)) - \mathbf{D}^{-1} \mathbf{u}.$$

Ersetzt man in (2.11) $\boldsymbol{\beta}$ durch \mathbf{u} und \mathbf{X} durch \mathbf{Z} ergibt sich

$$\frac{\partial h(\mathbf{u})}{\partial \mathbf{u}} = \frac{1}{a(\phi)} \mathbf{Z}^T \mathbf{W}(\boldsymbol{\beta}, \mathbf{u}) \boldsymbol{\Delta}(\boldsymbol{\beta}, \mathbf{u}) (\mathbf{y} - \boldsymbol{\mu}) - \mathbf{D}^{-1} \mathbf{u} = \mathbf{0}, \quad (4.8)$$

wobei \mathbf{W} und Δ wie in Abschnitt 2.4 definiert sind. Nochmaliges Differenzieren nach \mathbf{u} mit Hilfe der Kettenregel liefert

$$\begin{aligned} \frac{\partial^2 h(\mathbf{u})}{\partial \mathbf{u} \partial \mathbf{u}^T} &= \underbrace{\frac{1}{a(\phi)} \mathbf{Z}^T \frac{\partial \mathbf{W}(\boldsymbol{\beta}, \mathbf{u}) \Delta(\boldsymbol{\beta}, \mathbf{u})}{\partial \mathbf{u}^T} (\mathbf{y} - \boldsymbol{\mu})}_{\text{Restterm}} \\ &\quad - \frac{1}{a(\phi)} \mathbf{Z}^T \mathbf{W}(\boldsymbol{\beta}, \mathbf{u}) \Delta(\boldsymbol{\beta}, \mathbf{u}) \frac{\partial \boldsymbol{\mu}}{\partial \mathbf{u}^T} - \mathbf{D}^{-1}, \end{aligned} \quad (4.9)$$

wobei $\partial \boldsymbol{\mu} / \partial \mathbf{u}^T = (\partial \boldsymbol{\mu} / \partial \boldsymbol{\eta}) \cdot (\partial \boldsymbol{\eta} / \partial \mathbf{u}) = \Delta^{-1}(\boldsymbol{\beta}, \mathbf{u}) \mathbf{Z}$.

Für Binomial- oder Poissonmodelle ist

$$\mathbf{W}(\boldsymbol{\beta}, \mathbf{u}) \Delta(\boldsymbol{\beta}, \mathbf{u}) = \mathbf{I}_n.$$

Diese letzte Gleichung soll im Folgenden kurz bewiesen werden.

Beweis. Die Diagonaleinträge der Diagonalmatrix $\mathbf{W}(\boldsymbol{\beta}, \mathbf{u})$ lauten $\omega_i = \{v(\mu_i)[g'(\mu_i)]^2\}^{-1}$. Die Diagonalmatrix $\Delta(\boldsymbol{\beta}, \mathbf{u})$ hat die Einträge $\delta_i = \{g'(\mu_i)\}$. Da außerdem für Binomial- und Poissonmodelle $a(\phi) = 1$ ist, gilt

$$\text{Var}(y) = v(\mu_i) a(\phi) = b''(\theta_i) a(\phi) = \frac{\partial b'(\theta_i)}{\partial \theta_i} = \frac{\partial \mu_i}{\partial \theta_i} = (g'(\mu_i))^{-1}$$

und folglich $v(\mu_i) = \text{Var}(y)/a(\phi)$. Einsetzen in $\mathbf{W}(\boldsymbol{\beta}, \mathbf{u})$ und $\Delta(\boldsymbol{\beta}, \mathbf{u})$ ergibt

$$\mathbf{W}(\boldsymbol{\beta}, \mathbf{u}) \Delta(\boldsymbol{\beta}, \mathbf{u}) = \begin{pmatrix} (g'(\mu_1))^{-1} g'(\mu_1) & & 0 \\ & \ddots & \\ 0 & & (g'(\mu_n))^{-1} g'(\mu_n) \end{pmatrix} = \mathbf{I}_n$$

□

Damit verschwindet für Binomial- und Poissonmodelle in Gleichung (4.9) der erste Term. Im Allgemeinen hat der erste Term Erwartungswert 0 bezüglich der bedingten Verteilung von \mathbf{y} gegeben \mathbf{u} . Daher ist es angebracht diesen Term als Restterm zu betrachten. Damit ergibt sich

$$\begin{aligned} \left. \frac{\partial^2 h(\mathbf{u})}{\partial \mathbf{u} \partial \mathbf{u}^T} \right|_{\mathbf{u}=\mathbf{u}_0} &\approx -\frac{1}{a(\phi)} \mathbf{Z}^T \mathbf{W}(\boldsymbol{\beta}, \mathbf{u}_0) \Delta(\boldsymbol{\beta}, \mathbf{u}_0) \Delta^{-1}(\boldsymbol{\beta}, \mathbf{u}_0) \mathbf{Z} - \mathbf{D}^{-1} \\ &= -\frac{1}{a(\phi)} \mathbf{Z}^T \mathbf{W}(\boldsymbol{\beta}, \mathbf{u}_0) \mathbf{Z} - \mathbf{D}^{-1}. \end{aligned} \quad (4.10)$$

Setzt man diese Gleichung in (4.4) ein, ergibt sich

$$\begin{aligned} l(\boldsymbol{\beta}) &\approx \log f(\mathbf{y} | \mathbf{u}_0, \theta, \phi) - \frac{1}{2} \mathbf{u}_0^T \mathbf{D}^{-1} \mathbf{u}_0 - \frac{q}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{D}| + \\ &\quad \frac{q}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{Z}^T \mathbf{W}(\boldsymbol{\beta}, \mathbf{u}_0) \mathbf{Z} / a(\phi) + \mathbf{D}^{-1}| \\ &= \log f(\mathbf{y} | \mathbf{u}_0, \theta, \phi) - \frac{1}{2} \mathbf{u}_0^T \mathbf{D}^{-1} \mathbf{u}_0 - \frac{1}{2} \log |\mathbf{Z}^T \mathbf{W}(\boldsymbol{\beta}, \mathbf{u}_0) \mathbf{Z} / a(\phi) + \mathbf{D}^{-1}|. \end{aligned}$$

Um den Maximum-Likelihood-Schätzer $\hat{\beta}$ zu finden ist diese Gleichung bezüglich β zu maximieren. Dazu werden die Ableitungen von l nach jedem zu schätzenden Parameter β betrachtet. Damit ergeben sich die Scoregleichungen

$$\begin{aligned} \frac{\partial l(\beta)}{\partial \beta} &\approx \frac{\partial}{\partial \beta} \log(f(\mathbf{y}|\mathbf{u}_0, \theta, \phi)) + \frac{1}{2} \frac{\partial}{\partial \beta} \log |\mathbf{Z}^T \mathbf{W}(\beta, \mathbf{u}_0) \mathbf{Z} / a(\phi) + \mathbf{D}^{-1}| \\ &\stackrel{(2.11)}{=} \frac{1}{a(\phi)} \mathbf{X}^T \mathbf{W}(\beta, \mathbf{u}_0) \Delta(\beta, \mathbf{u}_0) (\mathbf{y} - \boldsymbol{\mu}) + \frac{1}{2} \frac{\partial}{\partial \beta} \log |\mathbf{Z}^T \mathbf{W}(\beta, \mathbf{u}_0) \mathbf{Z} / a(\phi) + \mathbf{D}^{-1}| \\ &\approx \frac{1}{a(\phi)} \mathbf{X}^T \mathbf{W}(\beta, \mathbf{u}_0) \Delta(\beta, \mathbf{u}_0) (\mathbf{y} - \boldsymbol{\mu}). \end{aligned} \quad (4.11)$$

Für die Approximation in Gleichung (4.11) wurde angenommen, dass $\mathbf{W}(\beta, \mathbf{u}_0)$ sich als Funktion von β nur unwesentlich ändert. Zu Lösen sind folglich die beiden Gleichungen

$$\frac{1}{a(\phi)} \mathbf{X}^T \mathbf{W}(\beta, \mathbf{u}_0) \Delta(\beta, \mathbf{u}_0) (\mathbf{y} - \boldsymbol{\mu}) = 0 \quad (4.12)$$

und

$$\frac{1}{a(\phi)} \mathbf{Z}^T \mathbf{W}(\beta, \mathbf{u}_0) \Delta(\beta, \mathbf{u}_0) (\mathbf{y} - \boldsymbol{\mu}) = \mathbf{D}^{-1} \mathbf{u} \quad (4.13)$$

für β und \mathbf{u} . Um \mathbf{D} zu schätzen ist eine weitere Methode erforderlich.

Ein weitere Methode ist die penalisierte Quasi-Likelihood Methode. Die Idee besteht darin, eine Straffunktion (engl. „penalty function“) zur Quasi-Likelihoodfunktion Q , vgl. (2.16), hinzuzufügen, so dass

$$PQ(\mathbf{y}, \beta, \mathbf{u}) = \sum_{i=1}^n Q(\mathbf{y}_i, \beta, \mathbf{u}_i) - \frac{1}{2} \mathbf{u}^T \mathbf{D}^{-1} \mathbf{u}, \quad (4.14)$$

mit $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)^T$, $\mathbf{y}_i = (u_{i1}, \dots, y_{iT_i})^T$ und $\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_n)^T$.

Die penalisierte Quasi-Likelihood Methode lässt sich in der Praxis nicht zufrieden stellend anwenden. Breslow und Lin (1995) und Lin und Breslow (1996) zeigen, dass diese Methode zu verzerrten Schätzern führt und daher inkonsistent ist. Die Inkonsistenz selbst ist noch nicht das Problem, solange die Verzerrung klein genug ist und kleine Stichproben noch gut sind. Es gibt Fälle, bei denen der penalisierte Quasi-Likelihood-Schätzer sehr schlecht ist. Eine Verbesserung wird erreicht, falls die bedingte Verteilung von \mathbf{y} gegeben \mathbf{u} annähernd normalverteilt ist. In der Praxis versucht man die Daten so zu transformieren, dass sie approximativ normalverteilt sind und greift dann auf das gemischte lineare Modell zurück.

Einen anderen Zugang zur flexiblen Modellierung bieten Bayesianische Ansätze. Mit Markov Chain Monte Carlo (MCMC) Schätzverfahren lassen sich Bayes-Analysen auch für komplexere Problemstellungen durchführen, und über die Maximum a posteriori Schätzung

bestehen enge Bezüge zur penalisierten Likelihood-Schätzung. Auf eine ausführliche Darstellung wird hier verzichtet. Weitere Methoden der Maximum-Likelihood-Schätzung für GLMMs finden sich in Kapitel 10 in McCulloch und Searle (2001) [22].

5 Anwendung auf Mobilitätsdaten

Im folgenden Kapitel soll das Verkehrsmittelwahlverhalten privater Haushalte in Alltag und Freizeit analysiert werden. Im Speziellen sollen neben wegespezifischen Merkmalen auch personen- und haushaltsbezogene Faktoren berücksichtigt werden. Dazu ist es notwendig ein der Realität entsprechendes Modell zu entwickeln, dass eine differenzierte Analyse der Mobilität nach Art, Umfang und Zweck der Ortsveränderung ermöglicht. Grundlage ist das in Kapitel 3 eingeführte multinomiale Logit-Modell.

5.1 Erhebung und Beschreibung der Daten

Die Daten sind der Erhebung Mobilität '97, eine Bayern weit repräsentative Studie zum Mobilitätsverhalten der Verbraucher, entnommen. Sie wurden im Auftrag der BMW AG München am Institut für Sozialökonomie des Haushalts an der Technischen Universität München-Weihenstephan in Zusammenarbeit mit der Abteilung Verkehrsforschung der Infratest Burke Wirtschaftsforschung GmbH & Co., München, erhoben. Die Grundgesamtheit der Studie bilden deutschsprachige Personen, die mindestens 10 Jahre alt sind und in einem privaten Haushalt im Freistaat Bayern leben.

Die Erhebung erfolgte in drei Wellen im März, Juni und Oktober um die saisonalen Schwankungen im Mobilitätsverhalten der Teilnehmer zu berücksichtigen. Jeder Teilnehmer berichtete zwei bzw. drei Tage lang über sämtlich durchgeführte Wege außer Haus.

Zur Beschreibung und Erklärung des Mobilitätsverhaltens kommt den Aktivitätenansätzen der Verkehrswissenschaft eine bedeutende Rolle zu. Sie lassen sich nach Axhausen (1995) [3] bis zum Jahr 1949 zurückverfolgen. Ziel dieser Ansätze war es, auf Grundlage der erfassten Daten die Entwicklung in den einzelnen Verkehrsbereichen abbilden zu können und Planungsempfehlungen abgeben zu können.

Die in dieser Arbeit zur Auswertung gelangende Stichprobe umfasst 167 Haushalte in München mit insgesamt 296 Haushaltsmitgliedern. Diese führten 2794 Wege durch.

Ein Weg führt von einem Standort P_0 mit der Aktivität A_0 und der Startzeit t_0 zu einem Zielort P_1 mit der Aktivität A_1 und der Ankunftszeit t_1 . Dabei wird innerhalb $t_1 - t_0$ die Strecke s_1 mit einem oder mehreren Verkehrsmitteln zurückgelegt. Etappen innerhalb eines Weges wurden bei dieser Erhebung nur qualitativ über die Nennung der benutzten Verkehrsmittel erfasst. Weitere Einflussgrößen werden durch subjektive Bewertungen für die *wegespezifische Verkehrsmittelwahl* erfasst (gekennzeichnet durch Merkmale wie

bequem, kostengünstig, dringend, gutes Wetter, usw.). Vergangenes bzw. grundsätzliches Mobilitäts- und Reiseverhalten spielen ebenfalls eine bedeutende Rolle bei der Verkehrsmittelwahl. Aktivitäten werden durch die Merkmale Erwerbsbereich, Unterhaltungsbereich, sonstige unterhaltswirtschaftliche Aktivitäten, Transferbereich und sonstige Aktivitäten im Modell abgebildet. Aktivitäten sind mobilitätsrelevant, wenn sie außerhalb des eigenen Wohngrundstückes ausgeübt werden. Eine ausführliche Beschreibung findet man in Zängler, Anhang D [29].

Neben der Protokollierung von Wegen und Aktivitäten wurden zusätzlich *erhebungsbezogene, personen- und haushaltsbezogene Merkmale* erfasst.

Die *Erhebungsebene* umfasst neben der Jahreszeit, den Mobilitätstag und Witterungseinflüsse wie die Tagestemperatur, die Sonnenscheindauer und die Niederschlagsmenge.

Personen werden mittels sozialökonomischen Merkmalen wie Alter, Geschlecht, Familienstand, Bildungsstand und Stellung im Beruf charakterisiert. Zu beachten ist, dass Merkmale wie beispielsweise der persönliche Beitrag zum Haushaltseinkommen das Mobilitätsverhalten darüber hinaus beeinflussen können. Ein wichtiges Merkmal für die Verkehrsmittelwahl ist der persönliche Zugang zu Gütern der Mobilität. Als Beispiel seien der Besitz des Führerscheins und die Nutzungsmöglichkeiten von haushaltseigenen und öffentlichen Verkehrsmitteln erwähnt. Daraus können aus dem Verkehrsmittelwahlverhalten der Vergangenheit verkehrsmittelbezogene Präferenzen oder Vorhersagen bestimmt werden.

Haushalte sind unter anderem durch Merkmale wie Anzahl der Personen im Haushalt (Haushaltsgröße), die Art der Zusammensetzung des Haushalts (Haushaltstyp) und das Haushaltseinkommen bzw. das geschätzte Pro-Kopf-Einkommen gekennzeichnet. Die Art der Wohnung und Eigentumsverhältnisse (einschließlich der Besitz eines Wochenendhauses, Wohnwagens oder Nutzgartens außerhalb des Wohngrundstückes) können Aussagen über den Lebensstil geben, aber auch das Verkehrsmittelwahlverhalten entscheidend beeinflussen. Interessant sind ferner Merkmale wie der Zugang zu wichtigen Einrichtungen und zu öffentlichen Verkehrsmitteln. Daher sind private, öffentliche und freie Güter entscheidende Voraussetzung zur Realisierung von Mobilität.

Als abhängige, zu erklärende Variable Y wird die Wahl des Verkehrsmittels mit den vier Responsekategorien

- Öffentlicher Personenverkehr (ÖV), bestehend aus dem öffentlichen Personennahverkehr (S-Bahn, U-Bahn, Straßen- und Stadtbahn, Busse des Nahverkehrs und Fähren/Boote) und dem öffentlichen Personenfernverkehr (Eisenbahn, Fernbus, Flugzeug),
- Verkehrsbereich bestehend aus Fußgängern,
- Verkehrsbereich bestehend aus Fahrrädern,
- Individualverkehr (IV), Verkehrsbereich bestehend aus Pkw, Firmenwagen und Kraftfahrzeugen, aber auch die Fahrt mit einem Taxi

zu Grunde gelegt. Der Response ist vom Typ einer Nominalskala, d.h. die Zahlenwerte stellen nur Etiketten dar und können beliebig vertauscht werden. Eine Anordnungsmöglichkeit ist zum Beispiel

$$Y = \begin{cases} 1 & \text{Fußgänger} \\ 2 & \text{Fahrräder} \\ 3 & \text{ÖV} \\ 4 & \text{IV.} \end{cases}$$

Die unabhängigen Variablen \boldsymbol{x} lassen sich wie beschrieben in

- erhebungsbezogene Variablen
- haushaltsbezogene Variablen
- wegebezogene Variablen
- personenbezogene Variablen

differenzieren. Das in Kapitel 3 eingeführte Multinomiale Logit-Modell ist ein Verfahren, um die Abhängigkeit einer solchen mehrkategorialen abhängigen Variable von anderen unabhängigen Variablen zu untersuchen. Es berechnet die Wahrscheinlichkeit, dass ein interessierendes Ereignis eintritt, wobei eine der vier Kategorien als Referenzkategorie gewählt wird, in Abhängigkeit von den Werten der unabhängigen Variablen. In den folgenden Abschnitten dient IV, als das am häufigsten benutzte Verkehrsmittel, als Referenzkategorie.

5.1.1 Struktur der Daten

In diesem Abschnitt wird die Struktur der Daten beschrieben. Es werden neben den erhebungs- und wegespezifischen Faktoren vor allem Merkmale der Haushalte und Personen in Abhängigkeit von den benutzten Verkehrsmitteln aufgeführt. Allerdings wird sich dieser Abschnitt auf wenige ausgesuchte Merkmale beschränken. Mit Hilfe der Funktion *prozent* (vgl. Anhang B.2) kann für alle weiteren Variablen die relative Häufigkeit angegeben werden. Durch Auf- und Abrunden der ersten Nachkommastelle der Prozentwerte

kann es zu Rundungenungenauigkeiten kommen. Eine Analyse der einzelnen Variablen findet sich in Zängler, Kapitel 5 [29]. Diese bezieht sich auf die gesamte Stichprobe Bayerns.

Erhebungsbezogene Merkmale:

Mobilitätstag:

	Beobachtungen: Mobilitätstag				Anteil in Prozent:			
	Fuß	IV	ÖV	Rad	Fuß	IV	ÖV	Rad
Mo	78	89	61	36	29.5	33.7	23.1	13.6
Di	57	126	56	24	23.5	45.8	19.2	11.5
Mi	122	238	100	60	26.4	39.6	22.4	11.7
Do	86	129	73	38	26.4	39.6	22.4	11.7
Fr	75	150	66	22	24.0	47.9	21.1	7.0
Sa	151	290	50	41	28.4	54.5	9.4	7.7
So	60	207	27	42	17.9	61.6	8.0	12.5

An allen Wochentagen ist IV der häufigsten benutzte Verkehrsbereich. Auffallend ist, dass die Nutzung von ÖV an den Wochenenden stark abnimmt. An den Werktagen Montag bis Freitag ist die Nutzung hingegen in etwa gleich. Am Wochenende finden keine Wege zum Ausbildungsplatz und nur 5 Wege zum Arbeitsplatz mit ÖV statt. Falls am Wochenende dienstliche oder geschäftliche Wege durchgeführt werden, werden diese mit IV zurückgelegt. Daran kann ein Zusammenhang mit dem häufigen Nutzungszweck von ÖV für Arbeits- und Ausbildungswege an Werktagen vermutet werden.

Tageszeit:

	Beobachtungen: Tageszeit				Anteil in Prozent:			
	Fuß	IV	ÖV	Rad	Fuß	IV	ÖV	Rad
Morgen (06.00 bis 11.59)	260	367	160	78	30.1	42.4	18.5	9.0
Nachmittag (12.00 bis 17.59)	240	505	186	128	22.7	47.7	17.6	12.1
Abend (18.00 bis 23.59)	88	278	57	45	18.8	59.4	12.2	9.6
Nacht (0.00 bis 5.59)	4	38	9	3	7.4	70.4	16.7	5.6

Zwischen der zeitlichen Lage einer Fahrt/eines Weges und der Verkehrsmittelwahl lassen sich deutliche Abhängigkeiten aufzeigen. Insbesondere nehmen nach 0:00 Uhr Fahrten mit dem Rad und Wege zu Fuß stark ab. Auffallend ist, dass zu allen Tageszeiten IV, im Vergleich zu allen anderen Verkehrsbereichen, am häufigsten genutzt wird. Die Anzahl der Wege, die zu Fuß zwischen 6 und 24 Uhr zurückgelegt werden sind höher als die Anzahl der Wege die mit öffentlichen Verkehrsmitteln zur selben Tageszeit zurückgelegt werden. Der größte Anteil der Mobilität findet zwischen 6 und 18 Uhr statt.

Wetter:

Beobachtungen: gutes Wetter					Anteil in Prozent:			
	Fuß	IV	ÖV	Rad	Fuß	IV	ÖV	Rad
nein	535	1172	410	194	23.2	50.7	17.7	8.4
ja	94	57	23	69	38.7	23.5	9.5	28.4

Unterschiedliche äußere Bedingungen lassen ein verändertes Verkehrsmittelwahlverhalten vermuten. Nur bei 10% aller durchgeführten Wege hatten die Befragten angegeben, dass das Wetter gut gewesen sei. Bei gutem Wetter werden die Verkehrsbereiche Fuß und Rad erwartungsgemäß häufiger genutzt.

Haushaltsbezogene Merkmale:**Haushaltsgröße:**

Beobachtungen: Haushaltsgröße					Anteil in Prozent:			
	Fuß	IV	ÖV	Rad	Fuß	IV	ÖV	Rad
1	174	190	193	68	27.8	30.4	30.9	10.9
2	234	444	110	54	27.8	52.7	13.1	6.4
3	151	338	65	90	23.4	52.5	10.1	14.0
4+	70	257	65	51	15.8	58.0	14.7	11.5

Aufgrund der Lebensumstände sind zwischen Single- und Mehrpersonenhaushalten Unterschiede bei der Verkehrsmittelwahl zu beobachten. Auffallend ist das im Vergleich mit der Nutzung der anderen Verkehrsbereiche mit zunehmender Haushaltsgröße Wege zu Fuß abnehmen, Wege mit Individualverkehr hingegen zunehmen. Außerdem zeigt sich, dass der größte Teil aller durchgeführten ÖV Wege von Singlehaushalten zurückgelegt werden. Für die Wege, die dem Rad zurückgelegt werden, ist keine Tendenz bei variierender Haushaltsgröße erkennbar.

Haushaltstyp:

Beobachtungen: Haushaltstyp					Anteil in Prozent:			
	Fuß	IV	ÖV	Rad	Fuß	IV	ÖV	Rad
AlleinMKind	27	103	0	0	20.8	79.2	0.0	0.0
ErwHH	86	182	47	54	23.3	49.3	12.7	14.6
Familie	123	309	68	81	21.2	53.2	11.7	13.9
MehrGenHH	11	90	20	8	8.5	69.8	15.5	6.2
Paar	208	355	105	52	28.9	49.3	14.6	7.2
Single	174	190	193	68	27.8	30.4	30.9	10.9

Untersucht man den Haushaltstyp so zeigt sich, dass prozentual gesehen insbesondere Alleinstehende mit Kind/ern am häufigsten IV nutzen. An der Anzahl der zurückgelegten

Wege erkennt man jedoch, dass die Anzahl Alleinstehender mit Kind/ern nur sehr gering ist. Wege mit ÖV werden am häufigsten von Singles zurückgelegt. Dieses Ergebnis kann auch bereits aus der Haushaltsgröße abgelesen werden. Familien und kinderlose Haushalte nutzen zu etwa 50% IV.

Haushaltseinkommen:

Beobachtungen: Haushaltseinkommen					Anteil in Prozent:				
-----	Fuß	IV	ÖV	Rad	-----	Fuß	IV	ÖV	Rad
1: < 2000 DM	80	104	47	26	31.1	40.5	18.3	10.1	
2: 2000 bis 3000 DM	108	137	97	30	29.0	36.8	26.1	8.1	
3: 3000 bis 4000 DM	133	191	92	49	28.6	41.1	19.8	10.5	
4: 4000 bis 5000 DM	61	194	40	40	18.2	57.9	11.9	11.9	
5: 5000 bis 7000 DM	135	294	85	77	22.8	49.7	14.4	13.0	
6: > 7000 DM	98	307	61	41	19.3	60.6	12.0	8.1	

Mobilität ist abgesehen von den Wegen die zu Fuß zurückgelegt werden mit Kosten verbunden. Eine besondere Bedeutung kommt daher dem Einkommen durch seinen Einfluss auf den Pkw-Besitz zu. Mit zunehmendem Haushaltseinkommen nimmt der Anteil der Wege die mit IV gegenüber einem anderen Verkehrsbereich zurückgelegt werden zu. Von den 167 befragten Haushalten fallen 36% in die Kategorie mit niedrigem Haushaltseinkommen (bis 3000 DM). Hierbei ist jedoch darauf zu achten, dass 1/4 dieser Haushalte Singlehaushalte sind. Diese werden bereits zur Hälfte der Kategorie niedriges Einkommen zugerechnet. 2-Personenhaushalte sind nur noch mit 32%, 3-Personenhaushalte mit 7% und 4- und Mehrpersonenhaushalte mit 2% der Kategorie niedriges Haushaltseinkommen zuzurechnen. Umgekehrt fallen nur 2% der Singlehaushalte, aber bereits 77% der 4- und Mehrpersonenhaushalte in die Kategorie hohes Haushaltseinkommen (mehr als 5000 DM). Daraus kann folglich nicht geschlossen werden, dass Haushalte mit niedrigem oder hohem Einkommen einen Einfluss auf die Verkehrsmittelwahl haben.

Pkw im Haushalt vorhanden:

Beobachtungen: Pkw im Haushalt vorhanden					Anteil in Prozent:				
-----	Fuß	IV	ÖV	Rad	-----	Fuß	IV	ÖV	Rad
nein	127	40	144	70	33.3	10.5	37.8	18.4	
ja	502	1189	289	193	23.1	54.7	13.3	8.9	

Die Auswertung des Haushaltfragebogens ergibt, dass von den 167 befragten Haushalten 77% über einen Pkw im Haushalt verfügen. Obige Auswertung zeigt, dass Haushalte mit einem Pkw Verkehrsbereich IV häufiger nutzen als Haushalte, die über keinen Pkw verfügen. Andererseits erkennt man deutlich, dass die Nutzung öffentlicher Verkehrsmit-

tel aber auch die Nutzung der Verkehrsbereiche Fuß und Rad höher sind, wenn kein Pkw vorhanden ist, d.h. wenn kein Fahrzeug zur Verfügung steht, kann es auch nicht genutzt werden. Daher steigen die Chancen ein anderes Verkehrsmittel zu nutzen.

Anbindung an den schienengebundenen Nahverkehr:

Beobachtungen: Anbindung an ÖV					Anteil in Prozent:			
-----					-----			
	Fuß	IV	ÖV	Rad	Fuß	IV	ÖV	Rad
nein	22	126	19	35	10.9	62.4	9.4	17.3
ja	607	1103	414	228	25.8	46.9	17.6	9.7

An den Daten erkennt man, dass bei 92% aller zurückgelegter Wege eine Anbindung an ÖV vorhanden ist. Nur 17.6% der zurückgelegten Wege werden dann tatsächlich mit ÖV durchgeführt. Falls keine Anbindung an ÖV existiert, wird zu 62.4% IV gewählt.

Entfernung zur nächstgelegenen Haltestelle in Kilometer ([km]):

Beobachtungen: km zur nächsten Haltestelle					Anteil in Prozent:			
-----					-----			
	Fuß	IV	ÖV	Rad	Fuß	IV	ÖV	Rad
0	54	43	55	19	31.6	25.1	32.2	11.1
0.1	122	156	58	44	32.1	41.1	15.3	11.6
0.2	91	192	95	32	22.2	46.8	23.2	7.8
0.3	122	304	61	67	22.0	54.9	11.0	12.1
0.4	29	65	11	4	26.6	59.6	10.1	3.7
0.5	124	255	99	64	22.9	47.0	18.3	11.8
0.6	4	16	4	2	15.4	61.5	15.4	7.7
0.7	2	7	0	0	22.2	77.8	0.0	0.0
0.8	2	16	4	4	7.7	61.5	15.4	15.4
1	67	113	29	27	28.4	47.9	12.3	11.4
2	5	41	0	0	10.9	89.1	0.0	0.0
2.5	1	2	5	0	12.5	25.0	62.5	0.0
3	0	14	0	0	0.0	100.0	0.0	0.0

Die Zugangsmöglichkeiten zu öffentlichen Verkehrsmitteln stellen eine wesentliche Voraussetzung zur Nutzung dieser dar. Die Verfügbarkeit der einzelnen öffentlichen Verkehrsmittel im Alltag ist abhängig vom Haushaltsstandort und dessen räumlicher Verkehrsinfrastruktur. Die Werte verdeutlichen dass die nächste Haltestelle häufig weniger als 0.5 km vom Ausgangspunkt entfernt ist. Dennoch ist die Nutzung öffentlicher Verkehrsmittel bei guter Haltestellenerreichbarkeit nicht außergewöhnlich hoch. Mit zunehmender Entfernung zur nächstgelegenen Haltestelle nimmt die Nutzung von IV weiter zu. Die Anzahl der Beobachtungen nimmt jedoch stark ab. Damit kann abermals ein Zusammenhang mit der guten Verkehrsinfrastruktur in München vermutet werden.

Zufriedenheit mit der Anbindung:

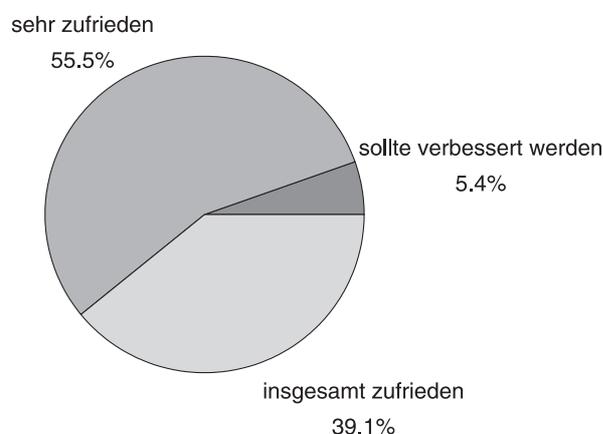


Abbildung 5: Zufrieden mit der Anbindung an ÖV

Die Betrachtung der Zufriedenheit im Zusammenhang mit der Haltestellenerreichbarkeit der verschiedenen öffentlichen Verkehrsmittel liefert einen Hinweis für die Akzeptanz der verfügbaren öffentlichen Verkehrsmittel. Von den 167 in München befragten Haushalten sind 55.5% mit der öffentlichen Verkehrsanbindung an ihrem Wohnort sehr zufrieden.

39.1% sind insgesamt zufrieden und nur 5.4% der Haushalte fordern eine Verbesserung. Letzteres hat daher kaum Einfluss auf die Verkehrsmittelwahl.

Wegebezogene Merkmale:

Entfernung des Weges [km]:

Beobachtungen: Entfernungsklasse

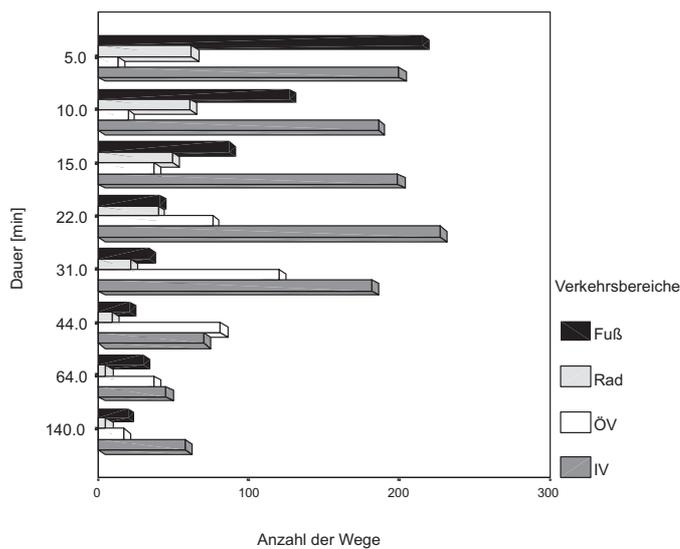
Anteil in Prozent:

Entfernungsklasse	Fuß	IV	ÖV	Rad	Fuß	IV	ÖV	Rad
< 1 km	460	95	9	70	72.6	15.0	1.4	11.0
1 – 5 km	124	434	141	147	14.7	51.3	16.7	17.4
5 – 10 km	12	300	137	33	2.5	62.2	28.4	6.8
10 – 15 km	2	156	64	6	0.9	68.4	28.1	2.6
15 – 20 km	2	36	23	5	3.0	54.5	34.8	7.6
20 – 50 km	0	100	18	0	0.0	84.7	15.3	0.0
> 50 km	0	82	14	1	0.0	84.5	14.4	1.0

Die Verkehrsmittelwahl hängt einerseits von dem Wegenetz des Verkehrsmittels, andererseits von der Distanz ab. Zu berücksichtigen sind Freizeitwege, bei denen die Weglänge häufig einen anderen Stellenwert einnimmt. Abhängigkeiten bestehen unter anderem mit Merkmalen, wie keine andere Möglichkeit, hat sich bewährt, kürzeste Fahrzeit, zur Bewegung, ist flexibel oder passt zur Entfernung. Der größte Teil der durchgeführten Wege liegt bei bis zu 8 km. Mehr als die Hälfte aller Wege ist kürzer als 5 km. Wege mit einer Länge von über 50 km haben einen Anteil von 4%. Während kurze Wege (< 1 km) überwiegend zu Fuß zurückgelegt werden (73%), wird der ÖV nur zu 1% in Anspruch genommen, d.h. der ÖV wird für kurze Wege kaum genutzt. Eine mögliche Ursache liegt sicher darin, dass

es häufig rentabler oder schneller ist, kurze Strecken zu Fuß zu gehen, anstatt den entgeltpflichtigen ÖV zu nutzen. Bei kurzen Strecken wird immerhin noch zu 15% IV genutzt. 11% der kurze Wege werden mit dem Fahrrad zurückgelegt. Die Verwendung des Autos auf Kurzstrecken kann darauf zurückzuführen sein, dass das Auto auf kurzen Teilstrecken zu Besorgungsfahrten genutzt wird oder weil man bereits vorher damit unterwegs war. Unabhängig von der Weglänge wird IV am häufigsten benutzt. Wege zwischen 5 und 20 km werden häufiger mit ÖV zurückgelegt (in etwa 30%). Ab einer Strecke von 20 km wird IV zu 85%, ÖV nur noch zu 15% genutzt. Ein Grund liegt sicher darin, dass mit zunehmender Entfernung die öffentliche Verkehrsinfrastruktur abnimmt.

Dauer der Fahrt/des Weges: Neben der Entfernung spielt die benötigte Zeit eine wesentliche Rolle. Die folgende Abbildung stellt einen Überblick über die Wegedauer und den dazu benutzten Verkehrsbereich dar.



Wege geringer Dauer werden in der Regel zu Fuß oder mit dem Auto zurückgelegt und dienen häufig sog. Besorgungsfahrten, wie Einkäufe. Lange Wege zu Fuß betreffen häufig sog. Rundwege, d.h. Wege die von der Wohnung ausgehend wieder in die Wohnung führen. Dabei werden bis zu 8 km zurückgelegt. 30% der ÖV Nutzer

Abbildung 6: Dauer des Weges/der Fahrt

sind etwa eine halbe Stunde unterwegs. Immerhin noch rund 20% der ÖV Nutzer sind eine dreiviertel Stunde unterwegs. Die folgenden Abbildungen 7 und 8 sollen diesen Zusammenhang noch einmal im Einzelnen veranschaulichen.

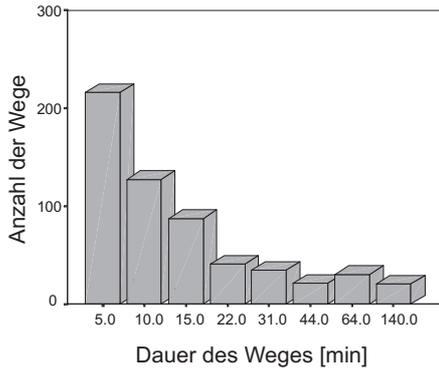


Abbildung 7: Wege zu Fuß

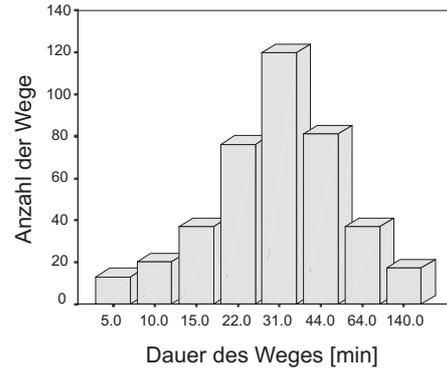


Abbildung 8: Wege mit ÖV

Streckentyp:

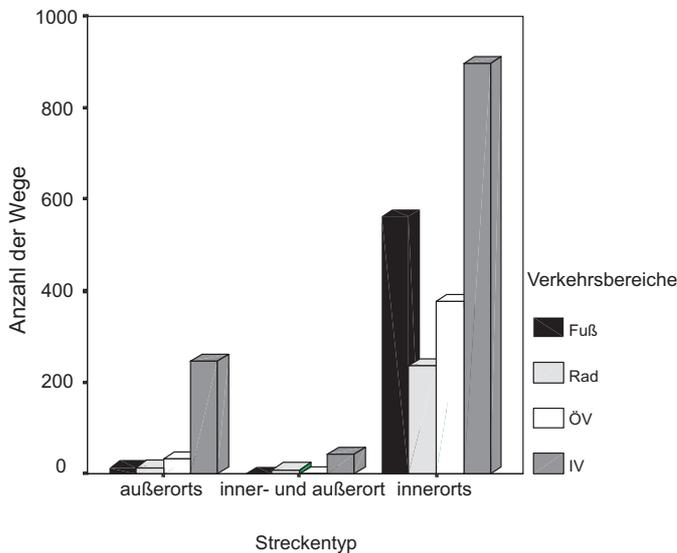


Abbildung 9: Streckentyp

85% der Wege werden innerorts, davon 43% mit IV, 27% zu Fuß, 18% mit ÖV und 11% mit dem Fahrrad zurückgelegt. Von den 15% Wegen die außerorts bzw. inner- und außerorts zurückgelegt werden, werden 81% mit IV, 4% zu Fuß, 9% mit ÖV und 6% mit dem Fahrrad zurückgelegt. Die Ursache für die wenige ÖV Nutzung liegt an der Tatsache, dass

die öffentliche Verkehrsinfrastruktur außerorts abnimmt und daher auf IV zurückgegriffen wird. Die häufige Verwendung von IV auf Wegen außerorts bzw. inner- und außerorts kann auch darauf zurückgeführt werden, dass 41% dieser Wege für Freizeitwecke unternommen werden, von denen wiederum 77% mit dem Auto zurückgelegt werden.

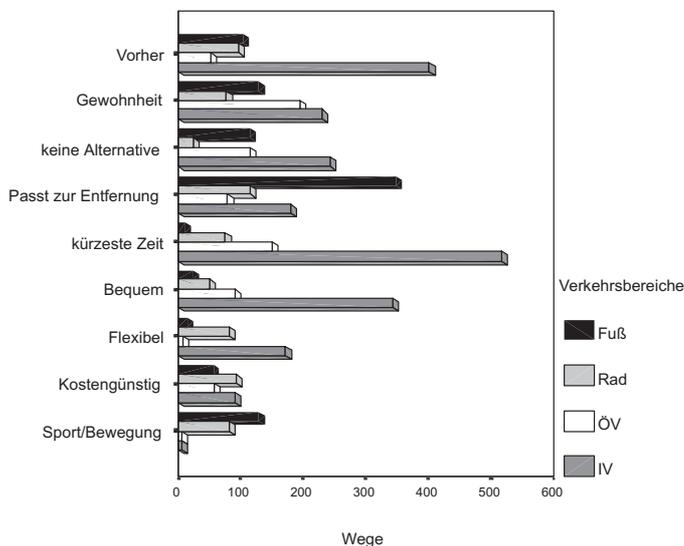
Weg allein oder mit mehreren Personen zurückgelegt:

Beobachtungen: Weg alleine oder mit mehreren	Anteil in Prozent:			
-----	-----			
	Fuß	IV	ÖV	Rad
alleine	387	608	332	216
mit mehreren	229	619	97	46
	25.1	39.4	21.5	14.0
	23.1	62.5	9.8	4.6

Sind mehrere Personen gemeinsam unterwegs, so kann die Verkehrsmittelwahl der gesamten Gruppe durch das individuelle Bedürfnis eines Mitfahrers geprägt werden. 61% aller

Wege werden alleine zurückgelegt. Wege, die nicht alleine zurückgelegt werden, werden zu 62.5% mit dem Verkehrsbereich IV zurückgelegt. Die Ergebnisse zeigen, dass IV Nutzer auf 50% der Wege alleine und auf 50% in Begleitung unterwegs sind. In öffentlichen Verkehrsmitteln sind auf 77% der Wege Personen ohne Begleitung unterwegs. Ein Grund ist, dass ein hoher Anteil der Wege mit öffentlichen Verkehrsmitteln für den Weg zur Arbeit zurückgelegt wird.

Motive der Verkehrsmittelwahl: Motive die das Entscheidungsverhalten bezüglich der Verkehrsmittelwahl beeinflussen können, zeichnen sich durch Angaben von Gründen der Verkehrsmittelwahl bereits zurückgelegter Wege ab. Die folgende Abbildung zeigt mögliche Gründe für die Verkehrsmittelwahl auf.



Die Tatsache, dass sich Personen im öffentlichen Verkehr weniger an das Verkehrsmittel gebunden fühlen als bei IV, zeigt sich an der geringen Nutzungshäufigkeit der öffentlichen Verkehrsmittel von 8% der Wege, bei denen Personen bereits vorher mit ÖV unterwegs sind. Die Entscheidung

Abbildung 10: Gründe für die Verkehrsmittelwahl

für das öffentliche Verkehrsmittel ist sehr häufig durch Gewohnheit geprägt. Die Gründe kürzeste Zeit und Bequemlichkeit werden für IV am häufigsten aufgeführt. Das Verkehrsmittel IV wird häufig auch gewählt, wenn keine anderen Alternativen existieren. Günstigere Kosten werden am häufigsten für die Wahl des Fahrrades genannt.

Eindruck der Fahrt/des Weges:

Beobachtungen: Eindruck der Fahrt des Weges

Anteil in Prozent:

	Fuß	IV	ÖV	Rad	Fuß	IV	ÖV	Rad
angenehm	224	388	72	103	28.5	49.3	9.1	13.1
durchschnittlich	288	714	312	136	19.9	49.2	21.5	9.4
unangenehm	15	49	19	7	16.7	54.4	21.1	7.8

Die Empfindungen während der Fahrt können ebenfalls ein wichtiges Kriterium für die Verkehrsmittelwahl darstellen. Die während einer Fahrt entstandenen Eindrücke können den Entscheidungsprozess beim künftigen Verkehrsmittelwahlverhalten beeinflussen. 34% der Wege mit IV haben beim Nutzer einen angenehmen Eindruck hinterlassen. Als durchschnittlich werden 62% und als unangenehm 4% der Wege mit IV empfunden. Häufige Ursache für unangenehme Wege sind Unfälle, Staus, Baustellen und Witterungsbedingungen. Die Wege zu Fuß werden mit 43% als angenehm, mit 55% als durchschnittlich und nur mit 3% als unangenehm empfunden. Bei 14 der 15 unangenehmen Wege zu Fuß haben die Betroffenen über schlechtes Wetter berichtet, was einen Einfluss auf den negativen Eindruck erkennen lässt. Im öffentlichen Verkehr tritt bei 18% der Wege ein angenehmer, bei 77% ein durchschnittlicher und bei 5% ein unangenehmer Eindruck auf. Als häufigste Ursache für unangenehme Fahrten kommen hier Überfüllung und Verspätungen in Frage.

Wegzweck:

Beobachtungen: Wegzweck	Anteil in Prozent:							
-----	Fuß	IV	ÖV	Rad	Fuß	IV	ÖV	Rad
Arbeit	64	167	117	49	16.1	42.1	29.5	12.3
Ausbildung	10	12	22	16	16.7	20.0	36.7	26.7
Dienstleistungen	74	171	54	40	21.8	50.4	15.9	11.8
Einkauf	261	251	75	59	40.4	38.9	11.6	9.1
Freizeit	138	442	139	74	17.4	55.7	17.5	9.3
Rundwege	45	8	4	3	75.0	13.3	6.7	5.0
Service	19	71	2	8	19.0	71.0	2.0	8.0
dienstl./geschäft.	4	74	15	4	4.1	76.3	15.5	4.1

Ein Weg kann zum Selbstzweck durchgeführt werden oder um einen Zielort zu erreichen, an dem eine bestimmte Aktivität (Arbeiten, Versorgung, Freizeit, usw.) ausgeführt wird. In Abhängigkeit vom Zweck wird das Verkehrsmittel gewählt. Zusammenhänge der Verkehrsmittelwahl existieren mit Merkmalen, wie Einkommen, Pkw-Verfügung, Verfügung über eine Netzkarte oder Bahncard. Mit einem Anteil von 32% dominieren die Wege für Freizeit. Danach folgen Wege zur Versorgung (Einkauf mit 26%) und zur Arbeit (16%). Auffallend ist das Wege für Freizeit zur Hälfte mit IV zurückgelegt werden. Den höchsten IV-Anteil weist mit 76% der Bereich der Dienst- und Geschäftswege auf. Der größte ÖV-Anteil findet sich bei Wegen im Bereich der Ausbildung. Dort nimmt der ÖV mit 37% den größten Anteil ein. Ein großer Anteil an Fußwegen findet sich im Bereich der Rundwege.

Dringlichkeit:

Beobachtungen: Dringlichkeit					Anteil in Prozent:			
	Fuß	IV	ÖV	Rad	Fuß	IV	ÖV	Rad
nein	463	778	294	160	27.3	45.9	17.3	9.4
ja	166	451	139	103	19.3	52.5	16.2	12.0

Diese Größe nimmt einen bedeutenden Stellenwert bei der Verkehrsmittelwahl ein. Abhängigkeiten bestehen zur zeitlichen Dauer, um den Weg zurückzulegen. Wege die dringend durchgeführt werden müssen, werden zu 52% mit IV zurückgelegt. Aber auch bei Wegen die als nicht dringend eingestuft werden, wird zu 46% Verkehrsbereich IV in Anspruch genommen. Die Nutzung von ÖV scheint in etwa gleich.

Personenbezogene Merkmale:**Nutzung und Alter:**

Beobachtungen: Nutzung					Anteil in Prozent:			
	Fuß	IV	ÖV	Rad	Fuß	IV	ÖV	Rad
Pkw-Hauptnutzer	280	909	137	83	19.9	64.5	9.7	5.9
Pkw-Mitnutzer	163	238	110	84	27.4	40.0	18.5	14.1
sonstiges	186	82	186	96	33.8	14.9	33.8	17.5

Erwartungsgemäß ist die Nutzung von IV von Pkw-Hauptnutzern am höchsten.

Beobachtungen: Alter					Anteil in Prozent:				
	Fuß	IV	ÖV	Rad	Fuß	IV	ÖV	Rad	
18+ bis 31	31	155	276	108	52	26.2	46.7	18.3	8.8
31+ bis 42	42	159	405	71	87	22.0	56.1	9.8	12.0
42+ bis 55	55	117	333	91	88	18.6	52.9	14.5	14.0
55+ bis 65	65	137	160	96	27	32.6	38.1	22.9	6.4
> 65		55	50	55	9	32.5	29.6	32.5	5.3

Prozentual gesehen nutzen insbesondere Personen ab 55 Jahren öffentliche Verkehrsmittel. In dieser Personengruppe werden auch am häufigsten Wege zu Fuß zurückgelegt. Die häufigste Nutzung von IV findet sich in der Altersgruppe der 31 bis 42 - jährigen. Die Nutzung des Fahrrades hat in der Altersgruppe der 42 bis 55 - jährigen ihre Spitze.

Vor allem Personen, die keinen Führerschein besitzen oder keinen Pkw zur Verfügung haben sind unter anderem durch die zunehmende Zentralisierung von Einrichtungen für das alltägliche Leben auf öffentliche Verkehrsmittel angewiesen. Diese Personen können den Verkehrsbereich IV nicht nutzen. Für diese Nutzergruppe (Kategorie *sonstiges*) steigt die Chance die drei Verkehrsbereiche Fuß, ÖV oder Rad zu nutzen. Folglich muss zwischen

Personen, die einen Pkw steuern dürfen von denen die keinen Pkw steuern dürfen und daher den Pkw nur als Mitfahrer nutzen können unterschieden werden. 82% der befragten Personen besitzen einen Führerschein. Dieser hohe Anteil ist darauf zurückzuführen, dass bei der Modellentwicklung Personen unter 18 Jahren eliminiert worden sind. Der Anteil der Führerscheinbesitzer sinkt in der Altersgruppe ab 55 Jahren jedoch wieder deutlich ab. Die Neigung, die Fähigkeit und die Art Wege außer Haus durchzuführen beeinflussen die Verkehrsmittelwahl und ändern sich mit dem Alter. Beziehungen mit Merkmalen, wie Führerscheinbesitz, Pkw-Verfügbarkeit, Erwerbstätigkeit oder auch das Einkommen sind durchaus denkbar. Die Altersgruppe der 31 bis 42-jährigen nutzt am häufigsten IV. Die folgende Abbildung stellt nochmal einen Überblick über die verwendeten Verkehrsbereiche in den verschiedenen Altersgruppen dar.

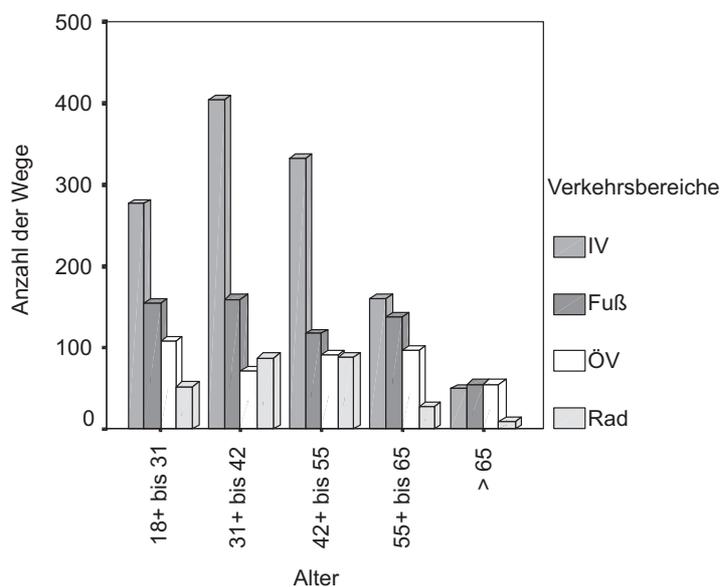


Abbildung 11: Anzahl der zurückgelegten Wege in den verschiedenen Altersgruppen nach Verkehrsbereichen

In den Altersgruppen der 18 bis 65-jährigen dominieren eindeutig Wege mit IV als das am häufigsten benutzte Verkehrsmittel. Ab der Altersgruppe der 55-jährigen nimmt die Anzahl der zurückgelegten Wege stark ab. Ein Grund liegt darin, dass ältere Menschen insgesamt weniger mobil sind.

Ab einem Alter von 65 Jahren gibt es kaum noch einen Unterschied zwischen der Anzahl der Wege die zu Fuß, mit öffentlichen Verkehrsmitteln oder mit dem Auto zurückgelegt werden. Das Fahrrad wird bei älteren Menschen fast nicht mehr benutzt. Die Ursache kann möglicherweise auf Unsicherheiten im Umgang mit dem Fahrrad zurückgeführt werden.

Geschlecht

Beobachtungen: Geschlecht

Anteil in Prozent:

	Fuß	IV	ÖV	Rad	Fuß	IV	ÖV	Rad
männlich	266	628	178	145	21.9	51.6	14.6	11.9
weiblich	36	601	255	118	27.2	45.0	19.1	8.8

Die Bedeutung des Merkmals Geschlecht hängt unter anderem mit den traditionellen Rollenbildern, wie die Stellung im Haushalt, die Erwerbstätigkeit, der Pkw-Führerschein oder -Besitz zusammen. Männer nutzen IV zu 52%, während Frauen IV mit 45% nutzen. Der Anteil der Wege, die zu Fuß und mit öffentlichen Verkehrsmitteln zurückgelegt werden, ist bei Frauen etwas höher als bei Männern.

Netzkarte:

Beobachtungen: Netzkarte

Anteil in Prozent:

	Fuß	IV	ÖV	Rad	Fuß	IV	ÖV	Rad
ja	251	250	277	71	29.6	29.4	32.6	8.4
nein	378	979	156	192	22.2	57.4	9.1	11.3

Man erkennt deutlich, dass Netzkartenbesitzer häufiger ÖV benutzen als Nicht-Netzkartenbesitzer. Von den Befragten die mit öffentlichen Verkehrsmitteln unterwegs waren besitzen 64% eine Netzkarte. Insgesamt besitzen nur 39% der Befragten Personen eine Netzkarte. Unter den Begriff Netzkarte fallen Monatskarten, Netzkarte und Grüne Karten.

Persönliches Einkommen:

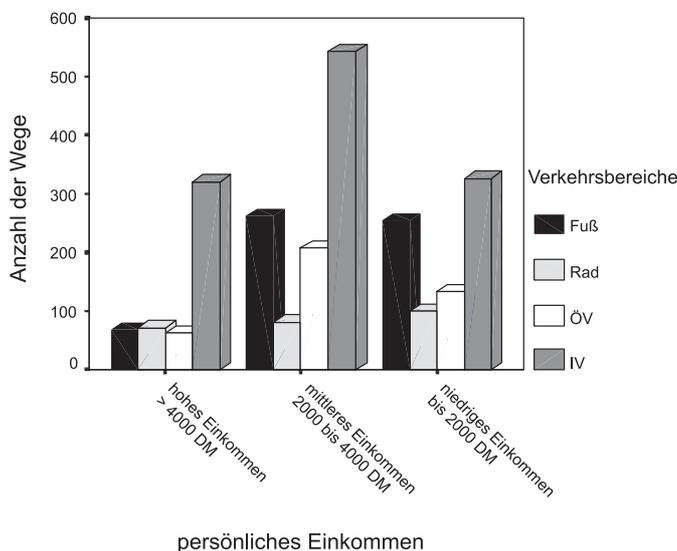


Abbildung 12 zeigt, dass bei mittlerem persönlichem Einkommen der größte Anteil der Wege mit IV zurückgelegt werden. Personen mit niedrigem persönlichem Einkommen legen anteilmäßig am häufigsten Strecken zu Fuß zurück.

Abbildung 12: Anzahl der zurückgelegten Wege in den verschiedenen Einkommensgruppen

Für das Verhalten Wege per Rad zurückzulegen ist keine Tendenz ersichtlich. Durch den Einfluss auf den Pkw- oder Rad-Besitz, die Fahrtenhäufigkeit oder den Bahncard-Besitz gewinnt diese Größe bei der Verkehrsmittelwahl an Bedeutung.

Rad-Verfügung:

Beobachtungen: Radverfügung					Anteil in Prozent:			
	Fuß	IV	ÖV	Rad	Fuß	IV	ÖV	Rad
nein	137	137	105	0	36.1	36.1	27.7	0.0
ja	492	1092	328	263	22.6	50.2	15.1	12.1

81% der befragten Personen sind im Besitz eines Fahrrades. Die Ergebnisse zeigen, dass bei Rad-Verfügung nur 12.1% aller zurückgelegten Wege mit dem Fahrrad durchgeführt werden.

Bahncard:

Beobachtungen: Bahncard					Anteil in Prozent:			
	Fuß	IV	ÖV	Rad	Fuß	IV	ÖV	Rad
ja	81	138	80	40	23.9	40.7	23.6	11.8
nein	548	1091	353	223	24.7	49.3	15.9	10.1

13% der befragten Personen sind im Besitz einer Bahncard. Bei Besitz einer Bahncard werden Wege, prozentual betrachtet, häufiger mit ÖV zurückgelegt, als bei Nichtbesitz. Bereits 47% der Bahncardbesitzer verfügen über ein mittleres persönliches Einkommen (2000 DM bis 4000 DM netto), 22% verfügen bereits über ein hohes persönliches Einkommen (mehr als 4000 DM netto).

5.1.2 Versuchsplanung und Datenaufbereitung

Grundsatz für die statistische Versuchsplanung ist die Verwendung untereinander unabhängiger Einflussparameter, damit Einflüsse richtig zugeordnet werden können. Wichtig ist auch das Einbeziehen aller möglichen Einflussparameter. Ist die Anzahl der möglichen Einflussparameter sehr groß, müssen zunächst Haupteinflussgrößen herausgefiltert werden. Der Datensatz, der in dieser Arbeit verwandt wird, umfasst weit mehr als 1000 Einflussgrößen. Zunächst können die Bestimmungsgrößen untersucht werden, von denen ein Einfluss auf die Verkehrsmittelwahl erwartet wird. Wichtige Merkmale wurden bereits im vorangegangenen Abschnitt vorgestellt. Die vorhandenen Einflussgrößen werden zunächst sinnvoll zu Kategorien zusammengefasst. Die Variablen wurden zur Beschreibung der Daten benannt und soweit es sich nicht um metrische Variablen handelte mit Labels versehen. Details können in Zängler [29] nachgelesen werden.

Die Kodierung einiger ausgewählter Variablen findet sich in Anhang C. Eine Übersicht aller kodierten Variablen findet sich auf beigefügter CD-ROM in der Datei Datensatzbeschreibung.pdf.

Durch die gleichzeitige Berücksichtigung mehrerer unabhängiger Variablen im Modell wird eine Überbewertung einzelner Merkmale entgegengewirkt. Je mehr Variablen in das Modell aufgenommen werden, desto größer ist die Gefahr der Multikollinearität zwischen den unabhängigen Variablen. Multikollinearität der Einflussgrößen heißt, dass die Spalten der Matrix \mathbf{Z} (vgl. Abschnitt 3.7) *fast* linear abhängig sind. Der vorliegende Datensatz enthält Variablen, die nicht notwendigerweise voneinander unabhängig sind. Daher mussten die Variablen geeignet (entsprechend der zu Grunde liegenden Fragestellung) zusammengefasst werden. Eine Person die zum Zeitpunkt der Befragung eine Schulausbildung absolviert, wird in der Regel nicht bereits Student, Angestellter oder Beamter sein. Folglich sind diese Größen nicht unabhängig voneinander und müssen mit großer Sorgfalt analysiert werden. Die Größe Tätigkeit beispielsweise, mit den drei Kategorien in Ausbildung, berufstätig und nicht erwerbstätig, wird aus 22 dichotomen Variablen (u.a. ohne Schulabschluss, Mittlere Reife, Abitur, ohne Berufsausbildung, Meister, Angestellter, Hochschulabschluss, usw.) generiert. Das Merkmal Pkw-Hauptnutzer beinhaltet neben dem Besitz eines entsprechenden Führerscheins auch die Verfügbarkeit über einen Pkw, d.h. die Variablen sind nicht unabhängig. Die dichotomen Größen Führerschein bzw. Pkw-Verfügbarkeit sollten daher nicht gemeinsam mit dem Merkmal Pkw-Hauptnutzer in das Modell aufgenommen werden. Besser ist es *eine* Variable zu verwenden, die diese Merkmale enthält, unter der Prämisse, dass die Variablen im Modell unabhängig sein müssen. Würden diese Größen gemeinsam in ein Modell mit mehreren Einflussgrößen aufgenommen, erhält man Schätzungen nahe 0. Aus diesem Grund wird die neue Variable Nutzung mit den drei Kategorien Pkw-Hauptnutzer, Pkw-Mitnutzer und sonstiges generiert. Damit wird das Modell einerseits übersichtlicher, andererseits lassen sich die zu schätzenden Parameter reduzieren. Wie die Variablen im Einzelnen zusammengefasst wurden, kann den Kodierungsplänen entnommen werden. Auf die Zusammenfassung wichtiger Variablen wird ggf. im Einzelnen noch einmal separat darauf hingewiesen. Neben einer empirischen Untersuchung (siehe Abschnitt 3.6) ist eine genaue Vorauswahl der erklärenden Variablen unabdingbar. Ist man andererseits an der Verkehrsmittelwahl von Ledigen, Studenten oder Führerscheinbesitzern interessiert, sollte ein Zugriff auf diese dichotomen Variablen weiterhin möglich sein. Der zu Grunde liegende Datensatz liefert häufig dichotomen Variablen mit nur sehr wenigen Beobachtungen, wie beispielsweise die Größe Wehrdienstleistender. Der Grund liegt darin, dass nur 267 Personen in die Stichprobe eingehen. Ein Einfluss auf die Verkehrsmittelwahl ist in einem solchen Fall nicht nachweisbar. Um den Einfluss zu überprüfen muss eine größere Stichprobe verwandt werden.

Der Pkw ist heute das Hauptverkehrsmittel der Münchener. 47% aller Wege werden mit dem Pkw oder anderen Verkehrsmitteln, die zum Individualverkehr hinzugefügt werden

zurückgelegt. Dagegen werden nur zu 18% öffentliche Verkehrsmittel benutzt. Immerhin wird fast jeder vierte Weg zu Fuß zurückgelegt. Wege die von Personen durchgeführt worden sind, die keinen Personenfragebogen abgegeben haben, werden eliminiert. Eine besondere Bedeutung nimmt der ÖV für Personen ein, die z.B. aus Altersgründen noch nicht motorisiert sind. Für diese Personengruppen ist eine spezifischere Analyse notwendig, die hier allerdings nicht betrachtet wird. Personen die nicht älter als 18 Jahre alt sind werden daher bei der Analyse vernachlässigt. Die Anzahl der Wege für diese Untersuchung reduziert sich damit von $n = 2794$ auf $n = 2554$ Wege.

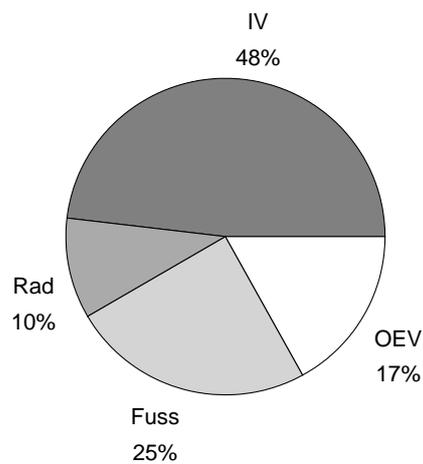


Abbildung 13: Anteil der Wege ($n=2554$ Wege) nach Verkehrsbereichen

5.2 Datenanalyse

5.2.1 Explorative Datenanalyse zum Erkennen von Haupteffekten

Mit Hilfe der explorativen Datenanalyse (vgl. Abschnitt 3.6) können Haupteffekte erkannt werden. Die explorative Analyse ist folglich ein hilfreiches Instrument zur Modellbildung. Die im vorangegangenen Abschnitt beschriebenen Variablen stellen nur ein sehr kleinen Teil der Größen dar, die tatsächlich zum Modellaufbau untersucht werden. Mit Hilfe der Plots wird eine Vorauswahl über mögliche Einflussgrößen getroffen, die in Tabelle 9 aufgeführt sind. Schließlich werden in etwa 40 Größen in die engere Auswahl einbezogen. Außerdem kann mit Hilfe der Plots erkannt werden, ob kategoriale Variablen noch weiter zusammengefasst werden können. Durch weiteres Zusammenfassen der Variablen und entsprechender Behandlung von Kollinearitäten zwischen den Variablen kann die Anzahl möglicher Einflussgrößen weiter reduziert werden. Dieses Vorgehen wurde in Abschnitt 3.6 bereits beschrieben. Einige Variablen, die aus sachwissenschaftlicher Sicht einen Einfluss haben (z.B. Fahrgemeinschaft), fallen heraus, da nur wenige Beobachtungen vorhanden sind. Hier ist es sinnvoll einen größeren Stichprobenumfang zu wählen, zum Beispiel ganz Bayern oder ganz Deutschland. In dieser Arbeit wird darauf allerdings verzichtet. Beispielfhaft sind die wichtigsten Plots in Anhang D.1 dargestellt.

Der Datensatz enthält neben kategorialen Variablen einige metrische Variablen wie Alter, Entfernung, Niederschlagsmenge, Sonnenscheindauer, Temperatur, Dauer des Weges und Geschwindigkeit. Ziel ist es diese beizubehalten. Mit Hilfe des in Abschnitt 3.7.3 eingeführten Akaike-Kriteriums wird geprüft, ob es sinnvoll ist die Variablen metrisch oder kategorial in das Modell aufzunehmen. Die Untersuchung ergibt, dass die Größen Alter, Entfernung, Wegedauer und Entfernung zu nächstgelegenen Haltstelle metrisch in das Modell eingehen können. Für die explorative Datenanalyse (vgl. Abschnitt 3.6) müssen diese Größe jedoch zunächst gruppiert werden. Welche Größe letztendlich in das Modell aufgenommen wird, hängt von der zu untersuchenden Fragestellung ab. Da die Mobilitätsstudie '97 zahlreiche Größen enthält, die für die Verkehrsmittelwahl nicht notwendig sind, konnten viele Größen im Vorfeld eliminiert werden.

Folgende Tabelle gibt einen Überblick über den Einfluss der getesteten Variablen. Die rechte Spalte gibt an, ob die Variable in die engere Wahl für das Haupteffekt-Modell aufgenommen wird.

Variable	Effekt
Erhebungsbezogene Variablen	
Befragungszeitraum (Welle)	sehr geringer Einfluss
Mobilitätstag	Einfluss
Haushaltsbezogene Variablen	
Hauhaltstyp	Einfluss
Haushaltsgröße	Einfluss
Anzahl der Kinder im Haushalt	Einfluss
Haushaltseinkommen	Einfluss
Mietstatus	kein Einfluss
Ausstattung mit Wochenendhaus	kein Einfluss
Ausstattung mit Garten	kein Einfluss
Ausstattung mit haushaltstechnischen Geräten	kein Einfluss
Ausstattung mit elektronischen Gräten	kein Einfluss
Besitz von mind. einem Hund	kein Einfluss (zu wenig Beobachtungen)
Besitz von mind. einem anderen Haustier	kein Einfluss
Anbindung an schienengebundenen Nahverkehr	Einfluss
Entfernung zur nächsten Haltestelle [km]	Einfluss
Dauer zur nächsten Haltestelle [min]	kein Einfluss
Zufriedenheit mit dem ÖV	kein Einfluss
Parkplatzsituation zu Hause	kein Einfluss
Ortsgröße gut	kein Einfluss

Wegebezogene Variablen	
Niederschlag	kein Einfluss
Sonne	geringer Einfluss
Temperatur	kein Einfluss
gutes Wetter	Einfluss
Entfernung	Einfluss
Streckentyp	Einfluss
Weg allein oder mit mehreren schon vorher damit unterwegs	Einfluss
Gewohnheit	Einfluss
keine andere Möglichkeit	geringer Einfluss
passt zur Entfernung	Einfluss
geringster Zeitaufwand	Einfluss
viel Gepäck	kein Einfluss (zu wenig Beobachtungen)
bequem	geringer Einfluss
flexibel	Einfluss
kostengünstig	Einfluss
Sport/Bewegung	Einfluss
Freizeiterlebnis	kein Einfluss (zu wenig Beobachtungen)
lag auf dem Weg	kein Einfluss (zu wenig Beobachtungen)
Improvisation	kein Einfluss (zu wenig Beobachtungen)
Stop war kurzfristig geplant	kein Einfluss (zu wenig Beobachtungen)
Variabel	kein Einfluss (zu wenig Beobachtungen)
Stop war dringlich	kein Einfluss
Stop war geplant	kein Einfluss
Eindruck der Fahrt	Einfluss
Ortstypen (Zielort)	Einfluss
Start-Tageszeit	Einfluss
Dauer der Fahrt/des Weges	Einfluss leicht quadratisch bei IV-ÖV
Personenbezogene Variablen	
Altersgruppe	leicht quadratischer Einfluss bei IV-ÖV
Geschlecht	geringer Einfluss
Familienstand	Einfluss
Nutzung (Haupt-, Mitnutzer, sonstige)	Einfluss
Rad-Verfügung	Einfluss
Motorrad-Verfügung	Einfluss
Motorrad-Führerschein	kein Einfluss
Tätigkeit	geringer Einfluss
persönliches Einkommen	Einfluss
Netzkarte	Einfluss
Bahncard	Einfluss

Bahnsonderangebote	geringer Einfluss
Fahrgemeinschaft	kein Einfluss (zu wenig Beobachtungen)
Ziel- / Wegzweck	Einfluss
Einstellung zum IV	geringer Einfluss
Einstellung zum ÖV	kein Einfluss

Tabelle 9: Ergebnisse der Explorative Analyse für marginale Effekte einzelner Kovariablen

5.2.2 Haupteffektmodell

Im Weiteren muss untersucht werden, ob die explorativ gefundenen Einflussgrößen tatsächlich zur Erklärung der Verkehrsmittelwahl beitragen. Wie in Abschnitt 3.7.6 erwähnt ist eine Untersuchung aller Submodelle unmöglich. Ein mögliches Vorgehen für die weitere Analyse wurde dort aufgezeigt. Mit Hilfe des Likelihood-Ratio-Tests (LRT) können sukzessive signifikante Größen hinzugefügt werden. Größen die keinen signifikanten Beitrag leisten sind sofort wieder zu eliminieren. Für die Analyse des Verkehrsmittelwahlverhaltens werden zunächst $k - 1 = 3$ Binomialmodelle entwickelt und schließlich alle signifikanten Größen in einem multinomialen Modell zusammengefasst.

Bemerkung 5.2.1. In der Regel besitzen metrische Variablen verschieden Werte. Dies führt häufig zu einer großen Anzahl von Freiheitsgraden. In Beispiel 3.7.6, Seite 66 wurde dieses Problem vorgestellt. Es konnte gezeigt werden, dass nach dem *Scoren* der Variablen die Anzahl der Freiheitsgrade deutlich geringer ist. Aus diesem Grund werden die metrischen Variablen Alter, Entfernung, Dauer des Weges und die Entfernung zur Haltestelle gescort. Man erkennt dies im Folgenden daran, dass diesen Größen ein **S** vor dem Variablennamen vorangestellt wird.

Bemerkung 5.2.2. Die Größe Alter wird außerdem für die weitere Betrachtung normiert. Die Normierung wird mit einem **N** im Variablennamen angedeutet. Die Variable Alter wird daher im Folgenden mit **S.ALTER.N** bezeichnet.

Nach der Durchführung der Vorwärtsanalyse ergibt sich für die Mobilitätsstudie folgendes Haupteffekt-Modell:

Effekt	Resid. Dev.	LRT	Freiheitsgrade	p-Wert
volles Modell	1734.624	.000	0	–
POLY.S.ALTER.N1	1739.416	4.792	3	.188
POLY.S.ALTER.N2	1743.331	8.707	3	.033
S.ENTF.KM	2223.465	488.841	3	.000
S.DAUERMIN	1948.478	213.854	3	.000
S.KM.OEV.MIN	1757.663	23.039	3	.000
NUTZUNG	1861.848	127.224	6	.000
HSSCHIEN	1749.090	14.466	3	.002
WEG.MIT	1789.984	55.360	3	.000
STR.TYP.NEU	1790.218	55.594	3	.000
HH.GROESSE	1799.580	64.956	9	.000
KOSTEN	1766.453	31.829	3	.000
FLEXIBEL	1778.387	43.763	3	.000
OHNE.ALT	1747.199	12.575	3	.006
GEWOHN	1788.627	54.003	3	.000
VORHER	1753.907	19.283	3	.000
MB.TAG	1812.620	77.996	18	.000
NETZKARTE	1795.493	60.869	3	.000
WETT.GUT	1773.657	39.033	3	.000
EINDRUCK	1769.002	34.378	6	.000
PERS.EINK.NEU	1766.883	32.259	6	.000
GUT.ENTF	1843.190	108.566	3	.000
BEQUEM	1772.156	37.532	3	.000
ZUFR.ANB.OEV	1755.549	20.925	6	.002
EINPKW.N	1747.779	13.155	3	.004
SPORT	1872.564	137.940	3	.000
KURZ.T	1821.169	86.545	3	.000
HH.EINK.NEU	1740.160	5.536	6	.477
BAHNCARD	1743.158	8.534	3	.036
RAD.VERF	1761.292	26.668	3	.000

Tabelle 10: Likelihood-Quotienten-Tests der Haupteinflussgrößen

Die p-Werte zeigen, dass die Variablen bis auf Alter und Haushaltseinkommen signifikant sind.

5.2.3 Explorative Datenanalyse zum Erkennen von Interaktionen

Parameter die das Mobilitätsverhalten beeinflussen können einer objektiv erkennbaren Handlungsebene oder einer subjektiv beurteilbaren Bewusstseinssebene entstammen. Nach der Wahl eines Haupteffekt-Modells können Beziehungen zueinander untersucht werden (Interaktionen). Die Wahrscheinlichkeit bei der Verkehrsmittelwahl das Auto zu benutzen hängt von der Verfügbarkeit eines Pkw ab. Es ist beispielsweise von Interesse, zu untersuchen, ob die Variation der Verfügbarkeit eines Pkw für die verschiedenen Alters-

gruppen identisches Verkehrsmittelwahlverhalten zur Folge hat. Neben der Einflussgröße Pkw-Nutzung wird die metrische Variable Alter als weitere Kovariable betrachtet. Beispiel 3.6.3 zeigt, wie man bei der Suche nach Interaktionen explorativ vorgehen kann.

Individuelle Lebenssituationen können bestimmte Aktivitäten und damit bestimmte Mobilität ermöglichen, aber auch untersagen. Insbesondere bei der Verkehrsmittelwahl können äußere Lebensbedingungen den Handlungsspielraum begrenzen oder sogar die Nutzung von Alternativen vollkommen unmöglich machen. Beispiele für Interaktionen, die untersucht werden können werden im Folgenden kurz dargestellt:

Alter: Alter ist eine wichtige Bestimmungsgröße für die Verkehrsmittelwahl. Die Fähigkeit und die Art, Aktivitäten außer Haus durchzuführen ändert sich mit dem Alter (Holz-Rau, 1990, S. 39 [18]). Es existieren starke Beziehungen mit Merkmalen, wie Pkw-Nutzung oder persönliches Einkommen.

Einkommen: Diese Größe ist insbesondere durch seinen Einfluss auf die Pkw-Nutzung und den Bahncard-Besitz von Bedeutung.

Pkw-Nutzung: Die Erlaubnis einen Pkw zu fahren teilt die Verkehrsteilnehmer in zwei Gruppen. Personen ohne Führerschein können den Pkw nur als Mitfahrer nutzen. Weitere Abhängigkeiten sind zu den Merkmalen wie Alter oder Einstellung zum IV zu vermuten.

Rad-Verfügbarkeit: Steht kein Rad zur Verfügung, kann es auch nicht genutzt werden. Interaktionen können zwischen der physischen Verfügbarkeit zu den Größen Einkommen, Haushaltsgröße, Alter, Pkw-Nutzung oder Einstellung zum IV bestehen.

Anbindung an den öffentlichen Nahverkehr: Voraussetzung für die Nutzung des ÖV ist das Vorhandensein einer Zugangsmöglichkeit zum öffentlichen Nahverkehr. Interaktionen können zwischen der ÖV-Anbindung und der Zufriedenheit mit öffentlichen Nahverkehr bestehen.

Beispiel 3.3.2 zeigt neben der explorativen Untersuchung eine weitere Methode Interaktionen festzustellen. In Abschnitt 3.7.4 wurden verschiedene Test-Methoden vorgestellt. Mit Hilfe der Funktion *interaktion* (vgl. Anhang B.3) und mit Hilfe des Multinomialen Modells werden oben aufgeführte Größen zunächst explorativ auf Interaktionen getestet.

Ergebnisse

Beispielhaft sind einige Interaktionplots in Anhang D.2 dargestellt. Mit der explorativen Datenanalyse ergeben sich Interaktionen zwischen:

- Alter und Nutzung, Haushaltseinkommen sowie persönliches Einkommen
- persönliches Einkommen und Bahncard
- Haushaltseinkommen und Bahncard sowie Pkw-Nutzung

Zwischen

- Alter und Rad-Verfügbarkeit
- Haushaltsgröße und Rad-Verfügbarkeit sowie Nutzung
- Haushaltseinkommen und Rad-Verfügbarkeit sowie Bahncard
- persönliches Einkommen und Rad-Verfügbarkeit
- persönliches Einkommen und Haushaltseinkommen
- Zufriedenheit mit der Anbindung an ÖV und der Haltestellenerreichbarkeit sowie positive Einstellung zum ÖV

können keine Interaktionen festgestellt werden.

5.2.4 Modell mit Interaktionen

Nicht alle festgestellten Interaktionen gehen tatsächlich in das Modell ein. Für die Mobilitätsstudie ergibt sich, nach der Anpassung (vgl. Abschnitt 3.7.5), das im folgenden beschriebene Modell. Die erste Tabelle (Likelihood-Quotienten-Tests) gibt einen Überblick über alle im Modell enthaltenen Interaktionen. Die zweite Tabelle enthält neben den Parameterschätzungen $\hat{\beta}$, die geschätzten Standardfehler für $\hat{\beta}$, die Wald-Statistiken, die p-Werte und die odds-ratios (vgl. Abschnitt 3.7). Sie ist zudem in erhebungsbezogene, haushaltsbezogene, wegebezogene und personenbezogene Merkmale untergliedert. Dies wird jeweils durch eine Trennzeile angezeigt.

Effekt	Resid. Dev.	LRT	Freiheits- grade	p-Wert
NUTZUNG*HH.EINK.NEU	1621.426	99.347	12	.000
PERS.EINK.NEU*HH.EINK.NEU	1589.170	67.091	9	.000
NUTZUNG*POLY.S.ALTER.N1	1562.789	40.710	6	.000
NUTZUNG*POLY.S.ALTER.N2	1541.584	19.505	6	.003
PERS.EINK.NEU*BAHNCARD	1571.326	49.247	6	.000

Tabelle 11: Likelihood-Quotienten-Tests der Interaktionen

Parameterschätzungen

Effekt	Verkehrsmittel	Schätzwert $\hat{\beta}$	geschätzter Std.fehler	Wald ω	p-Wert	exp(β)
Intercept	Fuß	6.621	1.745	14.379	0.000	–
	ÖV	-3.553	1.373	6.694	0.010	–
	Rad	0.229	2.002	0.013	0.909	–
MB.TAG=Di	Fuß	-2.482	0.661	14.094	0.000	0.1
	ÖV	-1.930	0.478	16.324	0.000	0.1
	Rad	-0.638	0.610	1.093	0.296	0.5
MB.TAG=Mi	Fuß	-1.283	0.549	5.465	0.019	0.3
	ÖV	-1.797	0.403	19.907	0.000	0.2
	Rad	-0.757	0.535	2.004	0.157	0.5
MB.TAG=Do	Fuß	-1.228	0.570	4.634	0.031	0.3
	ÖV	-1.886	0.431	19.192	0.000	0.2
	Rad	-1.171	0.600	3.807	0.051	0.3
MB.TAG=Fr	Fuß	-1.277	0.550	5.399	0.020	0.3
	ÖV	-2.193	0.423	26.831	0.000	0.1
	Rad	-1.083	0.573	3.565	0.059	0.3
MB.TAG=Sa	Fuß	-1.544	0.479	10.376	0.001	0.2
	ÖV	-3.167	0.426	55.376	0.000	0.0
	Rad	-1.431	0.508	7.937	0.005	0.2
MB.TAG=So	Fuß	-2.141	0.569	14.170	0.000	0.1
	ÖV	-3.250	0.500	42.298	0.000	0.0
	Rad	-1.063	0.570	3.475	0.062	0.3
HH.GROESSE=2	Fuß	0.140	0.500	0.079	0.779	1.2
	ÖV	-2.070	0.415	24.893	0.000	0.1
	Rad	-2.825	0.639	19.518	0.000	0.1
HH.GROESSE=3	Fuß	-0.725	0.557	1.690	0.194	0.5
	ÖV	-1.523	0.470	10.498	0.001	0.2
	Rad	-1.692	0.631	7.195	0.007	0.2
HH.GROESSE=4+	Fuß	-0.525	0.706	0.553	0.457	0.6
	ÖV	-1.045	0.530	3.916	0.048	0.4
	Rad	-1.937	0.737	6.902	0.009	0.1
HH.EINK.NEU (mittleres H.E.)	Fuß	-0.852	0.759	1.259	0.262	0.4
	ÖV	-0.022	0.645	0.001	0.973	1.0
	Rad	-0.726	0.973	0.556	0.456	0.5
HH.EINK.NEU (hohes H.E.)	Fuß	1.083	0.739	2.146	0.143	3.0
	ÖV	-1.246	0.780	2.552	0.110	0.3
	Rad	2.617	0.812	10.396	0.001	13.7
HSSCHIEN=1	Fuß	-1.045	0.593	3.105	0.078	0.4
	ÖV	1.864	0.575	10.521	0.001	6.5
	Rad	-1.472	0.464	10.064	0.002	0.2
S.KM.OEV.MIN	Fuß	-0.915	0.539	2.893	0.089	0.4
	ÖV	-1.859	0.487	14.575	0.000	0.2
	Rad	-1.123	0.506	4.930	0.026	0.3
ZUF.R.ANB.OEV (sehr zufrieden)	Fuß	1.308	0.348	14.127	0.000	3.7
	ÖV	0.202	0.266	0.577	0.447	1.2
	Rad	-1.489	0.352	17.856	0.000	0.2
ZUF.R.ANB.OEV (Verbesserung gewünscht)	Fuß	1.435	0.744	3.726	0.054	4.2
	ÖV	0.058	0.585	0.010	0.921	1.1
	Rad	-0.863	0.617	1.956	0.162	0.4

Effekt	Verkehrsmittel	Schätzwert $\hat{\beta}$	geschätzter Std.fehler	Wald ω	p-Wert	exp(β)
WETT.GUT=ja	Fuß	2.026	0.540	14.060	0.000	7.6
	ÖV	0.498	0.631	0.623	0.430	1.6
	Rad	3.023	0.510	35.107	0.000	20.5
S.ENTF.KM	Fuß	-1.773	0.136	170.937	0.000	0.2
	ÖV	-0.022	0.007	11.553	0.001	1.0
	Rad	-0.168	0.046	13.466	0.000	0.8
STR.TYP.NEU (innerorts)	Fuß	-0.870	0.803	1.174	0.279	0.4
	ÖV	3.415	0.543	39.570	0.000	30.4
	Rad	0.264	0.572	0.241	0.644	1.3
WEG.MIT (mit mehreren)	Fuß	-0.405	0.308	1.727	0.189	0.7
	ÖV	-0.639	0.244	6.852	0.009	0.5
	Rad	-2.306	0.340	46.106	0.000	0.1
VORHER	Fuß	-0.718	0.312	5.283	0.022	0.5
	ÖV	-0.848	0.278	9.294	0.002	0.4
	Rad	0.428	0.305	1.963	0.161	1.5
GEWOHN	Fuß	0.258	0.351	0.538	0.463	1.3
	ÖV	1.698	0.247	47.344	0.000	5.5
	Rad	-0.235	0.368	0.408	0.523	0.8
OHNE.ALT	Fuß	-0.216	0.339	0.403	0.525	0.8
	ÖV	0.495	0.283	3.068	0.080	1.6
	Rad	-0.932	0.421	4.916	0.027	0.4
GUT.ENTF	Fuß	2.998	0.319	88.532	0.000	20.1
	ÖV	1.097	0.287	14.624	0.000	3.0
	Rad	2.441	0.341	51.367	0.000	11.5
KURZ.T	Fuß	-4.580	0.612	56.032	0.000	0.0
	ÖV	-0.382	0.241	2.521	0.112	0.7
	Rad	-0.972	0.350	7.697	0.006	0.4
BEQUEM	Fuß	-2.008	0.532	14.223	0.000	0.1
	ÖV	-0.294	0.273	1.156	0.282	0.7
	Rad	-2.668	0.548	23.733	0.000	0.1
FLEXIBEL	Fuß	-0.051	0.735	0.005	0.945	1.0
	ÖV	-2.844	0.569	25.005	0.000	0.1
	Rad	0.540	0.569	0.901	0.343	1.7
KOSTEN	Fuß	1.484	0.631	5.532	0.019	4.4
	ÖV	0.704	0.453	2.417	0.120	2.0
	Rad	3.189	0.556	32.982	0.000	24.3
SPORT	Fuß	7.551	0.934	65.359	0.000	1902.4
	ÖV	0.324	1.144	0.080	0.777	1.4
	Rad	7.249	0.909	63.669	0.000	1407.4
S.DAUERMIN	Fuß	0.136	0.013	111.446	0.000	1.1
	ÖV	0.063	0.007	93.275	0.000	1.1
	Rad	0.010	0.012	0.753	0.386	1.0
EINDRUCK (durchschnittlich)	Fuß	-0.979	0.332	8.689	0.003	0.4
	ÖV	1.238	0.295	17.582	0.000	3.4
	Rad	0.039	0.327	0.014	0.904	1.0
EINDRUCK (unangenehm)	Fuß	0.474	0.715	0.440	0.507	1.6
	ÖV	0.618	0.566	1.190	0.275	1.9
	Rad	0.421	0.759	0.307	0.579	1.5
POLY.S.ALTER.N1	Fuß	-16.170	8.274	3.819	0.051	0.0
	ÖV	-25.709	6.313	16.586	0.000	0.0
	Rad	-17.893	7.818	5.239	0.022	0.0

Effekt	Verkehrsmittel	Schätzwert $\hat{\beta}$	geschätzter Std.fehler	Wald ω	p-Wert	exp(β)
POLY.S.ALTER.N2	Fuß	1.324	7.107	0.035	0.852	3.8
	ÖV	17.030	5.657	9.064	0.003	2.5e+7
	Rad	0.609	6.930	0.008	0.930	1.8
NUTZUNG (Mitnutzer)	Fuß	1.478	1.032	2.051	0.152	4.4
	ÖV	2.581	0.890	8.401	0.004	13.2
	Rad	3.026	1.106	7.485	0.006	20.6
NUTZUNG (sonstiges)	Fuß	-0.112	0.666	0.028	0.867	0.9
	ÖV	3.857	0.543	50.370	0.000	47.3
	Rad	5.448	0.838	42.312	0.000	232.3
RAD.VERF	Fuß	-0.611	0.471	1.685	0.194	0.5
	ÖV	-0.398	0.353	1.275	0.259	0.7
	Rad	3.872	1.295	8.948	0.003	48.1
PERS.EINK.NEU (mittleres P.E.)	Fuß	0.058	1.168	0.002	0.960	1.1
	ÖV	-0.621	0.928	0.448	0.504	0.5
	Rad	-10.644	1.612	43.395	0.000	0.0
PERS.EINK.NEU (hohes P.E.)	Fuß	-2.728	1.067	6.536	0.011	0.1
	ÖV	0.145	0.777	0.035	0.852	1.2
	Rad	-2.285	0.745	9.410	0.002	0.1
NETZKARTE=nein	Fuß	-0.615	0.337	3.318	0.069	0.5
	ÖV	-1.708	0.248	47.308	0.000	0.2
	Rad	1.294	0.405	10.195	0.001	3.6
BAHNCARD	Fuß	-0.125	0.936	0.018	0.894	0.9
	ÖV	-0.187	0.676	0.076	0.782	0.8
	Rad	-2.865	0.901	10.108	0.001	0.1
EINPKW.N	Fuß	1.920	0.464	17.123	0.000	6.8
	ÖV	0.698	0.336	4.312	0.038	2.0
	Rad	-0.007	0.430	0.000	0.988	1.0
NUTZUNG (Mitnutzer)*	Fuß	-2.843	1.217	5.454	0.020	0.1
HH.EINK.NEU (mittleres H.E.)	ÖV	-2.764	1.017	7.394	0.007	0.1
	Rad	-1.587	1.233	1.657	0.198	0.2
NUTZUNG (sonstiges)*	Fuß	2.839	0.956	8.813	0.003	17.1
	ÖV	-2.115	0.758	7.786	0.005	0.1
HH.EINK.NEU (mittleres H.E.)	Rad	-2.201	1.211	3.301	0.069	0.1
	Fuß	-0.574	1.129	0.258	0.611	0.6
NUTZUNG (Mitnutzer)*	ÖV	-0.942	0.937	1.011	0.315	0.4
	Rad	-4.905	1.182	17.068	0.000	0.0
HH.EINK.NEU (hohes H.E.)	Fuß	-2.006	1.046	3.681	0.055	0.1
	ÖV	-1.990	0.878	5.135	0.023	0.1
HH.EINK.NEU (hohes H.E.)	Rad	-4.924	1.121	19.304	0.000	0.0
	Fuß	-0.989	0.865	1.308	0.253	0.4
PERS.EINK.NEU (mittl. P.E.)*HH.EINK.NEU (mittleres H.E.)	ÖV	0.972	0.710	1.872	0.171	2.6
	Rad	3.848	1.143	11.326	0.001	46.9
PERS.EINK.NEU (hohes P.E.)*HH.EINK.NEU (mittleres H.E.)	Fuß	0.311	0.725	0.184	0.668	1.4
	ÖV	-0.267	0.565	0.223	0.637	0.8
PERS.EINK.NEU (mittl. P.E.)*HH.EINK.NEU (hohes H.E.)	Rad	-2.172	0.774	7.879	0.005	0.1
	Fuß	-2.810	0.914	9.443	0.002	0.1
PERS.EINK.NEU (mittl. P.E.)*HH.EINK.NEU (hohes H.E.)	ÖV	2.100	0.756	7.719	0.005	8.2
	Rad	0.296	1.127	0.069	0.793	1.3
PERS.EINK.NEU (hohes P.E.)*HH.EINK.NEU (hohes H.E.)	Fuß	-3.039	0.779	15.201	0.000	0.0
	ÖV	0.412	0.655	0.396	0.529	1.5
(hohes H.E.)	Rad	-0.113	0.657	0.029	0.864	0.9

Effekt	Verkehrsmittel	Schätzwert $\hat{\beta}$	geschätzter Std.fehler	Wald ω	p-Wert	$\exp(\beta)$
NUTZUNG(Mitn.)*	Fuß	20.077	5.215	14.820	0.000	5.2e+8
POLY.S.ALTER.N1	ÖV	0.993	8.619	0.013	0.908	2.7
	Rad	-182.123	4.243	1842.796	0.000	0.0
NUTZUNG(sonst.)*	Fuß	59.001	4.310	187.421	0.000	4.2e+25
POLY.S.ALTER.N1	ÖV	59.665	6.245	91.284	0.000	8.2e+25
	Rad	91.615	5.334	294.989	0.000	6.1e+39
NUTZUNG(Mitn.)*	Fuß	4.061	6.030	0.453	0.501	58.0
POLY.S.ALTER.N2	ÖV	22.320	7.630	8.556	0.003	4.9e+9
	Rad	-134.303	5.837	529.325	0.000	0.0
NUTZUNG(sonst.)*	Fuß	-1.470	5.374	0.075	0.784	0.2
POLY.S.ALTER.N2	ÖV	-31.528	6.516	23.415	0.000	0.0
	Rad	-32.888	5.335	37.997	0.000	0.0
PERS.EINK.NEU	Fuß	-0.680	1.138	0.357	0.550	0.5
(mittleres P.E.)*	ÖV	-0.802	0.814	0.971	0.324	0.4
BAHNCARD	Rad	7.564	1.422	28.277	0.000	1926.9
PERS.EINK.NEU	Fuß	2.470	1.615	2.338	0.126	11.8
(hohes P.E.)*	ÖV	0.733	1.212	0.366	0.545	2.1
BAHNCARD	Rad	1.509	1.147	1.731	0.188	4.5

Tabelle 12: Parameterschätzungen für das Modell mit Interaktionen

Die Tabelle des Likelihood-Quotienten-Tests zeigt sehr deutlich, dass alle Interaktionen signifikant sind. Dazu genügt es die p-Werte zu betrachten. Zu berücksichtigen ist jedoch, dass nicht alle Responsekategorien gleichermaßen signifikant sind. Bei Betrachtung des p-Wertes, zum Beispiel für die Variable HH.GROESSE für die Responsekategorie zu Fuß gegenüber IV, zeigt sich, dass kein Einfluss besteht. Hierbei sei nochmals auf das Einführen von kategorienspezifischen Variablen (vgl. Abschnitt 3.4) hingewiesen.

5.3 Interpretation der Parameter

Haupteffekte

Zunächst werden die Haupteffekte, die nach erhebungsbezogenen, haushaltsbezogenen, wegebezogenen und personenbezogenen Merkmalen gegliedert sind, interpretiert. In Abschnitt 3.3, Seite 25 ff. finden sich bereits eine Zusammenfassung und Beispiele über die Interpretation geschätzten Parameter. Der Einfachheit halber wird die Vorgehensweise hier noch einmal kurz skizziert. Im Fall eines kategorialen Merkmals K mit den möglichen Ausprägungen $1, \dots, I$ sind $I - 1$ Dummy-Variablen $x_{K(1)}, \dots, x_{K(I-1)}$ notwendig. Mit $\beta_{K(i),r}$ wird der Parameter der Dummy-Variable $x_{K(i)}$ beschrieben und im Folgenden nur noch mit $K = i$ bezeichnet. r steht für die Kategorien Fuß, IV, ÖV und Rad. Falls $K = I$ liegt die kategoriale Variable in der Referenzkategorie vor. Im Fall positiver Parameterschätzungen wird das Chancenverhältnis (*relative Chance, odds-ratio*)

$$\frac{P(Y = r | K = i, \mathbf{x})/P(Y = r_0 | K = i, \mathbf{x})}{P(Y = r | K = I, \mathbf{x})/P(Y = r_0 | K = I, \mathbf{x})} = e^{\beta_{K(i), r}},$$

im Fall negativer Schätzungen in der Regel das umgekehrte Chancenverhältnis

$$\frac{P(Y = r_0 | K = i, \mathbf{x})/P(Y = r | K = i, \mathbf{x})}{P(Y = r_0 | K = I, \mathbf{x})/P(Y = r | K = I, \mathbf{x})} = e^{-\beta_{K(i), r}},$$

für r gegenüber $r_0 = IV$ zwischen $K = i$ und $K = I$ betrachtet. Mit \mathbf{x} werden alle übrigen Einflussgrößen erfasst. Da keine Interaktionen betrachtet werden, ist das Chancenverhältnis unabhängig von \mathbf{x} . Der Einfachheit halber wurden in Tabelle 12 diese odds-ratios als letzte Spalte angefügt, auf die im Folgenden immer wieder zurückgegriffen wird. Im Fall positiver Parameterschätzungen kann das Chancenverhältnis folglich direkt aus der Tabelle abgelesen werden. Im Fall negativer Parameterschätzung wird der reziproke Wert ($1/e^{\beta_{K(i), r}}$) verwendet.

Zur Erklärung der Effekte ist es häufig sinnvoll die Struktur der Daten (Abschnitt 5.1.1) noch einmal näher zu betrachten.

Erhebungsebene

Mobilitätstag: Die Schätzungen für die erhebungsbezogene Variable Mobilitätstag (MB.TAG) mit $I = \text{Montag}$ als Referenzkategorie sind negativ. Daher wird das umgekehrte Chancenverhältnis, d.h.

$$\frac{P(Y = IV | \text{MB.TAG} = i, \mathbf{x})/P(Y = r | \text{MB.TAG} = i, \mathbf{x})}{P(Y = IV | \text{MB.TAG} = \text{Montag}, \mathbf{x})/P(Y = r | \text{MB.TAG} = \text{Montag}, \mathbf{x})} = e^{-\beta_{\text{MB.TAG}=i, r}}$$

betrachtet, wobei r die Verkehrsmittel Fuß, ÖV oder Rad und i die Wochentage Dienstag bis Sonntag darstellen. Die Negativität der Schätzer besagt, dass im Vergleich zu Montag an den Wochentagen Dienstag bis Sonntag die Verkehrsmittel Fuß, ÖV bzw. Rad nicht so häufig wie IV gewählt werden. Die umgekehrten odds-ratios lauten:

Mobilitätstag	Fuß	ÖV	Rad
Di	11.97	6.89	1.89
Mi	3.61	6.03	2.13
Do	3.41	6.59	3.22
Fr	3.58	8.96	2.95
Sa	4.69	23.73	4.18
So	8.51	25.78	2.89

Tabelle 13: Odds-ratios der Mobilitätstage

Einfacher ist es diese mit Hilfe einer Grafik zu erläutern. Die Chancen Verkehrsmit-

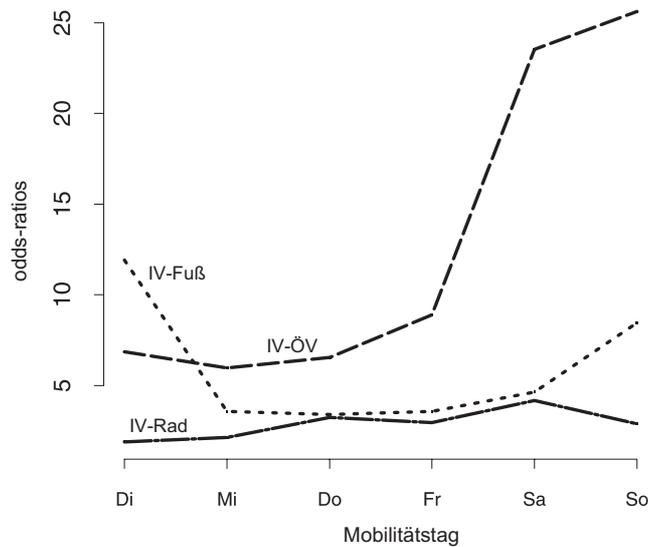


Abbildung 14: Geschätzte Chancenverhältnisse in Abhängigkeit vom Mobilitätstag

tel IV gegenüber Verkehrsmittel Fuß zu wählen betragen am Dienstag fast das 12-fache der Chancen am Montag. Der Wert $e^{-\beta_{\text{MB.TAG=Mittwoch, ÖV}}} = e^{1.797} = 6.03$ besagt, dass am Mittwoch die Chance IV gegenüber ÖV etwa das 6-fache der Chance am Montag beträgt. Das Chancenverhältnis von IV und Fuß ist an den Wochentagen Mittwoch bis Freitag fast identisch. Bei IV und ÖV liegen die odds-ratios an den Wochentagen Dienstag bis Freitag eng beieinander. D.h. an diesen Tagen betragen die Chancen IV gegenüber ÖV das 6 bis 9-fache der Chancen am Montag. Am Wochenende wird ÖV seltener genutzt, IV dafür umso stärker, was an der Grafik durch einen deutlichen Anstieg der odds-ratios zu erkennen ist. Daran kann ein Zusammenhang mit dem häufigen Nutzungszweck für Arbeits- und Ausbildungswege an Werktagen vermutet werden. Die p-Werte zwischen IV und Rad besagen, dass an den Tagen Dienstag, Mittwoch, Freitag und Sonntag kaum Signifikanz vorliegt. Die Chance zwischen IV und Rad beträgt am Samstag etwas mehr als das 4-fache der Chance am Montag.

Haushaltsebene

Für die haushaltsbezogenen Variablen Haushaltsgröße (HH.GROESSE), Haushaltseinkommen (HH.EINK.NEU), Anbindung an den schienengebundenen Nahverkehr (HSSCHIEN), Entfernung zur nächsten Haltestelle [min] (S.KM.OEV.MIN) und Zufriedenheit mit der Anbindung an ÖV (ZUFR.ANB.OEV) kann wie folgt argumentiert werden.

Haushaltsgröße: Zwischen der Verkehrsmittelwahl IV und Fuß ist die Haushaltsgröße nicht relevant. Die umgekehrten relativen Chancen IV gegenüber ÖV bzw. Rad zu wählen lauten:

Haushaltsgröße	ÖV	Rad
2	7.92	16.85
3	4.59	5.43
4+	2.85	6.94

Tabelle 14: Odds-ratios der Haushaltsgröße

Damit ergibt sich folgender Plot:

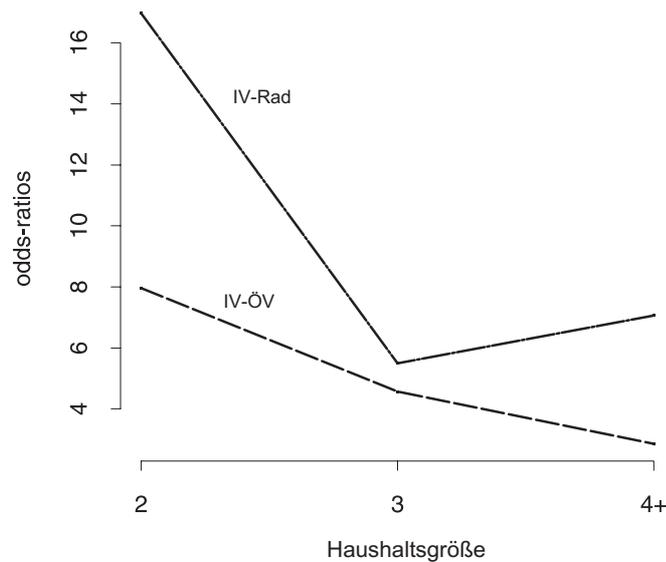


Abbildung 15: Geschätzte Chancenverhältnisse in Abhängigkeit von der Haushaltsgröße

Die Analyse der Haushaltsgröße zeigt, dass die Chance ÖV gegenüber IV bei einem Mehrpersonenhaushalt geringer als bei einem Singlehaushalt ist. Im Einzelnen heisst das, dass die Chancen Verkehrsmittel IV gegenüber ÖV zu wählen bei einem 2-Personenhaushalt in etwa das 8-fache, bei einem 3- bzw. 4- und Mehrpersonenhaushalt das 4.5- bzw. 3-fache der Chance eines Singlehaushaltes beträgt. Daran erkennt man, dass die Chance ÖV gegenüber IV zu benutzen ab einem 3-Personenhaushalt im Vergleich zu einem 2-Personenhaushalt wieder leicht steigt. In der Grafik zeigt sich dieser Zusammenhang durch das Fallen des Verlaufs IV-ÖV. Die Chance Rad gegenüber IV zu benutzen ist bei den 2-Personenhaushalten am geringsten. An der Grafik erkennt man dies daran, dass die Chance eines 2-Personenhaushaltes IV gegenüber Rad zu wählen fast das 17-fache der Chance eines Singlehaushaltes beträgt. Es zeigt sich, dass Wege per Rad bzw. ÖV am häufigsten von Singles zurückgelegt werden.

Haushaltseinkommen: Das Haushaltseinkommen wird erst innerhalb der Interaktionen interpretiert.

Anbindung an den schienengebundenen Nahverkehr und Entfernung zur nächstgelegenen Haltestelle: Um erkennen zu können inwieweit sich die Verfügbarkeit auf die Verkehrsmittelwahl auswirkt wird die Anbindung an den schienengebundenen Nahverkehr

bzw. die Entfernung zur nächstgelegenen Haltestelle analysiert. An den p-Werten erkennt man, dass die Anbindung und die Entfernung auf die Wahl von Fuß gegenüber IV kaum einen Einfluss besitzen. Falls eine Anbindung an den schienengebundenen Nahverkehr vorhanden ist, beträgt die Chance ÖV gegenüber IV zu wählen das 6.5-fache der Chance ohne Anbindung. Ein Grund hierfür ist, dass die Haushaltsmitglieder in München über eine gute Verkehrsinfrastruktur verfügen. Hingegen beträgt die Chance Rad gegenüber IV zu wählen nur in etwa das 0.2-fache der Chance ohne Anbindung an den schienengebundenen Nahverkehr.

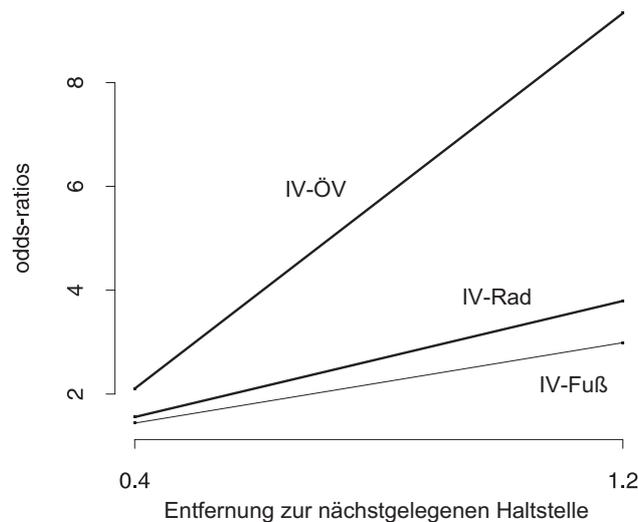


Abbildung 16: Geschätzte Chancenverhältnisse in Abhängigkeit von der Entfernung zur nächstgelegenen Haltestelle

Mit zunehmender Entfernung zur nächsten Haltestelle nimmt die Chance IV gegenüber ÖV oder Rad zu wählen zu. Abbildung 16 verdeutlicht diesen Zusammenhang durch das Steigen des Verlaufs bei zunehmender Entfernung.

Zufriedenheit mit der Anbindung: Die Chance Fuß gegenüber IV zu wählen bei allgemeiner Zufriedenheit der Anbindung steigt um das 3.7-fache, falls die Anbindung sehr zufriedenstellend ist. Es kann kein Einfluss festgestellt werden, dass Haushalte die mit der Anbindung zufrieden sind verstärkt mit öffentlichen Verkehrsmitteln unterwegs sind.

Wegeebene

Die Wegeebene umfasst die wegebezogenen Variablen gutes Wetter (WETT.GUT), Entfernung (S.ENTF.KM), Streckentyp (STR.TYP.NEU), alleine oder mit mehreren Personen unterwegs (WEG.MIT), vorher mit diesem VM unterwegs (VORHER), aus Gewohnheit damit unterwegs (GEWOHN), keine andere Möglichkeit (OHNE.ALT), passt zur Entfernung (GUT.ENTF), geringster Zeitaufwand (KURZ.T), bequem (BEQUEM), flexibel (FLEXIBEL), kostengünstig (KOSTEN), Sport/Bewegung (SPORT), Dauer der Fahrt/des Weges (S.DAUERMIN) und der Eindruck der Fahrt/des Weges (EINDRUCK).

gutes Wetter: Die Chance bei gutem Wetter Fuß oder Rad gegenüber IV zu wählen steigt um das 7.6 bzw. 20.5-fache der Chance bei schlechtem Wetter. Es wird kein Einfluss auf das Verkehrsmittelwahlverhalten öffentlicher Verkehrsmittel gegenüber IV festgestellt. Ein Grund dafür kann sein, dass die Befragungen zu Zeitpunkten (März, Juni und Oktober) stattgefunden haben, an denen kaum außergewöhnliche Witterungsbedingungen (Schnee, Eis, usw.) vorlagen.

Entfernung:

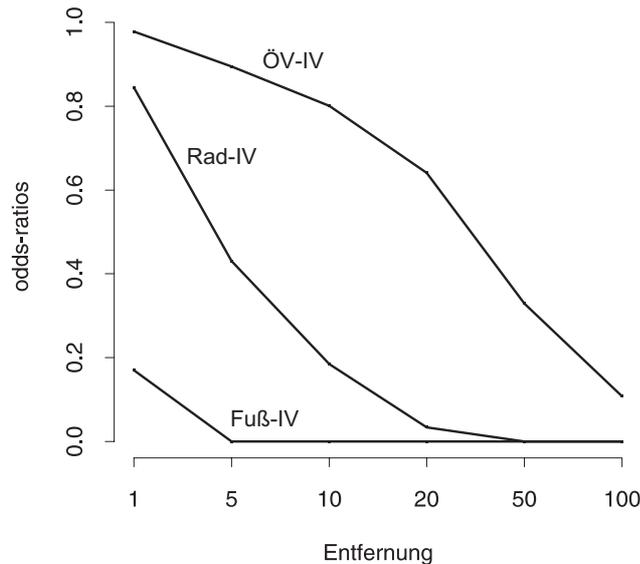


Abbildung 17: Geschätzte Chancenverhältnisse in Abhängigkeit von der Entfernung [km]

Die Entfernung ist eine sehr wichtige Einflussgröße zur Erklärung der Verkehrsmittelwahl. Erwartungsgemäß nimmt die Chance Fuß gegenüber IV zu wählen mit zunehmender Entfernung ab. Daher wird hier nicht das umgekehrte Chancenverhältnis betrachtet. Abbildung 17 zeigt, dass die Chance ÖV gegenüber IV zu wählen ebenfalls mit zunehmender Entfernung abnimmt, aber relativ langsam. Die Chance, dass bei einer Weglänge von einem Kilometer IV gegenüber Rad gewählt wird ist fast identisch, bei einer Weglänge von 10 km etwa 5.4 mal so hoch.

Streckentyp: Die Chance ÖV gegenüber IV zu wählen steigt um das 30-fache, wenn der Weg innerorts durchgeführt wird. Hieran erkennt man, dass die öffentliche Verkehrsinfrastruktur innerhalb Münchens sehr gut ist. Bei den Verkehrsmitteln Fuß und Rad wird kein Einfluss auf das Verkehrsmittelwahlverhalten festgestellt.

Weg alleine oder mit mehreren Personen zurückgelegt: Die Chance IV gegenüber ÖV zu wählen, falls der Weg mit mehreren Personen zurückgelegt wird beträgt fast das doppelte der Chance den Weg alleine zurückzulegen. Die Chance IV gegenüber Rad zu wählen, falls der Weg mit Begleitung zurückgelegt wird beträgt das 10-fache der Chance die Strecke alleine zu fahren. Für die Wege zu Fuß wird kein Einfluss auf das Verkehrsmittelwahlverhalten festgestellt.

Motive der Verkehrsmittelwahl: Häufig fällt die Verkehrsmittelwahl auf den Bereich, mit dem man bereits unterwegs war. Die Chance IV gegenüber Fuß oder ÖV zu wählen steigt in etwa auf das doppelte der Chance nicht vorher mit diesen Verkehrsmitteln unterwegs gewesen zu sein. Die Entscheidung für das öffentliche Verkehrsmittel ist sehr häufig durch Gewohnheit geprägt, d.h. die Chance ÖV gegenüber IV zu wählen steigt um das 5.5-fache, wenn das Verkehrsmittel gewohnheitsbedingt genutzt wird. Der Verkehrsbe- reich Fuß wird häufig gewählt, wenn die Entfernung angemessen ist. Die Chancen eines der Verkehrsbereiche Fuß, ÖV oder Rad gegenüber IV zu wählen steigen bei passender Entfernung auf das 20-, 3- bzw. 11.5-fache an. Die Chance Fuß gegenüber IV zu wählen ist nahezu 0, wenn kurze Zeiten als Grund angegeben werden, die Chance Rad zu wählen beträgt in diesem Fall gerade 0.4. Falls Bequemlichkeit als Grund angegeben wird, steigt die Chance Verkehrsmittel IV gegenüber Fuß bzw. Rad zu wählen auf das 7.5- bzw. 14.4-fache. Das Verkehrsmittel IV wird häufig auch gewählt, wenn keine anderen Alternativen existieren. Die Chance Rad gegenüber IV zu wählen beträgt hier ebenfalls nur 0.4. Die geringe Flexibilität des ÖV gegenüber dem IV sind als Defizite einzustufen. Die Chance der Verkehrsmittelwahl IV gegenüber ÖV steigt um das 17.2-fache, wenn Flexibilität als Grund der Verkehrsmittelwahl angegeben wird. Günstigere Kosten werden am häufigsten für die Wahl des Fahrrades genannt. Die Chance Rad oder Fuß gegenüber IV zu wählen beträgt das 24.3- bzw. 4.4-fache der Chance ohne Berücksichtigung von Kosten. Abbil- dung 10 zeigt, dass insbesondere Wege zu Fuß bzw. mit dem Fahrrad der Bewegung bzw. der sportlichen Aktivität dienen. In diesem Fall ist die Chance IV gegenüber Fuß bzw. Rad zu wählen fast 0.

Dauer der Fahrt/des Weges: Die folgende Abbildung zeigt die umgekehrten odds-ratios.

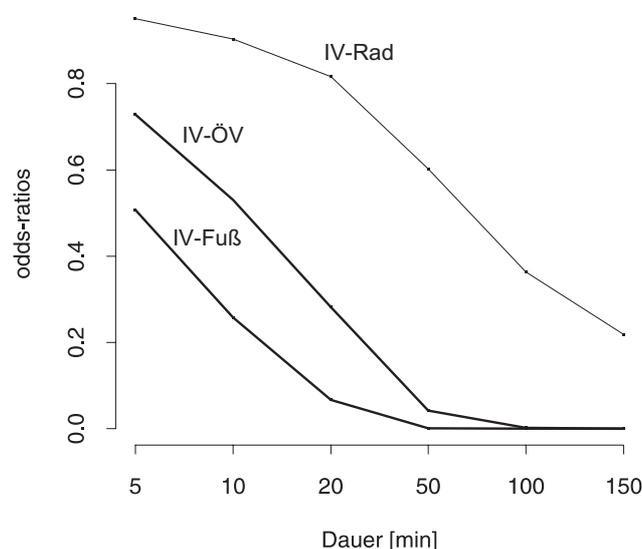


Abbildung 18: Geschätzte Chancenverhältnisse in Abhängigkeit von der Wegedauer [min]

Der positive Parameterschätzer $\beta_{S,ENTF.KM,\ddot{O}V}$ besagt, dass mit zunehmender Wegedauer die Chance $\ddot{O}V$ gegenüber IV zu wählen zunimmt. Lange Wege die zu Fuß zurückgelegt werden umfassen in der Regel nicht mehr als 5 km. Der positive Parameterschätzer für die Wahl zwischen Fuß und IV ist folglich darauf zurückzuführen, dass lange Wegezeiten zu Fuß in der Regel nur kurze Entfernungen betreffen. Diese werden folglich nicht mit dem Auto, sondern zu Fuß oder anderen Verkehrsmitteln (Fahrrad) zurückgelegt. Die Chance Verkehrsbereich zu Fuß gegenüber IV zu wählen nimmt somit bei zunehmender Wegedauer zu. Die Dauer des Weges hat bei kurzen Strecken nur geringen Einfluss auf die Wahl zwischen Rad und IV.

Eindruck der Fahrt/des Weges: Die Chance bei durchschnittlich entstandenen Ein-

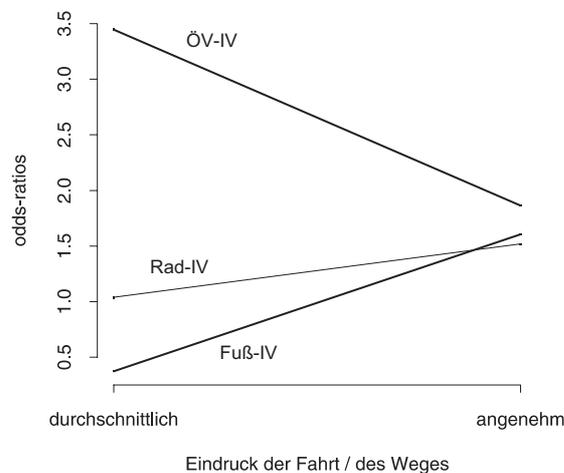


Abbildung 19: Geschätzte Chancenverhältnisse in Abhängigkeit vom Eindruck der Fahrt / des Weges

drücken während der Fahrt/des Weges IV gegenüber Fuß zu wählen beträgt das 2.7-fache der Chance bei angenehmen Eindruck. Die Chance bei durchschnittlicher Empfindung $\ddot{O}V$ gegenüber IV zu wählen steigt auf das 3.4-fache der Chance bei angenehmer Empfindung. Für die Wahl zwischen Rad und IV existiert kein Einfluss.

Personenebene

Die Personenebene umfasst die Variablen Alter (S.ALTER.N), Nutzung (NUTZUNG), Rad-Verfügung (RAD.VERF), persönliches Einkommen (PERS.EINK.NEU), Besitz einer Netzkarte (NETZKARTE), Besitz einer Bahncard (BAHNCARD) und die Einstellung zum Individualverkehr (EINPKW.N). Die Variablen Alter, Nutzung, persönliches Einkommen und Bahncard werden innerhalb der Interaktionen interpretiert.

Rad-Verfügung: Die Chance Rad gegenüber IV zu wählen bei Besitz eines Fahrrades beträgt in etwa das 50-fache der Chance ohne den Besitz eines Fahrrades. Die Ursache liegt darin, dass eine Person, die kein Fahrrad besitzt, dieses Verkehrsmittel folglich auch

nicht wählen kann. In München können seit einiger Zeit Fahrräder gemietet werden. Zum Zeitpunkt der Befragung gab es diese Möglichkeit allerdings noch nicht. Für die Verkehrsbereiche Fuß bzw. ÖV und IV wird kein Einfluss festgestellt.

Besitz einer Netzkarte: Ohne den Besitz einer Netzkarte beträgt die Chance IV gegenüber ÖV zu benutzen in etwa das 5.5-fache der Chance mit Netzkarte. Die Chance das Fahrrad gegenüber IV zu wählen beträgt das 3.6-fache der Chance mit Netzkarte. Der Besitz einer Netzkarte hat keine Einfluss auf die Wahl zwischen Fuß und IV.

Einstellung zu Individualverkehr: Auch die Einstellung zu bestimmten Verkehrsmitteln kann ein Kriterium für deren Wahl darstellen. Ist die Einstellung zum Individualverkehr negativ, so ist die Chance Fuß oder ÖV gegenüber IV zu wählen fast 7 bzw. 2 mal so hoch wie die Chance bei positiver Einstellung. Für die Entscheidung zwischen Rad und IV ist in diesem Kontext kein Einfluss erkennbar.

Interaktionseffekte

Im Folgenden werden die Interaktionseffekte interpretiert. Dazu werden folgende Annahmen getroffen:

Annahme 1: Es wird angenommen, dass mit Ausnahme der Interaktionsterme für alle im Modell enthaltenen *kategorialen Variablen* die Referenzkategorie gewählt wird ($K = I$).

Aus Abschnitt 3.3 ist bekannt, dass die I -te Kategorie (Referenzkategorie) implizit durch $x_{K(1)} = \dots = x_{K(I-1)} = 0$ erfasst wird.

Annahme 2: Für die *metrischen Größen*, die nicht im Interaktionsterm enthalten sind, wird jeweils der Mittelwert angenommen. Für Alter (A) ergibt sich ein Mittelwert von 43 Jahren, die mittlere Entfernung (E) beträgt 12 km, die Wegedauer (D) beträgt im Durchschnitt 25 Minuten und die Entfernung zur nächstgelegenen Haltestelle (DH) beträgt im Durchschnitt 0.5 km.

Annahme 3: In \mathbf{x} werden alle, nicht im Interaktionsterm enthaltenen Variablen erfasst.

Nutzung und Haushaltseinkommen: Mit Hilfe der explorativen Analyse wurde eine Interaktion zwischen Haushaltseinkommen und Nutzung festgestellt. Die p-Werte der Interaktion zwischen Nutzung (N) und Haushaltseinkommen (H.E.) legen nahe, dass die Interaktion nicht zu vernachlässigen ist. Bereits aus den Haupteffekten $\beta_{N=Pkw-Mitnutzer, r}$ und $\beta_{N=sonstiges, r}$ ist ersichtlich, dass die Wahl für IV gegenüber $r = \text{ÖV}$ oder $r = \text{Rad}$ für Pkw-Hauptnutzer zunimmt. Für $r = \text{Fuß}$ ist kein Einfluss erkennbar. Um den Einfluss der Kovariablen Haushaltsgröße zu verdeutlichen, ist es hilfreich das Modell für alle drei

Nutzungskategorien separat anzusehen. Zunächst wird grafisch festgehalten, dass Interaktionen vorliegen (Abbildung 20). Dazu werden die odds betrachtet. Mit den zu Beginn getroffenen Annahmen für alle kategorialen und metrischen Variablen, gilt

$$\frac{P(Y = r \mid x_{\text{Nutzung}}, x_{\text{Haushaltseinkommen}}, \mathbf{x})}{P(Y = IV \mid x_{\text{Nutzung}}, x_{\text{Haushaltseinkommen}}, \mathbf{x})} = e^{\beta_{0r} + \beta_{N,r} + \beta_{H.E.,r} + \beta_{N*H.E.,r}} e^{\beta_{A,r}*43 + \beta_{E,r}*12 + \beta_{D,r}*25 + \beta_{DH,r}*0.5},$$

mit $r = \text{Fuß, ÖV und Rad}$. Die Referenzkategorie der Variable Nutzung (N) ist Pkw-Hauptnutzer, die der Variable Haushaltseinkommen (H.E.) ist niedriges Haushaltseinkommen. Für die odds ergeben sich

	Fuß	ÖV	Rad
niedriges HH-Einkommen, Pkw-Hauptnutzer	0	0.04	0.12
mittleres HH-Einkommen, Pkw-Hauptnutzer	0	0.04	0.06
hohes HH-Einkommen, Pkw-Hauptnutzer	0	0.01	1.68
niedriges HH-Einkommen, Pkw-Mitnutzer	0	0.56	2.53
mittleres HH-Einkommen, Pkw-Mitnutzer	0	0.03	0.25
hohes HH-Einkommen, Pkw-Mitnutzer	0	0.06	0.26
niedriges HH-Einkommen, sonstiges	0	2.00	28.46
mittleres HH-Einkommen, sonstiges	0	0.24	1.53
hohes HH-Einkommen, sonstiges	0	0.08	2.84

Tabelle 15: Geschätzte odds in Abhängigkeit von der Nutzung und vom HH-Einkommen
Grafisch lassen sich die Chancen wie folgt darstellen:

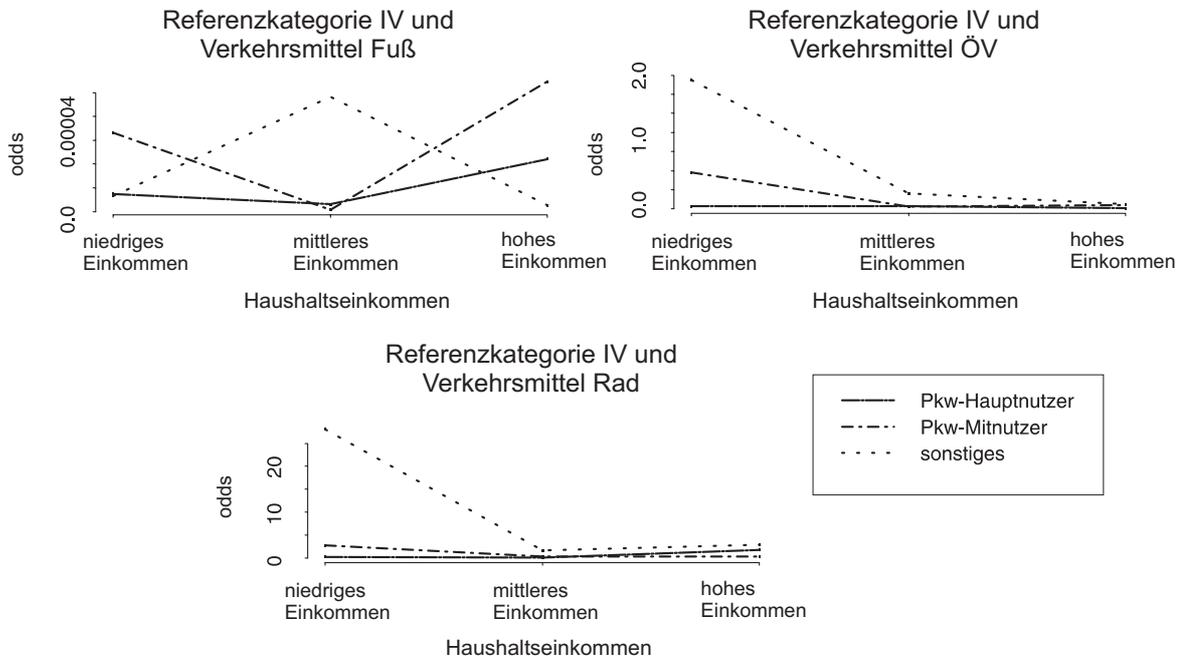


Abbildung 20: Geschätzte odds in Abhängigkeit von Nutzung und Haushaltseinkommen
Vergleicht man beispielsweise in der ersten Grafik den Verlauf für Pkw-Hauptnutzer mit den für Pkw-Mitnutzer oder sonstiges so zeigt sich, dass hier Nichtparallelitäten vorhanden sind und folglich von Interaktionen auszugehen ist. Für die beiden Grafen ÖV und

Rad gegenüber IV ist es zudem hilfreich die odds zu vergleichen um Nichtparallelitäten aufzudecken. Betrachtet man beispielsweise in der dritten Grafik den Verlauf für Pkw-Mitnutzer so erkennt man an den odds, dass beim Übergang zu mittleren Einkommen der Verlauf stark abfällt, während er beim Übergang zu hohen Einkommen fast konstant bleibt. Der Verlauf für Pkw-Hauptnutzer hingegen fällt zunächst nur schwach, während er beim Übergang zu hohen Einkommen stärker steigt. Die erste Grafik von Abbildung 20 zeigt, dass die Chance zu Fuß zu gehen nahezu Null ist. Die Ursache liegt darin, dass als durchschnittliche Entfernung der Wert 12 km gewählt wurde. Wählt man, bei ansonsten unveränderten gegebenen Einflussgrößen, eine durchschnittliche Entfernung von 2 km ändert sich nur die y-Achsen Skalierung. Die Struktur der Verläufe bleibt gleich. Bei Betrachtung der odds ist zu berücksichtigen, dass alle übrigen im Modell enthaltenen kategorialen Variablen ebenfalls über den konstanten Term ($e^{\beta_{0r}}$) in die Berechnung eingehen. Daher ist es schwierig die Chancen zu interpretieren. Es müssen alle Variablen betrachtet werden. Besser ist es das Chancenverhältnis zu betrachten.

Um das Chancenverhältnis zu beschreiben werden die odds-ratios benötigt. Verglichen wird das Verhältnis der Pkw-Hauptnutzer (Referenzkategorie I_N) zum Pkw-Mitnutzer bzw. zu den Nutzern der Kategorie sonstiges für alle Haushaltseinkommensgruppen (HH-Einkommen), d.h.

$$\begin{aligned} & \frac{P(Y = r \mid x_{\text{Nutzung}=i_N}, x_{\text{HH-Einkommen}=i_H}, \mathbf{x}) / P(Y = IV \mid x_{\text{Nutzung}=i_N}, x_{\text{HH-Einkommen}=i_H}, \mathbf{x})}{P(Y = r \mid x_{\text{Nutzung}=I_N}, x_{\text{HH-Einkommen}=i_H}, \mathbf{x}) / P(Y = IV \mid x_{\text{Nutzung}=I_N}, x_{\text{HH-Einkommen}=i_H}, \mathbf{x})} \\ &= \frac{\exp(\beta_{0r} + \beta_{\text{Nutzung}=i_N, r} + \beta_{\text{HH-Einkommen}=i_H, r} + \beta_{\text{Nutzung}=i_N * \text{HH-Einkommen}=i_H, r} + \mathbf{x}^T \boldsymbol{\beta}_r)}{\exp(\beta_{0r} + \beta_{\text{Nutzung}=I_N, r} + \beta_{\text{HH-Einkommen}=i_H, r} + \beta_{\text{Nutzung}=I_N * \text{HH-Einkommen}=i_H, r} + \mathbf{x}^T \boldsymbol{\beta}_r)} \\ &= \exp(\beta_{\text{Nutzung}=i_N, r} + \beta_{\text{Nutzung}=i_N * \text{HH-Einkommen}=i_H, r}), \end{aligned}$$

wobei r die Verkehrsbereiche Fuß, ÖV bzw. Rad, i_N die Kategorien Pkw-Mitnutzer bzw. sonstiges und i_H die Kategorien niedriges, mittleres bzw. hohes Haushaltseinkommen beschreiben. Für die odds-ratios ergeben sich folglich:

		Haushaltseinkommen		
		niedriges	mittleres	hohes
Fuß-IV	Hauptnutzer - Mitnutzer	4.38	0.26	2.47
	Hauptnutzer - sonstiges	0.89	15.30	0.12
ÖV-IV	Hauptnutzer - Mitnutzer	13.21	0.83	5.15
	Hauptnutzer - sonstiges	47.30	5.70	6.47
Rad-IV	Hauptnutzer - Mitnutzer	20.62	4.22	0.15
	Hauptnutzer - sonstiges	232.29	25.72	1.69

Tabelle 16: Geschätzte odds-ratios in Abhängigkeit von der Nutzung und vom HH-Einkommen

Damit ergeben sich folgende Plots:

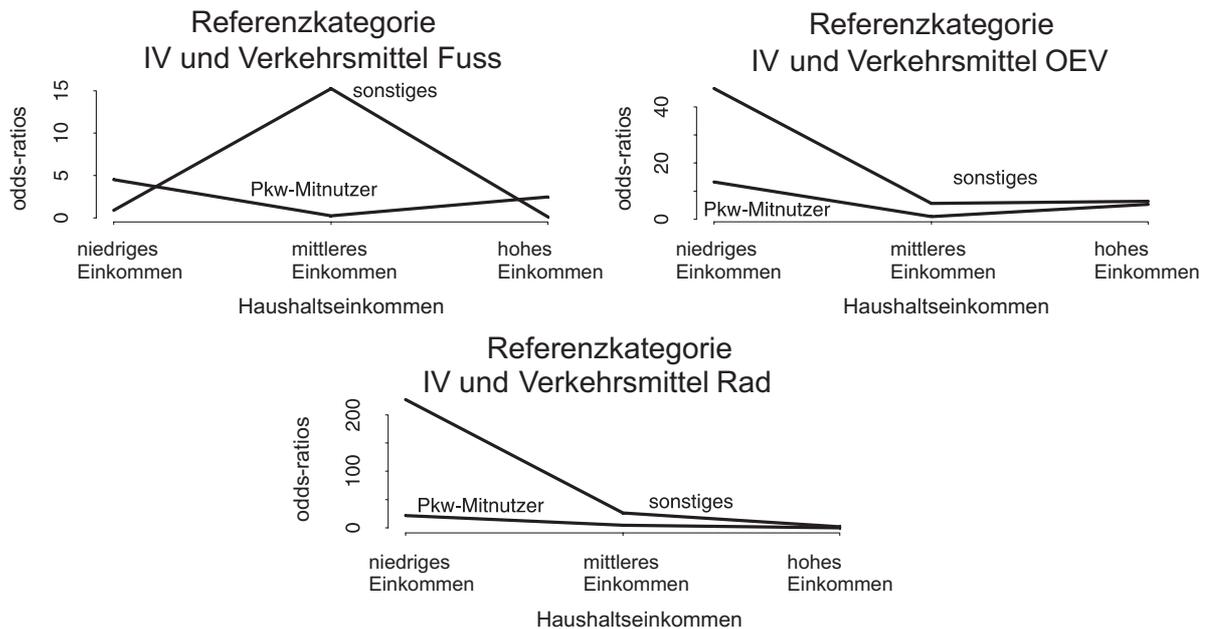


Abbildung 21: Geschätztes Chancenverhältnis in Abhängigkeit von Nutzung und HH-Einkommen

Die erste Grafik gibt an, dass die Chance eines Pkw-Mitnutzers den Verkehrsbereich Fuß gegenüber IV zu wählen bei niedrigem Haushaltseinkommen das 4.4-fache der Chance eines Pkw-Hauptnutzers beträgt. Bei mittlerem Haushaltseinkommen ist die Chance IV zu wählen fast 4 mal so hoch der Chance zu Fuß zu gehen, d.h. Pkw-Hauptnutzer nutzen bei höherem Haushaltseinkommen verstärkt IV. Bei Pkw-Mitnutzern die über ein hohes Haushaltseinkommen verfügen steigt die Chance Fuß gegenüber zu wählen um etwa das 2.5-fache. Hier ist jedoch zu berücksichtigen, dass der Einfluss hier nicht mehr signifikant ist (p-Wert: 0.611). Personen die der Kategorie sonstiges zugeordnet sind, können den Verkehrsbereich IV nur als Mitfahrer benutzen, da sie entweder keinen Führerschein oder kein entsprechendes Fahrzeug zur Verfügung haben. Daher ist die Chance eines Nutzers in Kategorie sonstiges den Verkehrsbereich Fuß gegenüber IV zu wählen mehr als das 15-fache höher der Chance eines Pkw-Hauptnutzers. Die Chance eines Pkw-Mitnutzers bei niedrigem Haushaltseinkommen öffentliche Verkehrsmittel anstatt IV zu benutzen ist mehr als das 13-fache der Chance eines Pkw-Hauptnutzers. Bei mittlerem Haushaltseinkommen dagegen steigt die Chance statt ÖV den Verkehrsbereich IV zu wählen um 1.2, d.h. auch hier ist mit steigendem Haushaltseinkommen die Chance IV zu nutzen für Pkw-Hauptnutzer größer. Für hohes Haushaltseinkommen wird kein Einfluss der Interaktion festgestellt. Für Nutzer der Kategorie sonstiges steigt die Chance Verkehrsbereich ÖV zu wählen erwartungsgemäß. Die dritte Grafik (Wahl zwischen Rad und IV) zeigt deutlich, dass die Chance IV zu wählen für Pkw-Hauptnutzer mit zunehmendem Haushaltseinkommen gegenüber Pkw-Mitnutzer bzw. sonstiges steigt. Alle drei Grafen zeigen folglich, dass

mit zunehmendem Einkommen die Chance IV zu wählen für Pkw-Hauptnutzer größer als für Pkw-Mitnutzer ist.

Bahncard, persönliches und Haushaltseinkommen: Die p-Werte lassen vermuten, dass Interaktionen zwischen Bahncard (BC) und persönlichem Einkommen sowie zwischen persönlichem und Haushaltseinkommen (H.E.) bestehen. An den Haupteffekten lässt sich ein Einfluss des persönlichen Einkommens auf die Verkehrsmittelwahl feststellen. Im weiteren wird grafisch bestätigt (Abbildung 22), dass Interaktionen vorliegen. Unter den getroffenen Annahmen 1, 2 und 3 ergeben sich aus den odds für die Interaktionen zwischen Bahncard, persönlichem und Haushaltseinkommen folgende Plots:

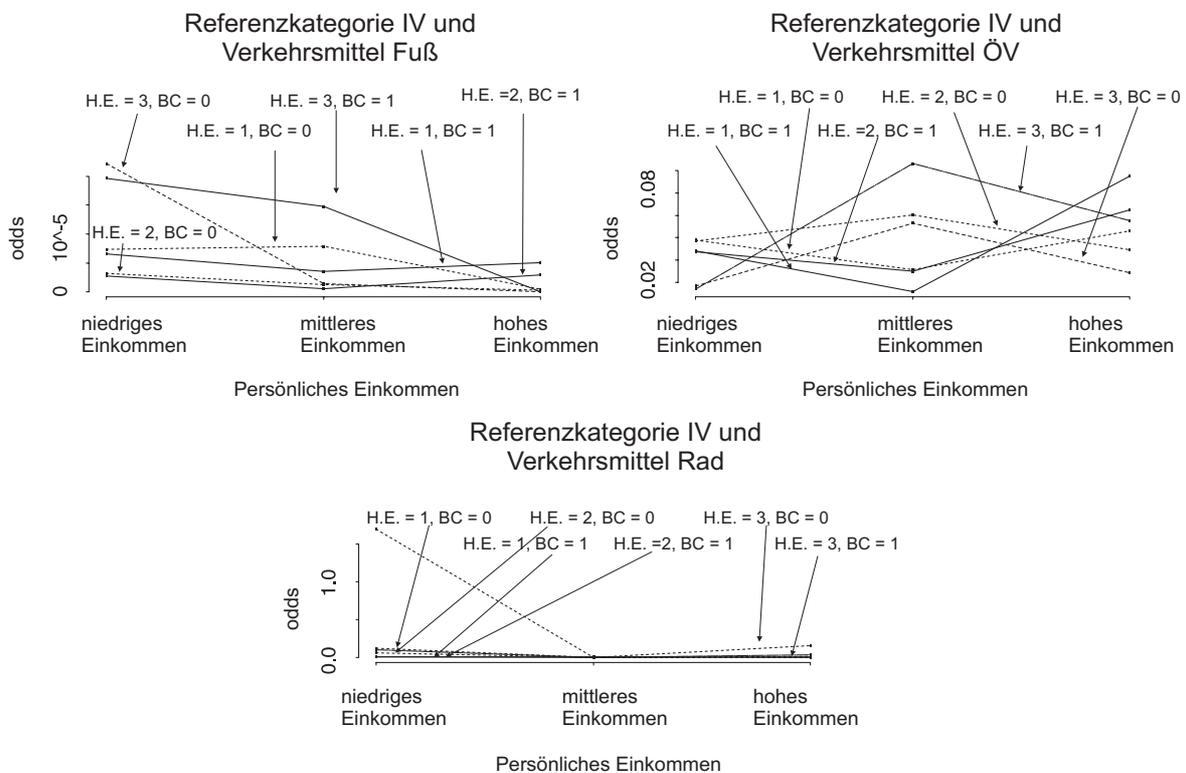


Abbildung 22: Geschätzte odds in Abhängigkeit von Haushaltseinkommen, persönlichem Einkommen und Bahncard-Besitz

Die Verläufe der ersten beiden Plots weisen eindeutig Nichtparallelitäten auf. Die erste Grafik verdeutlicht auch hier, dass die Chance Verkehrsbereich Fuß gegenüber Verkehrsbereich IV zu wählen bei einer durchschnittlichen Entfernung von 12 km in etwa Null ist. Da sich die Struktur bei einer angemessenen Entfernung nicht ändert, kann für angemessene Entfernung, beispielsweise von 1 bis 5 km, wie folgt argumentiert werden. Vergleicht man beispielsweise in der ersten Grafik den Verlauf einmal mit Besitz, einmal ohne Besitz einer Bahncard bei niedrigem Haushaltseinkommen erkennt man eindeutig die unterschiedlichen Steigungen. Während bei Nichtbesitz einer Bahncard der Verlauf

mit zu zunehmendem persönlichem Einkommen steigt, fällt er bei Besitz einer Bahncard. Umgekehrt steigt bei Nichtbesitz einer Bahncard der Verlauf beim Übergang zu hohem persönlichem Einkommen, er fällt bei Besitz einer Bahncard. Am häufigsten gehen Personen mit niedrigem persönlichem, aber hohem Haushaltseinkommen zu Fuß. Diese Personen sind nicht im Besitz einer Bahncard. Eine mögliche Ursache kann sein, dass diesen Personen nicht immer ein Fahrzeug zu Verfügung steht. Um Interaktionen im dritten Plot zu erkennen ist es günstiger die log-odds zu betrachten. Es ergibt sich folgende Grafik:

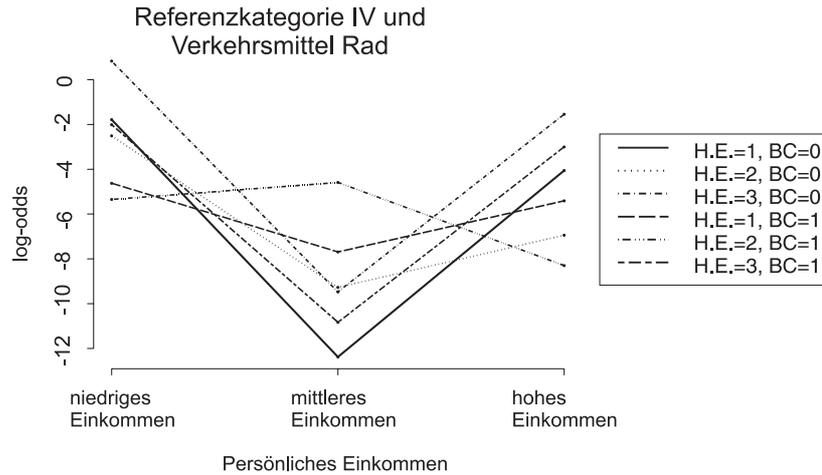


Abbildung 23: Geschätzte log-odds in Abhängigkeit von Bahncard-Besitz und Einkommen

Vergleicht man beispielsweise, unter der Annahme das eine Bahncard vorhanden ist, die Verläufe bei niedrigem Haushaltseinkommen mit denen bei mittleren oder hohem Haushaltseinkommen, wird deutlich, dass diese nicht parallel verlaufen. Folglich ist auch hier von Interaktionen auszugehen.

Im Folgenden soll untersucht werden, ob Personen mit niedrigem persönlichem Einkommen ein hohes Haushaltseinkommen zu Verfügung haben und daher andere Verkehrsmittelwahlentscheidungen treffen. Dazu werden die relativen Chancen berechnet, d.h. für alle Haushaltseinkommensgruppen werden die odds für mittleres und hohes persönlichem Einkommen gegenüber niedrigem persönlichem Einkommen betrachtet. Das Vorgehen wird an einem Beispiel erläutert. Für einen Vergleich niedrigem persönlichem Einkommens (n. P.E.) (als Referenzkategorie) mit hohem persönlichem Einkommen (h. P.E.) für alle Haushaltseinkommensgruppen (mit niedrigem Haushaltseinkommen als Referenzkategorie) ergibt sich

$$\begin{aligned} & \frac{P(Y = r \mid x_{\text{HH-Einkommen}=K(i)}, x_{\text{h. P.E.}}, \mathbf{x}) / P(Y = IV \mid x_{\text{HH-Einkommen}=K(i)}, x_{\text{h. P.E.}}, \mathbf{x})}{P(Y = r \mid x_{\text{HH-Einkommen}=K(i)}, x_{\text{n. P.E.}}, \mathbf{x}) / P(Y = IV \mid x_{\text{HH-Einkommen}=K(i)}, x_{\text{n. P.E.}}, \mathbf{x})} \\ &= \frac{\exp(\beta_{0r} + \beta_{K(i),r} + \beta_{\text{h. P.E.},r} + \beta_{K(i)*\text{h. P.E.},r} + \mathbf{x}^T \boldsymbol{\beta}_r)}{\exp(\beta_{0r} + \beta_{K(i),r} + \beta_{\text{n. P.E.},r} + \beta_{K(i)*\text{n. P.E.},r} + \mathbf{x}^T \boldsymbol{\beta}_r)} = \exp(\beta_{\text{h. P.E.},r} + \beta_{K(i)*\text{h. P.E.},r}) \end{aligned}$$

Damit ergeben sich folgende Plots:

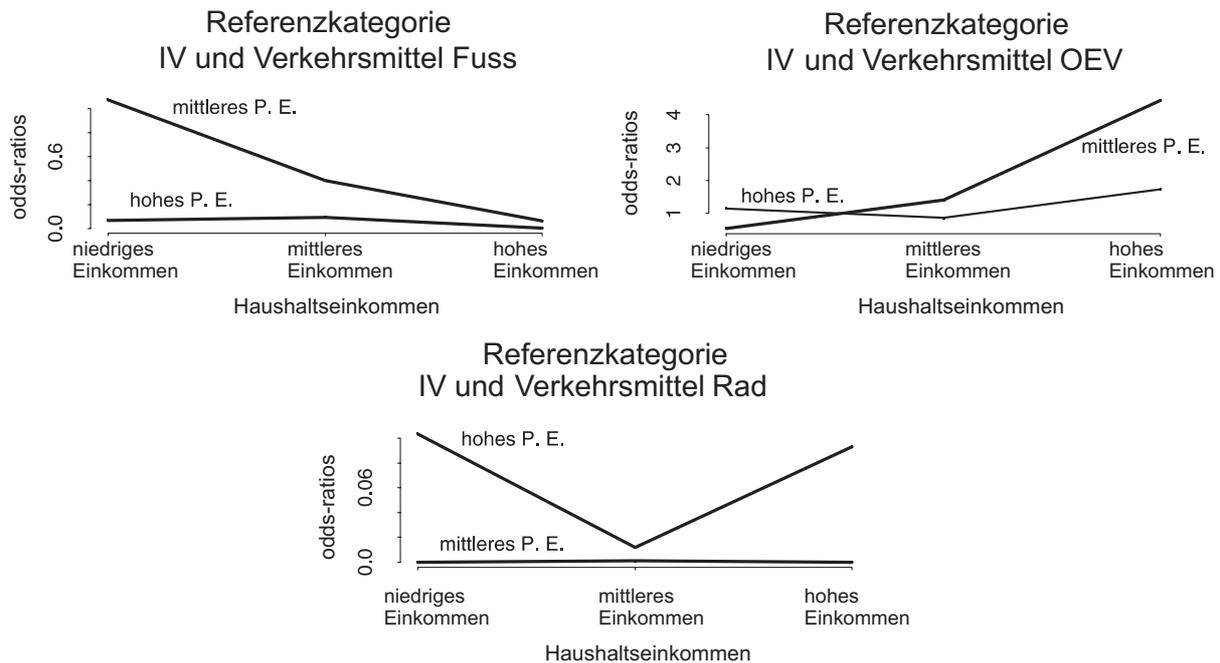


Abbildung 24: Geschätzte Chancenverhältnisse in Abhängigkeit von Haushalts- und persönlichen Einkommen

Zu beachten ist zunächst, dass Singlehaushalte mit hohem persönlichen Einkommen bereits in die Kategorie mittleres Haushaltseinkommen fallen. Die Chance bei einem mittleren Haushaltseinkommen Verkehrsbereich Fuß gegenüber IV zu wählen bei mittlerem persönlichen Einkommen ist niedriger als die Chance bei niedrigem persönlichen Einkommen. Bei hohem Haushaltseinkommen ist die Chance Fuß gegenüber IV zu wählen bei hohem persönlichen Einkommen ebenfalls geringer als bei niedrigem persönlichen Einkommen. Dieser Zusammenhang verdeutlicht, dass mit zunehmendem Einkommen vermehrt IV benutzt wird. Für die Wahl zwischen ÖV und IV kann beim Übergang zu hohem persönlichen Einkommen kein Einfluss festgestellt werden. Bei einem mittleren persönlichen Einkommen wächst die Chance ÖV gegenüber IV zu wählen mit zunehmendem Haushaltseinkommen im Vergleich der Chance bei niedrigem persönlichen Einkommen. In diesem Fall handelt es sich in der Regel um Mehrpersonenhaushalte. Eine mögliche Ursache kann sein, dass nicht alle Personen eines Haushalts ständig einen Pkw zur Verfügung haben und folglich öffentliche Verkehrsmittel nutzen. Beide Grafiken verdeutlichen, dass variierendes persönliches Einkommen zu unterschiedlichen Verkehrsmittelwahlverhalten führt. Im Vergleich zu niedrigem persönlichen Einkommen werden Wege zu Fuß, bei mittlerem persönlichen Einkommen, aber steigendem Haushaltseinkommen, seltener. Bei Wegen die mit öffentlichen Verkehrsmitteln zurückgelegt werden ist ein umgekehrter Trend zu erkennen.

Analog wird dieses Vorgehen für die Interaktion persönliches Einkommen mit Bahncard durchgeführt. Die Ergebnisse zeigen, dass ein Einfluss zwischen Rad und IV besteht. Bei

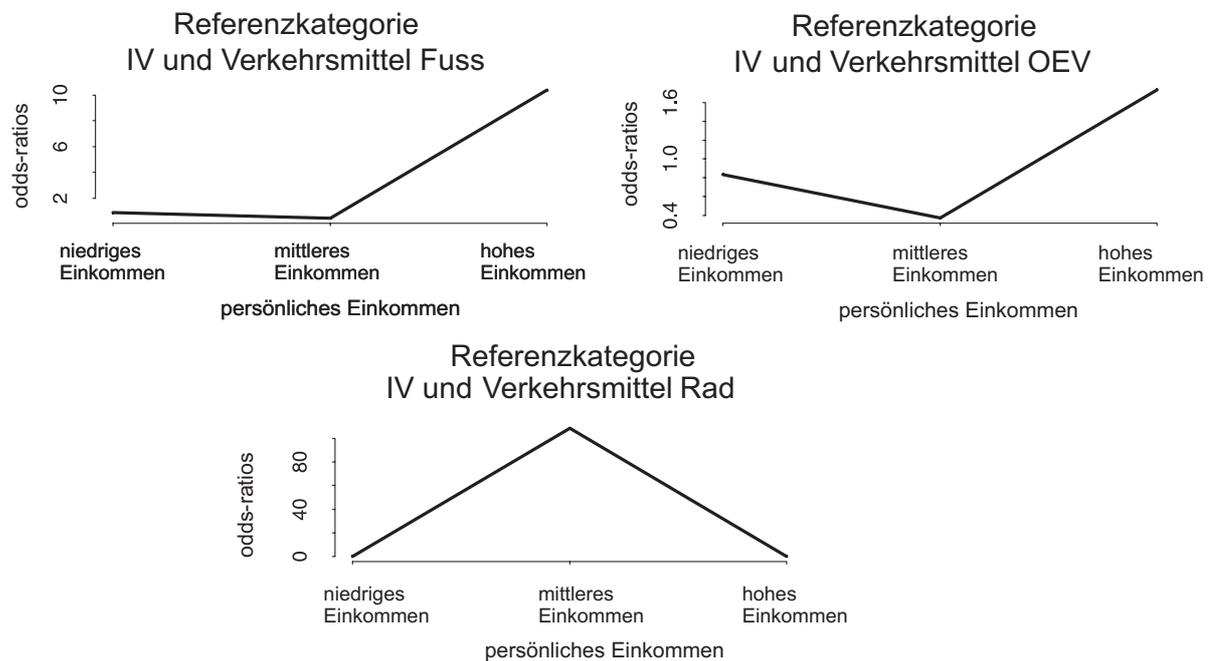


Abbildung 25: Geschätzte Chancenverhältnisse in Abhängigkeit von persönlichen Einkommen und Bahncard-Besitz

Besitz einer Bahncard steigt die Chance Rad gegenüber IV zu nutzen bei einem mittleren persönlichen Einkommen gegenüber niedrigem persönlichen Einkommen. Eine mögliche Ursache kann sein, dass Wege in der Freizeit zum einem Teil mit dem Fahrrad und zum anderen Teil mit der Bahn zurückgelegt werden.

Alter und Nutzung: Zunächst sei noch einmal darauf hingewiesen, dass Alter nicht linear sondern quadratisch in das Modell eingeht. Der Befehl `poly(S.ALTER.N, 2)` in S-Plus erzeugt eine Basis für polynomiale Regression, der zweite Ausdruck gibt den Grad des Polynoms an. Es ergeben sich folglich zwei Vektoren die orthogonal aufeinander stehen.

Mit den zu Beginn getroffenen Annahmen ergeben sich aus den odds folgende Plots:

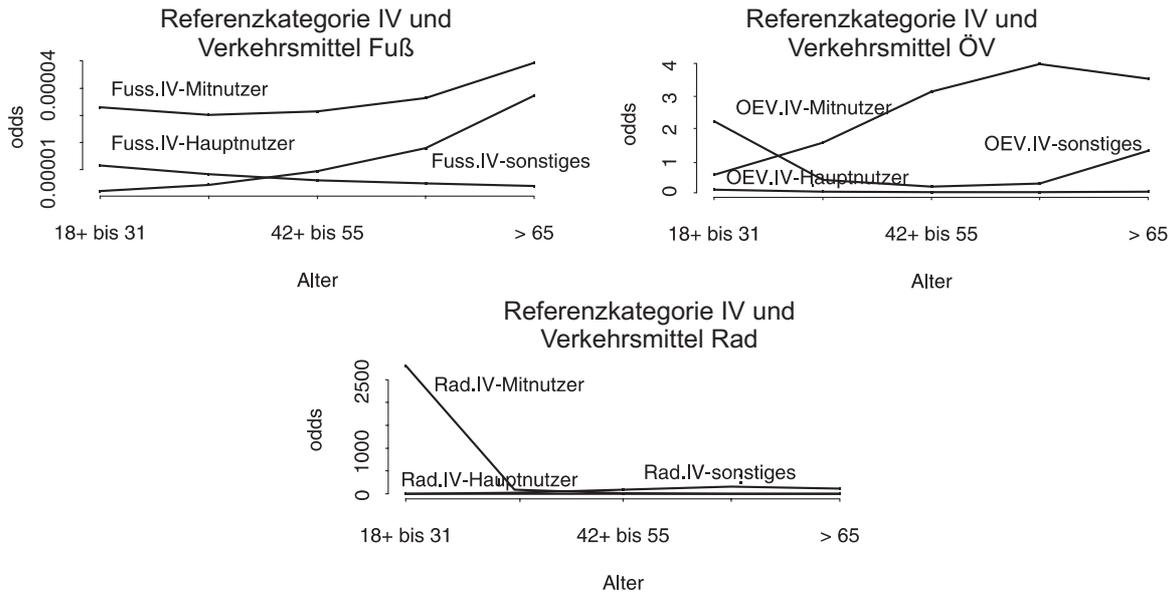


Abbildung 26: Geschätzte odds in Abhängigkeit von Alter und Nutzung

Es wird hier festgestellt, dass der Verlauf nicht parallel ist. Damit wird noch einmal bestätigt, dass Interaktionen zwischen Alter und Nutzung vorliegen. Auch hier sei noch einmal darauf hingewiesen, dass als durchschnittliche Entfernung 12 km gewählt wurde. Daher ist die Chance Fuß gegenüber IV zu nutzen nahezu Null. Wird die Entfernung, unter ansonsten unveränderten gegebenen Einflussgrößen, geändert, ändert sich nur die y-Achsen Skalierung. Grafik 1 in Abbildung 26 zeigt das Pkw-Hauptnutzer mit zunehmendem Alter verstärkt IV benutzen. Am häufigsten gehen Pkw-Mitnutzer Wege zu Fuß. Von den Pkw-Hauptnutzern gehen die jüngeren Personengruppen am häufigsten Wege zu Fuß. Von den Personen, die keinen Pkw zur Verfügung haben oder aber keinen Führerschein besitzen wird in den jüngeren Altersgruppen am wenigsten zu Fuß gegangen. Aus allen drei Grafiken ist ersichtlich, dass Pkw-Hauptnutzer die Verkehrsbereiche Fuß, ÖV bzw. Rad am wenigsten benutzen. Die Grafik für den Vergleich von IV und ÖV zeigt den quadratischen Effekt, wobei zu beachten ist, dass die Skala für Alter nicht äquidistant ist. Um das Chancenverhältnis darzustellen werden die odds-ratios berechnet, wobei Pkw-Hauptnutzer als Referenzkategorie dient.

		Alter				
		18+ bis 31	31+ bis 42	42+ bis 55	55+ bis 65	> 65
Fuß-IV	Mitnutzer	2.84	3.48	4.78	7.02	12.01
	sonstiges	0.19	0.54	1.49	3.54	9.01
ÖV-IV	Mitnutzer	20.65	9.82	8.74	13.66	40.7
	sonstiges	5.17	41.75	139.28	187.53	113.48
Rad-IV	Mitnutzer	134.69	586.95	59.01	0.28	0
	sonstiges	10.77	157.06	910.08	1940.04	1894.44

Tabelle 17: Geschätzte odds-ratios in Abhängigkeit von Alter und Nutzung

Damit ergeben sich folgende Plots:

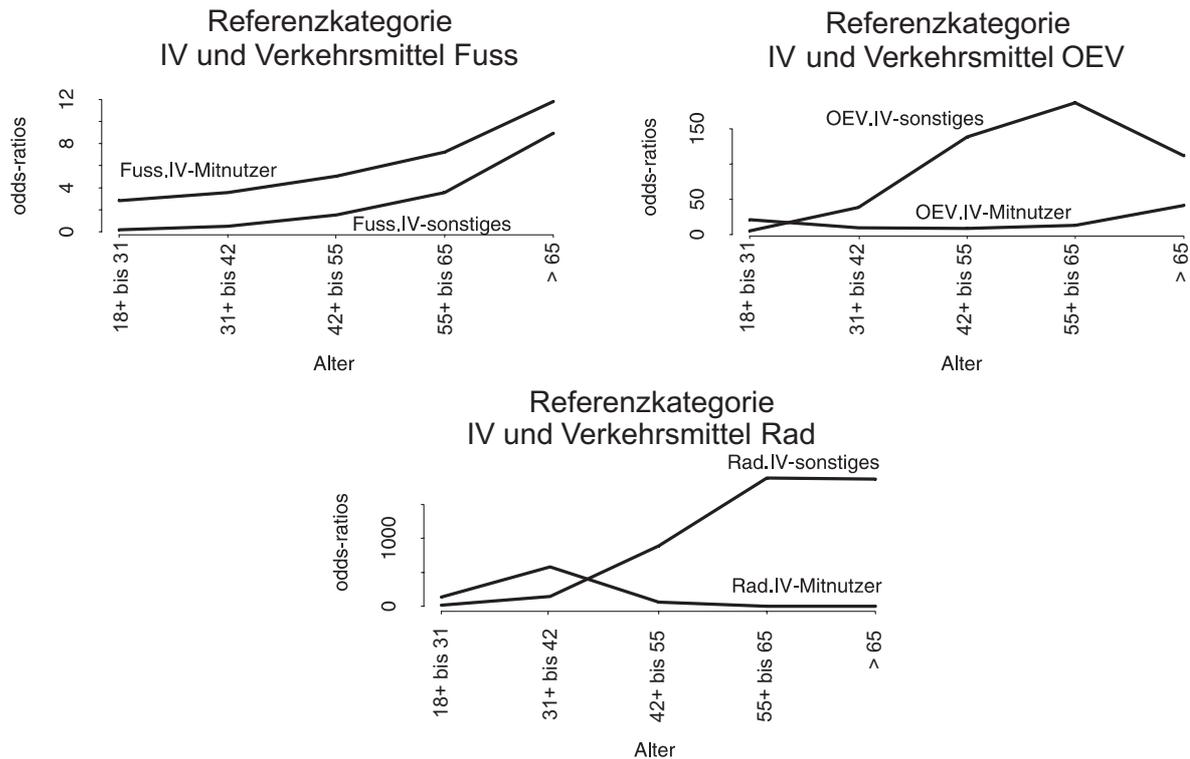


Abbildung 27: Geschätzte Chancenverhältnisse in Abhängigkeit von Alter und Nutzung

Die Chance eines Pkw-Mitnutzers Kategorie Fuß gegenüber IV zu wählen beträgt in der Altersgruppe der 18 bis 31-jährigen in etwa das 3-fache der Chance eines Pkw-Hauptnutzers. Abbildung 27 zeigt, dass mit zunehmendem Alter diese Chance weiter wächst. Analog kann für die Chance der Nutzer in Kategorie sonstiges argumentiert werden. Die zweite Grafik in Abbildung 27 zeigt, dass mit zunehmendem Alter die Chancen einer Person in Kategorie sonstiges Verkehrsbereich Fuß gegenüber IV zu wählen bis zu der Altersgruppe der 55 bis 65-jährigen steigt, in der Altersgruppe der über 65-jährigen jedoch wieder rückläufig ist. Im Vergleich zur Chance der Pkw-Hauptnutzer ist die Chance für Pkw-Mitnutzer zwischen den Verkehrsbereichen ÖV und IV zu wählen zunächst rückläufig, steigt aber ab einem Alter von 42 Jahren wieder an. Die dritte Grafik zeigt, dass die Chance Rad gegenüber IV zu wählen für Nutzer der Kategorie sonstiges mit zunehmendem Alter stark ansteigt, während die Chance der Pkw-Mitnutzer mit Ausnahme der Altersgruppe der 31 bis 42-jährigen abnimmt.

Beurteilung des Modells

Die Residuen geben wichtige Hinweise zur Güte Anpassung (vgl. Abschnitt 3.7.2). Die folgenden Abbildungen stellen die Abweichungen $p_{ir} - \hat{\pi}_{ir}$, für $i = 1, \dots, g$ und $r = \text{Fuß, ÖV und Rad}$ dar, einmal über η_r und einmal über i abgetragen.

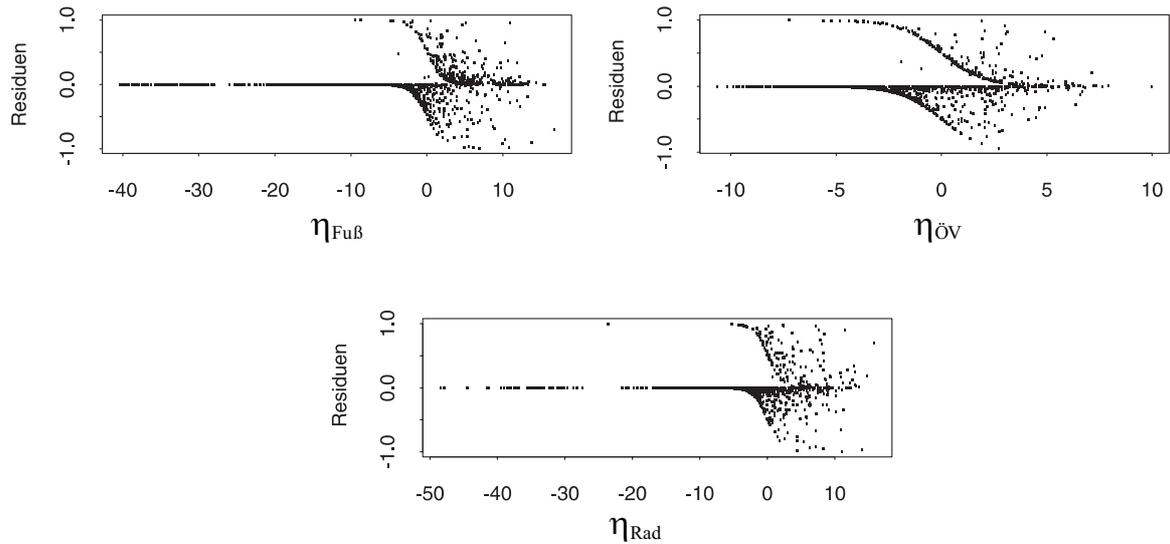


Abbildung 28: Residuenplot in Abhängigkeit von $\eta_r = \mathbf{x}^T \boldsymbol{\beta}_r$

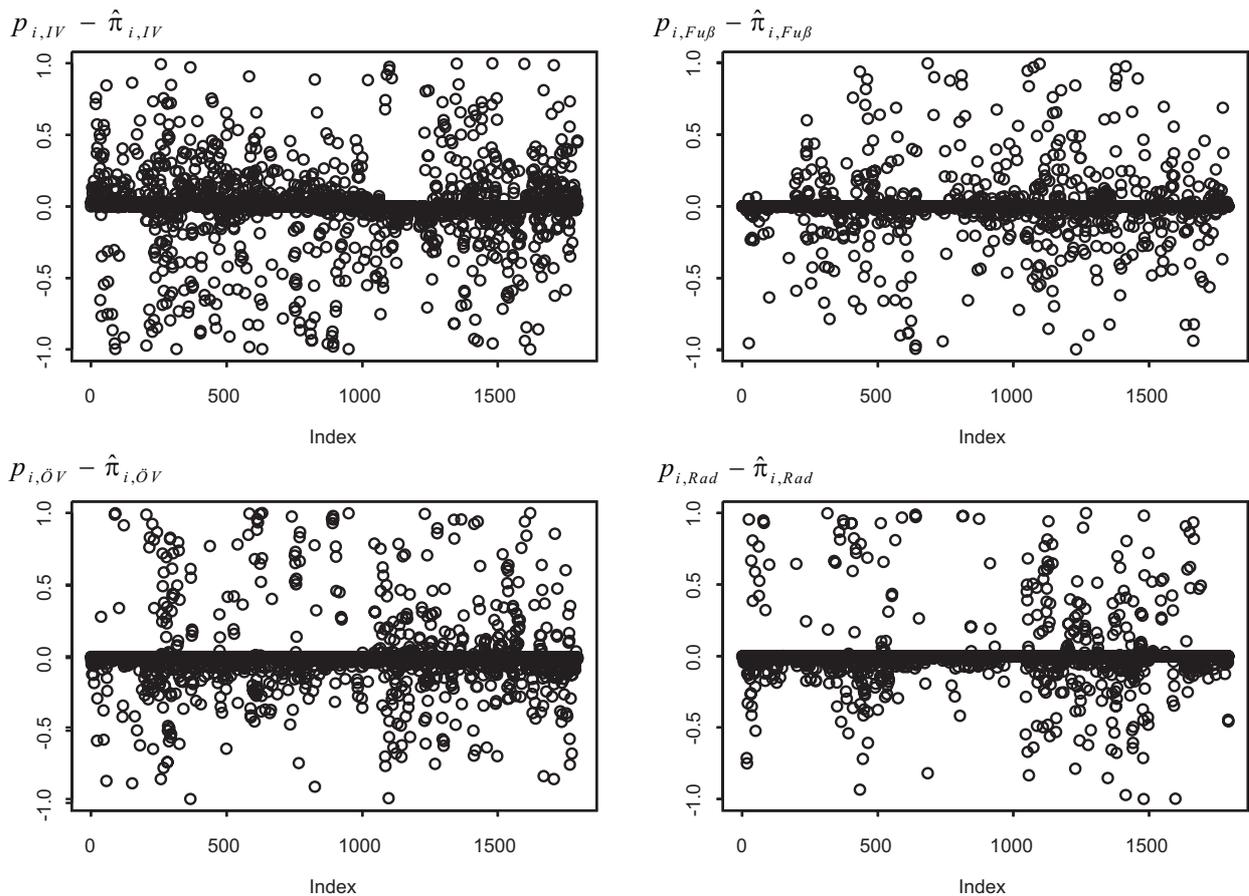


Abbildung 29: Residuenplot in Abhängigkeit von den Beobachtungen

Auffallend ist, dass bei positiven η_r die Residuen stärker um Null schwanken, bei negativen η_r hingegen so gut wie gar nicht. Ein Grund liegt darin, dass bei $\hat{\pi}_{i_r}$ ein positives η_r um

so stärker wirkt, als ein negatives η_r . Das Modell wird den Daten einigermaßen gerecht. Der größte Anteil der Residuen streut um Null. Abbildung 29 zeigt diesen Zusammenhang sehr deutlich. Das in Abschnitt 5.2.4 entwickelte Modell zum Verkehrsmittelwahlverhalten weist 1793 nichtleere Zeilen auf. Insgesamt werden nur 2554 Wege zurückgelegt. Daran erkennt man, dass pro Zelle nur sehr wenige Beobachtungen existieren. Das Modell liefert eine Residual Deviance von 1529.629. Damit ergibt sich folglich ein p-Wert von 1.

Nach Abschnitt 3.7.7 kann die Effektivität des Modells mit Hilfe des Pseudo R^2 von McFadden erklärt werden. Für das multinomiale Logitmodell ergibt sich ein McFadden Pseudo R^2 von 0.688. Die hohe Zahl spricht abermals dafür, dass der Informationsgehalt, der in den erklärenden Variablen steckt, relativ hoch ist. Das Modell wird den Daten einigermaßen gerecht.

5.4 Profile der Verkehrsmittelwahl

Im Folgenden ist man an der Verkehrsmittelwahl von Personen mit einem bestimmten Profil interessiert. Einerseits können Profile direkt aus der Designmatrix abgelesen werden (Beispiele 5.4.1 und 5.4.2). Andererseits können beliebige Profile betrachtet werden.

Beispiel 5.4.1. *Untersucht werden soll die Verkehrsmittelwahl eines 31- bis 42-jährigen Singles, der Freitags alleine, innerorts einen etwa 3.1 km langen Weg in ca. einer Stunde zurücklegt. Diese Person soll über ein hohes Haushaltseinkommen und folglich über ein hohes persönliches Einkommen verfügen, eine Bahncard und ein Fahrrad besitzen. Die Person ist Pkw-Hauptnutzer, hat aber eine negative Einstellung zum IV. Eine Anbindung an den schienengebundenen Nahverkehr ist vorhanden. Die Entfernung zur nächsten Haltestelle soll weniger als 1 km betragen, die Anbindung an ÖV zufriedenstellend sein. Der Eindruck des Weges soll durchschnittlich sein. Es genügt zu den vier Verkehrsmitteln drei Logits zu parametrisieren:*

$$\begin{aligned} \log \left(\frac{P(Y = r | \mathbf{x})}{P(Y = IV | \mathbf{x})} \right) = & \beta_{0r} + \beta_{\text{Anb. an } \ddot{O}V, r} + \beta_{\text{innerorts}, r} + \beta_{\text{Freitag}, r} + \beta_{\text{durchschnittl. Eindruck}, r} \\ & + \beta_{\text{hohes P.E.}, r} + \beta_{\text{sehr zufr. m. Anb.}, r} + \beta_{\text{neg. Einst. IV}, r} + \beta_{\text{Sport}, r} \\ & + \beta_{\text{Alter}, r} * \text{Alter} + \beta_{\text{Alter}^2, r} * \text{Alter}^2 + \beta_{\text{Entfernung}, r} * \text{Entfernung} \\ & + \beta_{\text{Wegedauer}, r} * \text{Wegedauer} + \beta_{\text{Entf. zum } \ddot{O}V, r} * \text{Entf. zum } \ddot{O}V \\ & + \beta_{\text{hohes H.E.}, r} + \beta_{\text{Bahncard}, r} + \beta_{\text{Rad-Verf.}, r} , \end{aligned}$$

mit $r = \text{Fuß}$, ÖV und Rad . Einsetzen der Koeffizienten in diese Gleichung ergibt

$$\log\left(\frac{P(Y = \text{Fuß} | \mathbf{x})}{P(Y = \text{IV} | \mathbf{x})}\right) = 13.242$$

$$\log\left(\frac{P(Y = \text{ÖV} | \mathbf{x})}{P(Y = \text{IV} | \mathbf{x})}\right) = 5.624$$

$$\log\left(\frac{P(Y = \text{Rad} | \mathbf{x})}{P(Y = \text{IV} | \mathbf{x})}\right) = 6.300.$$

Damit können im Folgenden die einzelnen Wahrscheinlichkeiten berechnet werden, d.h.

$$P(Y = \text{IV} | \mathbf{x}) = \frac{1}{1 + \exp(13.242) + \exp(5.624) + \exp(6.3)} = 0$$

$$P(Y = \text{Fuß} | \mathbf{x}) = \frac{\exp(13.242)}{1 + \exp(13.242) + \exp(5.624) + \exp(6.3)} = 0.998$$

$$P(Y = \text{ÖV} | \mathbf{x}) = \frac{\exp(5.624)}{1 + \exp(13.242) + \exp(5.624) + \exp(6.3)} = 0.001$$

$$P(Y = \text{Rad} | \mathbf{x}) = \frac{\exp(6.3)}{1 + \exp(13.242) + \exp(5.624) + \exp(6.3)} = 0.001.$$

Die Wahrscheinlichkeit, dass die im vorliegenden Beispiel genannte Person den Verkehrsbereich IV wählt ist Null, die Wahrscheinlichkeit, dass die Person zu Fuß geht beträgt fast 100%. Die Wahrscheinlichkeiten für die Verkehrsbereiche Rad und ÖV sind mit 0.1% verschwindend gering.

Beispiel 5.4.2. Eine 42 bis 55 Jahre alte Person, die in einem 4- bzw. Mehrpersonenhaushalt mit mittlerem Haushaltseinkommen lebt und Pkw-Hauptnutzer ist, möchte mit weiteren Personen am Sonntag bei gutem Wetter einen etwa 13 km langen Weg unter Umständen auch außerorts in 140 Minuten zurücklegen. Ein Fahrrad ist vorhanden. Die Person besitzt eine Bahncard aber keine Netzkarte. Sie verfügt über ein mittleres persönliches Einkommen. Eine Anbindung an öffentliche Verkehrsmittel ist vorhanden mit der die Person sehr zufrieden ist. Die nächstgelegene Haltestelle ist ca. 0.4 km entfernt. Die Person wird mit 76.4% den Verkehrsbereich IV, mit 18.9% ÖV, mit 3.44% Fuß und mit 1.26% Rad wählen.

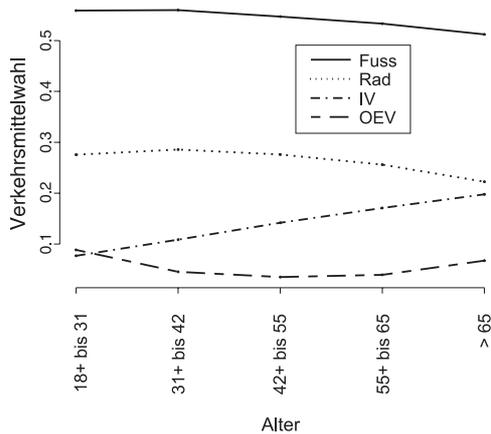
Im Folgenden werden einige ausgewählte Profile betrachtet. Tabelle 18 stellt einen Überblick über alle betrachteten Profile dar. Dabei bedeutet ein Strich, dass die zugehörige Variable variiert.

Variable	Beschreibung	Profilklasse 1	Profilklasse 2	Profilklasse 3	Profilklasse 4
MB.TAG	Mobilitätstag	Mi	-	Mi	Mi
HH.GROESSE	Haushaltsgröße [Personen]	2	2	2	1
HH.EINK.NEU	Haushaltseinkommen	niedrig	mittel	mittel	-
HSSCHIEN	Anbindung an schienengeb. ÖV	ja	ja	ja	ja
S.KM.OEV.MIN	Entfernung zur nächsten Haltestelle [km]	0,4	0,4	0,4	0,4
ZUFRA.ANB.OEV	Zufriedenheit mit der Anbindung an ÖV	insgesamt zufrieden	insgesamt zufrieden	insgesamt zufrieden	insgesamt zufrieden
WETT.GUT	Gutes Wetter	nein	nein	nein	nein
S.ENTF.KM	Entfernung [km]	-	-	6	1
STR.TYP.NEU	Streckentyp	innerorts	innerorts	innerorts	innerorts
WEG.MIT	Weg alleine oder mit mehreren	-	alleine	alleine	alleine
VORHER	Vorher damit unterwegs	ja	ja	ja	ja
GEWOHN	Gewohnheit / bewährt	ja	ja	ja	ja
OHNE.ALT	Keine andere Möglichkeit	nein	nein	nein	nein
GUT.ENTF	Paßt zur Entfernung	ja	ja	ja	ja
KURZ.T	Geringer Zeitaufwand	ja	ja	ja	ja
BEQUEM	Bequem	nein	nein	-	nein
FLEXIBEL	Flexibel	nein	nein	nein	nein
KOSTEN	Kostengünstig	ja	ja	ja	nein
SPORT	Als Sport / zur Bewegung	nein	nein	nein	nein
S.DAUERMIN	Dauer des Wegs [min]	5	30	-	10
EINDRUCK	Eindruck	angenehm	angenehm	angenehm	angenehm
S.ALTER.N	Alter	-	37	37	-
NUTZUNG	PKW-Nutzung	-	-	-	-
RAD.VERF	Rad-Verfügung	ja	ja	ja	ja
PERS.EINK.NEU	Persönliches Einkommen	niedrig	niedrig	hoch	-
NETZKARTE	Besitz einer Netzkarte	ja	-	ja	ja
BAHNCARD	Besitz einer Bahncard	nein	nein	nein	nein
EINPKW.N	Negative Einstellung zum IV	nein	-	nein	nein

Tabelle 18: Profile

Profilklasse 1: Um das Vorgehen besser zu verstehen, wird dieses Profil ausführlich beschrieben.

Zunächst wird das Verkehrsmittelwahlverhalten einer Person eines 2 Personenhaushaltes betrachtet. Diese Person soll Pkw-Hauptnutzers sein, einen Weg von ca. 0.6 km in 5 Minuten innerorts am Mittwoch alleine zurücklegen. Sie soll der Kategorie mit niedrigem persönlichen Einkommen und niedrigem Haushaltseinkommen zugeordnet sein, ein Fahrrad besitzen und in etwa in 0.4 km eine Haltestelle des ÖV erreichen. Eine Anbindung an den schienengebundenen Nahverkehr soll vorhanden sein, eine Netzkarte oder Bahncard nicht. Die Anbindung an öffentliche Verkehrsmittel soll zufrieden stellend sein. Motive für die Wahl sollen durch Gewohnheit geprägt sein und zur Entfernung passen. Die Strecke soll innerhalb kürzester Zeit zurückgelegt werden. Zudem soll eine kostengünstige Alternative in Vordergrund stehen. Die Person kann vorher mit diesem Verkehrsmittel bereits unterwegs gewesen sein. Betrachtet wird dieses Profil über alle Altersklassen.



Das Streckenprofil von 0.6 km in 5 Minuten lässt bereits vermuten, dass in diesem Fall die Bereiche Rad und Fuß bevorzugte Wahl sein werden. Abbildung 30 bestätigt diese Vermutung. Auffallend ist, dass das Verkehrsmittelwahlverhalten in den verschiedenen Altersklassen sich nicht stark unterscheidet.

Abbildung 30: Profil – Pkw-Hauptnutzer alleine unterwegs

Variiert man die Nutzergruppen so ergibt sich für Pkw-Mitnutzer ein Verkehrsmittelwahlverhalten gemäß Abbildung 31. Das Verkehrsmittelwahlverhalten dieses Profils in den verschiedenen Altersklassen ist sehr unterschiedlich. Die ersten drei Altersgruppen wählen verstärkt das Fahrrad, während die letzten beiden verstärkt den Bereich Fuß wählen. Für Nutzer der Kategorie sonstiges erhält man Abbildung 32. Dieses Profil wählt verstärkt das Fahrrad, da die Nutzung von IV eingeschränkt ist und ein kostengünstiges Verkehrsmittel bevorzugt wird. Das Verkehrsmittelwahlverhalten unterscheidet sich, wie schon bei den Pkw-Hauptnutzern, in den verschiedenen Altersklassen nur sehr gering.

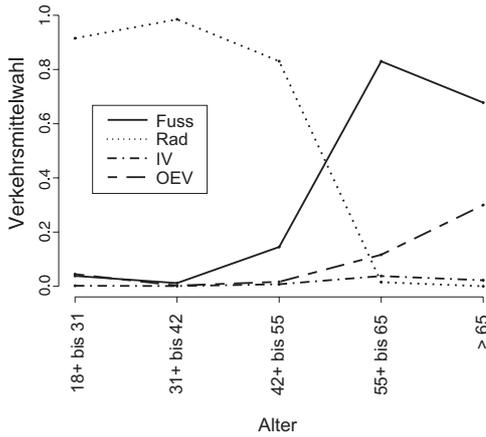


Abbildung 31: Profil – Pkw-Mitnutzer alleine unterwegs

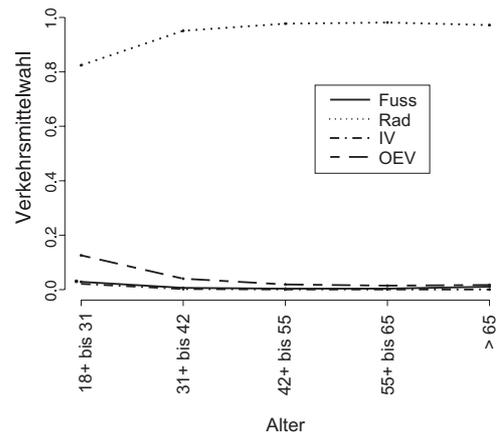
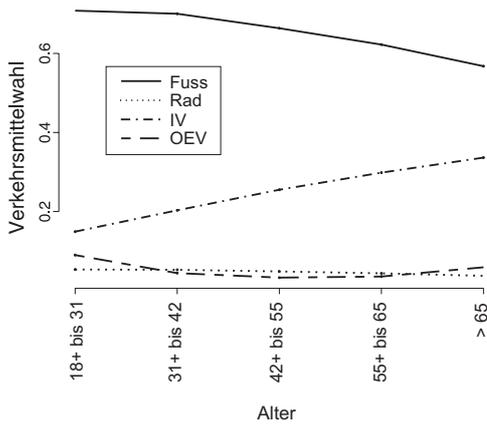
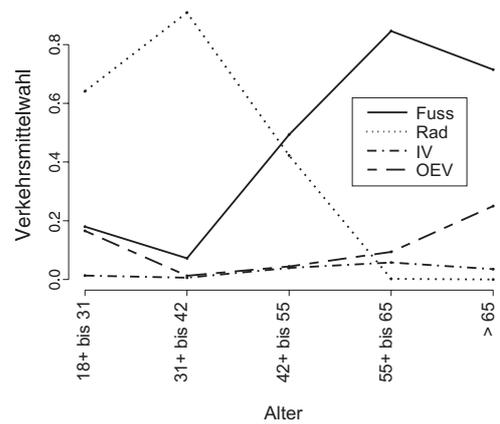


Abbildung 32: Profil – Nutzer der Kategorie sonstiges alleine unterwegs

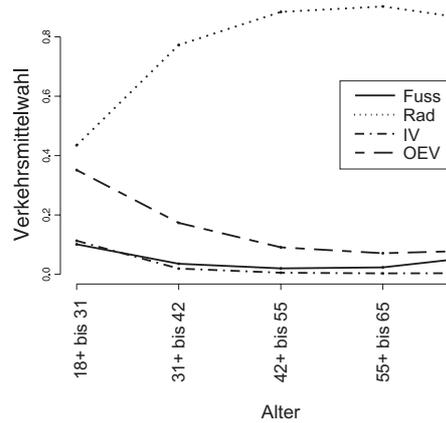
Betrachtet man dieses Profil für einen Weg, wobei dieser nun in Begleitung zurückgelegt wird, ergibt sich hierfür die in Abbildung 33 dargestellte Verkehrsmittelwahl.



Profil – Pkw-Hauptnutzer mit mehreren Personen unterwegs



Profil – Pkw-Mitnutzer mit mehreren Personen unterwegs



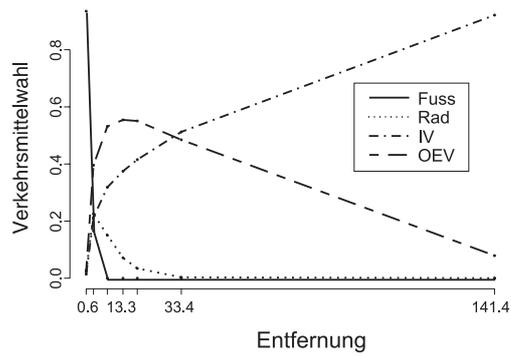
Profil – Nutzer der Kategorie
sonstiges mit mehreren Personen
unterwegs

Abbildung 33: Profilkategorie 1 mit mehreren unterwegs

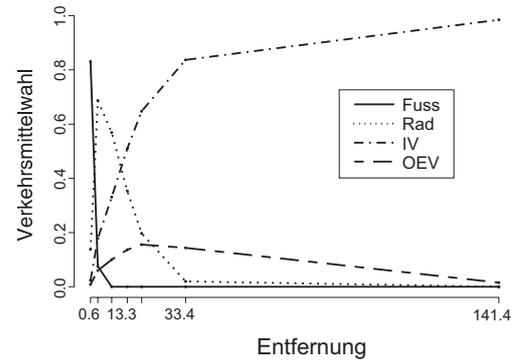
Die Abbildung zeigt, dass beispielsweise Pkw-Hauptnutzer die mit mehreren Personen unterwegs sind für kurze Strecken den Pkw häufiger wählen, als wenn sie alleine unterwegs sind. Die Verkehrsmittelwahl in Abhängigkeit des Alters unterscheidet sich jedoch nicht grundsätzlich für Pkw-Hauptnutzer, die alleine oder in Begleitung unterwegs sind. So nehmen mit zunehmendem Alter die Fahrten mit IV zu, sowohl wenn die Person alleine als auch mit Begleitung unterwegs ist. Die Wege, die zu Fuß zurückgelegt werden nehmen mit zunehmendem Alter leicht ab. Pkw-Mitnutzer wählen insbesondere in den ersten beiden Altersgruppen verstärkt das Fahrrad, in den älteren Altersgruppen wird verstärkt der Bereich Fuß gewählt. Das Verkehrsmittel IV spielt eine untergeordnete Rolle. Nutzer der Kategorie sonstiges wählen verstärkt das Fahrrad. Bei Wegestrecken, die mit mehreren Personen zurückgelegt werden, wird insbesondere in den jüngeren Altersgruppen häufiger ÖV gewählt, als bei Wegestrecken die alleine zurückgelegt werden. Zusammenfassend wird festgestellt, dass für diese kurze Entfernung nur sehr wenig ÖV und IV genutzt wird. IV wird nennenswert nur von Pkw-Hauptnutzern gewählt.

Profilkategorie 2: Im Folgenden werden Profile bei unterschiedlicher Entfernung betrachtet. Es wird eine Wegedauer von 30 Minuten zu Grunde gelegt. Die Person soll hier über ein niedriges persönliches Einkommen, aber ein mittleres Haushaltseinkommen verfügen. Außerdem wird ein Durchschnittsalter von 37 Jahren angenommen. Verglichen werden soll zudem einerseits die Verkehrsmittelwahl für den Mobilitätstag Mittwoch zum Mobilitätstag Sonntag, andererseits wird das Profil einmal für den Fall betrachtet, dass eine Netzkarte vorhanden ist und zum anderen für den Fall das keine Netzkarte vorhanden ist. Zum weiteren Vergleich wird ein Profil betrachtet, bei dem die Person eine negative Einstellung zum IV hat. Das Modell liefert für diese Profile, je nach Nutzergruppe, die in

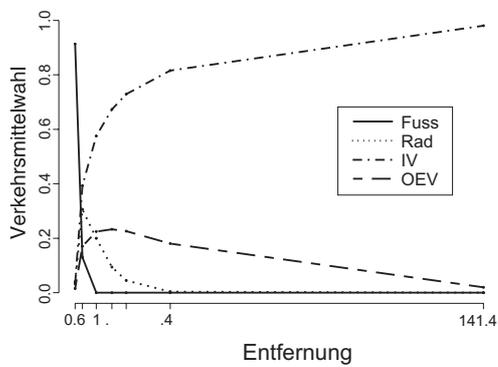
Abbildung 34 bis 36 dargestellten Ergebnisse.



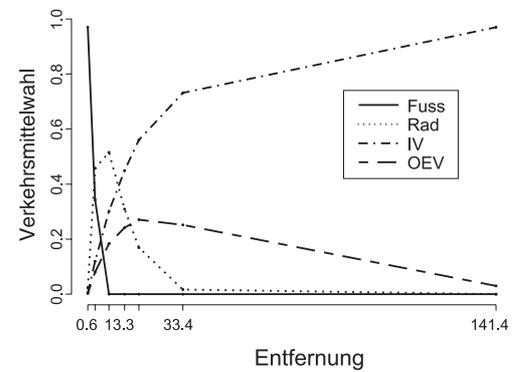
Profil – Pkw-Hauptnutzer, Mittwoch,
Netz Karte



Profil – Pkw-Hauptnutzer, Mittwoch,
keine Netz Karte

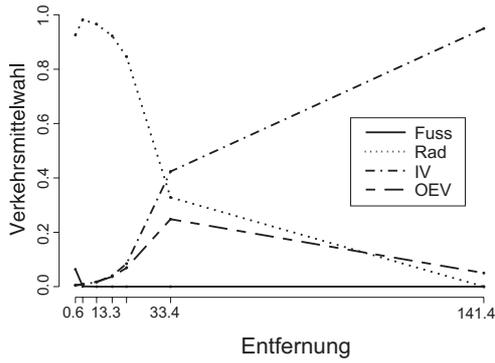


Profil – Pkw-Hauptnutzer, Sonntag,
Netz Karte

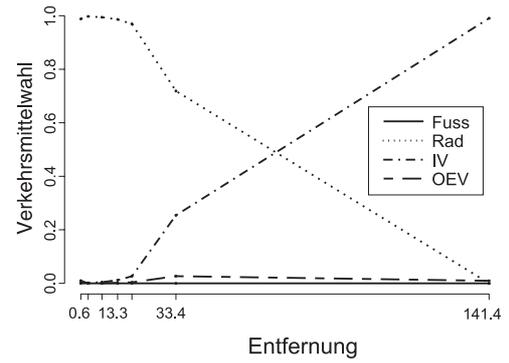


Profil – Pkw-Hauptnutzer, Mittwoch,
keine Netz Karte, negative Einstellung
zum IV

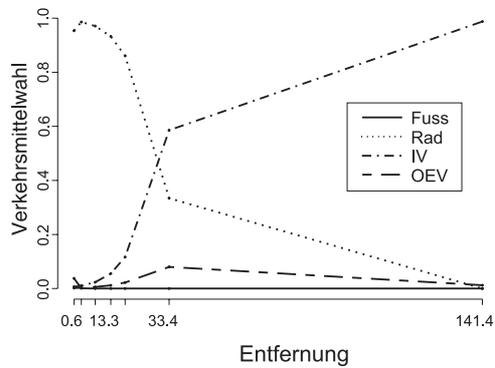
Abbildung 34: Profilkategorie 2 für Pkw-Hauptnutzer bei unterschiedlicher Entfernung und Wochentag



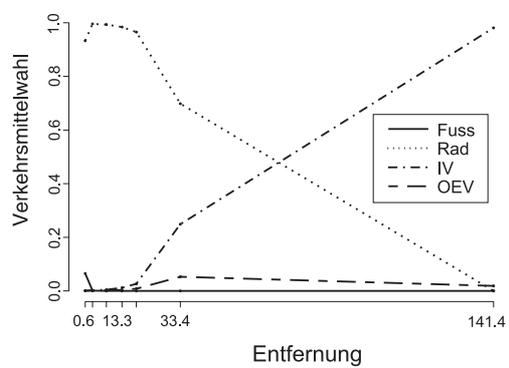
Profil – Pkw-Mitnutzer, Mittwoch, Netzkarte



Profil – Pkw-Mitnutzer, Mittwoch, keine Netzkarte

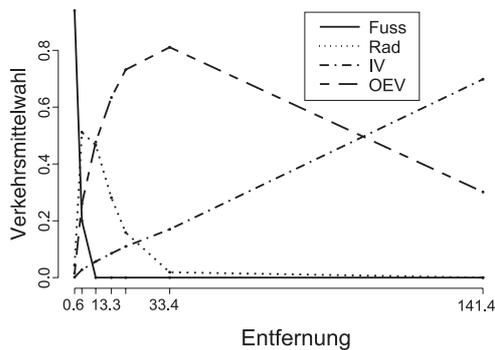


Profil – Pkw-Mitnutzer, Sonntag, Netzkarte

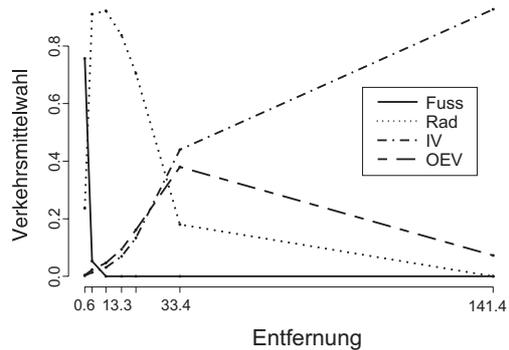


Profil – Pkw-Mitnutzer, Mittwoch, keine Netzkarte, negative Einstellung zum IV

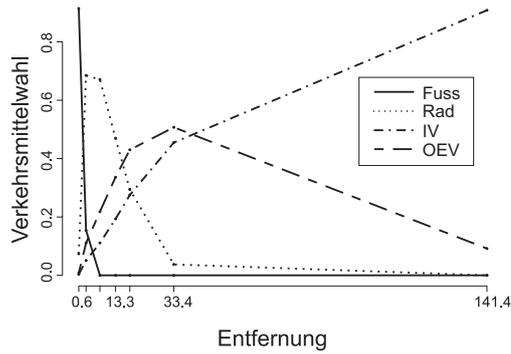
Abbildung 35: Profilklasse 2 für Pkw-Mitnutzer bei unterschiedlicher Entfernung und Wochentag



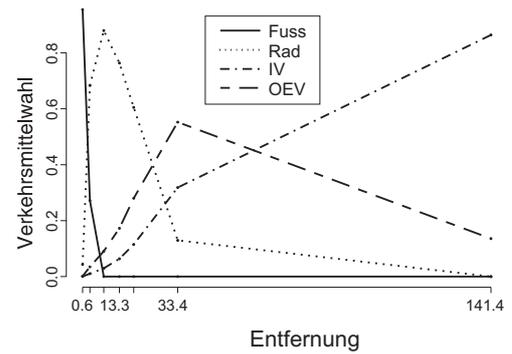
Profil – Nutzer der Kategorie sonstiges, Mittwoch, Netzkarte



Profil – Nutzer der Kategorie sonstiges, Mittwoch, keine Netzkarte



Profil – Nutzer der Kategorie
sonstiges, Sonntag, Netzkarte

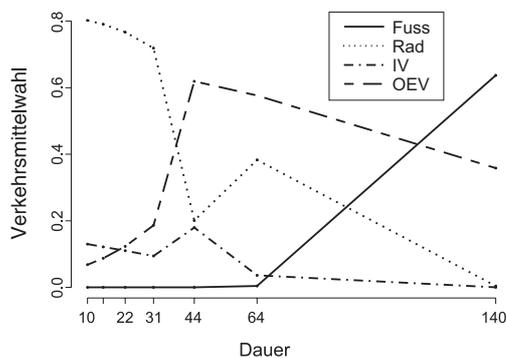


Profil – Nutzer der Kategorie
sonstiges, Mittwoch, keine Netzkarte,
negative Einstellung zum IV

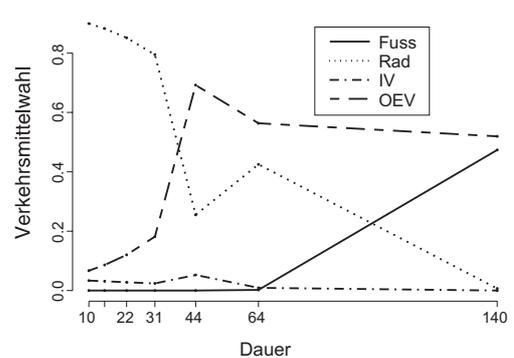
Abbildung 36: Profilkategorie 2 für Kategorie sonstige bei unterschiedlicher Entfernung und Wochentag

Bei niedriger Entfernung wird noch vermehrt zu Fuß gegangen oder das Fahrrad benutzt, mit zunehmender Entfernung sind IV und ÖV die am häufigsten verwendeten Verkehrsmittel. Besitzer einer Netzkarte nutzen insbesondere wochentags verstärkt ÖV (z.B. für Arbeitswege). Am häufigsten wird ÖV von Nutzern der Kategorie sonstige gewählt, sofern eine Netzkarte vorhanden ist. Die Nutzung von IV bei negativer Einstellung zum IV ist nicht bedeutend niedriger als bei positiver Einstellung.

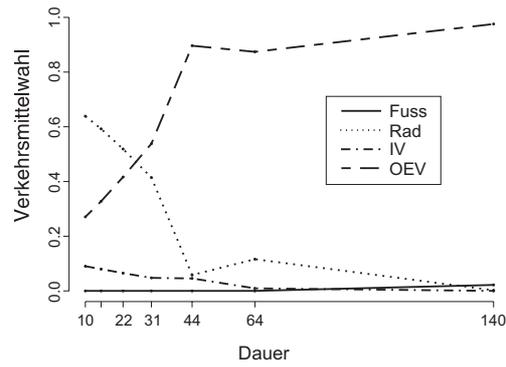
Profilkategorie 3: Nachfolgende Profile (Abbildung 37 und 38) zeigen das Verkehrsmittelwahlverhalten bei unterschiedlicher Wegedauer erneut für alle drei Nutzergruppen, einmal mit der subjektiven Größe Bequemlichkeit, d.h. das Verkehrsmittel wird gewählt weil es bequem ist und einmal ohne diese Größe. Als Entfernung soll 6 km zu Grunde gelegt werden. Als Mobilitätstag wird der Mittwoch gewählt. Die Person soll über eine Netzkarte verfügen. Es wird außerdem festgelegt, dass die Person sowohl über ein hohes persönliches als auch über ein hohes Haushaltseinkommen verfügen soll. Alle übrigen Größen bleiben unverändert (siehe Tabelle 18).



Profil – Pkw-Hauptnutzer



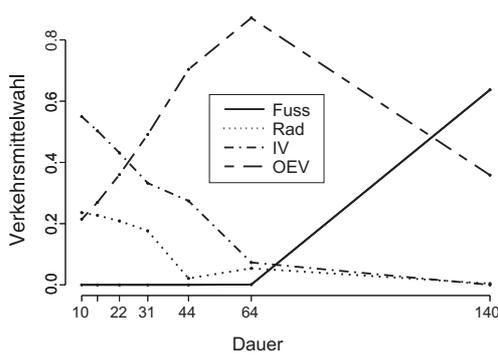
Profil – Pkw-Mitnutzer



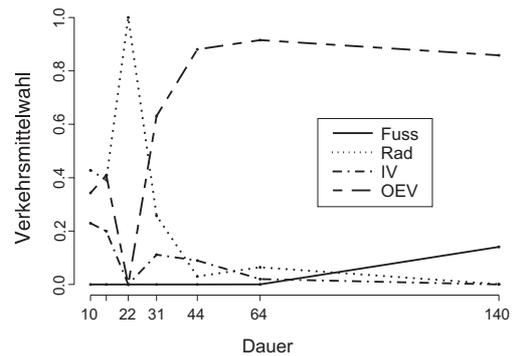
Profil – Nutzer der Kategorie
sonstiges

Abbildung 37: Profilklasse 3 bei unterschiedlicher Wegedauer, falls der Weg als nicht bequem empfunden wurde

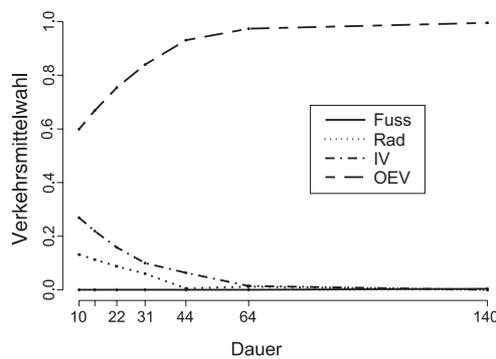
Fügt man die subjektive Größe Bequemlichkeit hinzu, ergeben sich folgende Abbildungen:



Profil – Pkw-Hauptnutzer



Profil – Pkw-Mitnutzer



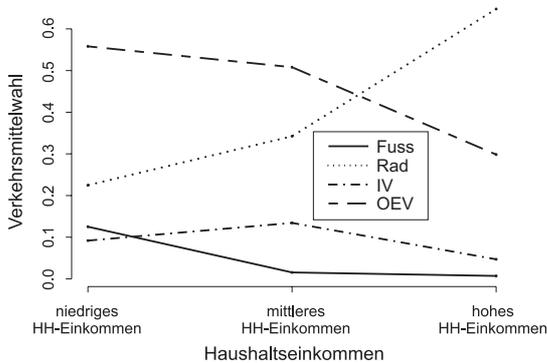
Profil – Nutzer der Kategorie
sonstiges

Abbildung 38: Profilklasse 3 bei unterschiedlicher Wegedauer, falls der Weg als bequem empfunden wurde

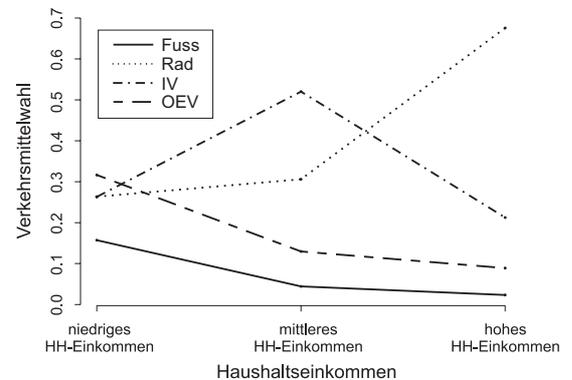
Für die gewählte Wegstrecke von 6 km ist es plausibel, dass die Nutzung von IV mit

zunehmender Dauer fallen wird. Im Vergleich dazu steigt die Dauer stark an, falls diese Strecke zu Fuß zurückgelegt wird. Ab einer durchschnittlichen Dauer von 40 Minuten werden sehr häufig öffentliche Verkehrsmittel genutzt. Für kurze Zeiten wird häufig das Rad gewählt. Die Größe Bequemlichkeit hat einen positiven Einfluss auf die Nutzung von IV, insbesondere bei Wegen bis zu einer drei viertel Stunde. Beispielsweise wählen Pkw-Hauptnutzer bei einer Wegedauer von 10 Minuten zu 13% IV. spielt die Bequemlichkeit eine Rolle bei der Verkehrsmittelwahl, so wählen bereits 55% der Pkw-Hauptnutzer IV. Auffällig ist, dass die Größe Bequemlichkeit für die Wahl öffentlicher Verkehrsmittel ebenfalls einen positiven Einfluss hat, der zum Teil sogar stärker ist, als für der Wahl von IV.

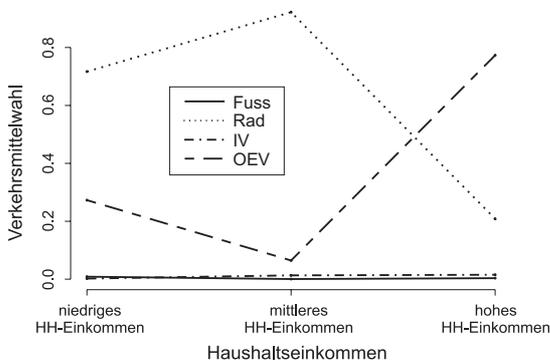
Profilklasse 4 Abschließend wird das Verkehrsmittelwahlverhalten eines Singles, über die verschiedenen Haushaltseinkommensklassen, betrachtet. Variiert wird das Alter. Es soll eine Strecke von 1 km innerhalb von 10 Minuten zurückgelegt werden, wobei eine Netzkarte vorhanden ist. Die Bequemlichkeit soll keine entscheidende Rolle spielen. Kostengünstige Verkehrsmittel werden in diesem Profil nicht bevorzugt gewählt. Alle anderen Größen bleiben unverändert (vgl. Tabelle 18). Damit ergeben sich nachfolgende Abbildungen



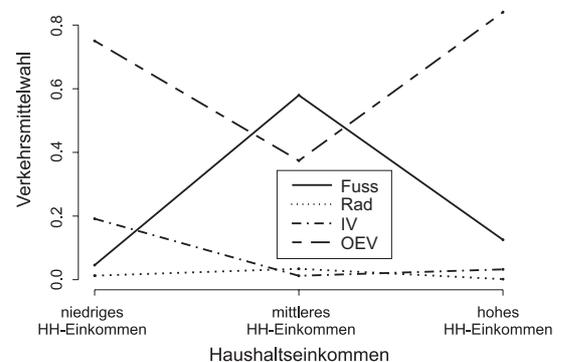
Profil – Pkw-Hauptnutzer, Alter: 25



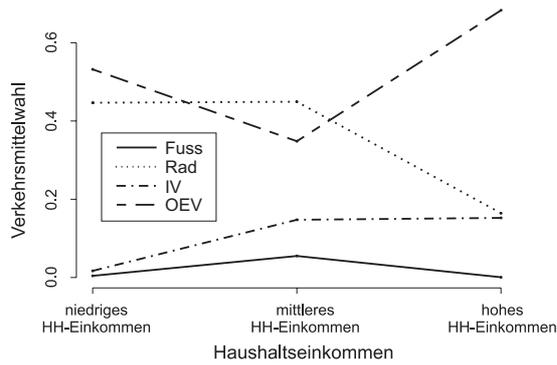
Profil – Pkw-Hauptnutzer, Alter: 60



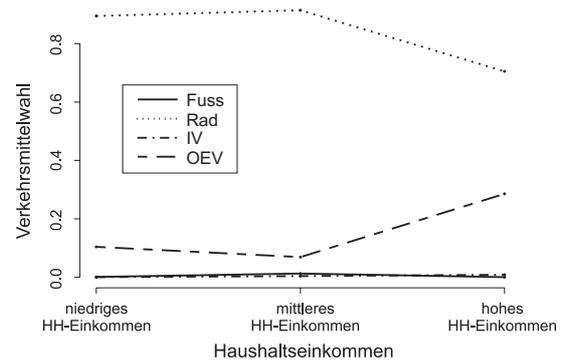
Profil – Pkw-Mitnutzer, Alter: 25



Profil – Pkw-Mitnutzer, Alter: 60



Profil – Nutzer der Kategorie
sonstiges, Alter: 25



Profil – Nutzer der Kategorie
sonstiges, Alter: 60

Abbildung 39: Profilkategorie 4 bei unterschiedlichem Alter

Allen Grafen ist gleich, dass mit zunehmendem Haushaltseinkommen kein eindeutiger Trend vorliegt. Beispielsweise steigt die Wahrscheinlichkeit einer 25-jährigen bzw. einer 60-jährigen Person, die Pkw-Hauptnutzer ist, IV zu wählen mit mittlerem Einkommen, während sie bei hohem Einkommen wieder sinkt. Die Wahl für das Fahrrad hingegen steigt mit zunehmendem Einkommen. Für die verstärkte Nutzung des Fahrrades kann die kurze Entfernung eine Ursache sein. Analog kann für die verbleibenden Nutzergruppen argumentiert werden.

Die Ergebnisse können nicht mit den Ergebnissen in Zängler [29] verglichen werden. Dazu müsste die Grundgesamtheit ebenfalls auf ganz Bayern ausgedehnt werden. Eine Untersuchung mit dem Statistikpaket S-Plus führt bei großen Datenmengen jedoch zu erheblichen Rechenzeiten.

6 Zusammenfassung und Ausblick

In der vorliegenden Arbeit wird ein multinomiales Modell zur Beschreibung und Erklärung des Verkehrsmittelwahlverhaltens privater Haushalte in Alltag und Freizeit entwickelt. Grundlage für die Untersuchung ist die Theorie der multinomialen Logit-Modelle.

Die Entscheidung, welche Einflussgrößen zur Erklärung des Mobilitätsverhaltens beitragen können, kommt dabei eine besondere Bedeutung zu und erfordert ausreichend Erfahrung.

Eine sehr wichtige Einflussgröße für die Verkehrsmittelwahl ist die Entfernung des Weges. Das Modell zeigt insbesondere, dass bei kurzen Strecken (bis 1 km) der Bereich ÖV kaum gewählt wird, die Bereiche Fuß und Rad hingegen sehr häufig. Bei der Erklärung des Wahlverhaltens zeigt sich auch, dass sozioökonomische Merkmale wie das Alter eine ebenso wichtige Rolle spielen wie die Nutzungsmöglichkeit (Pkw-Hauptnutzer, Pkw-Mitnutzer oder sonstiges), der Wochentag oder die Entfernung zur nächstgelegenen Haltestelle. Darüber hinaus wird das Mobilitätsverhalten durch den persönlichen Beitrag zum Haushaltseinkommen, aber auch durch subjektive Merkmale wie bequem, kostengünstig, kürzeste Zeitdauer und Gewohnheit beeinflusst. Die Wahl der Bereiche Fuß bzw. Rad, im Speziellen, hängen zudem von den Witterungsbedingungen ab, die Wahl für öffentliche Verkehrsmittel vom Besitz einer Netzkarte, Bahncard und ganz entscheidend von den Zugangsmöglichkeiten zu öffentlichen Verkehrsmitteln.

Im Hinblick auf eine Verbesserung der Modellerklärung wäre es von Interesse weiterführende Untersuchungen mit Hilfe kategorienspezifischer Variablen durchzuführen. Die durchgeführten Untersuchungen basieren auf der Betrachtung aller Kategorien. Für den Anwender ist es jedoch von Bedeutung wie sich die Ergebnisse verändern, wenn nur die Kategorien, die einen Einfluss auf die Verkehrsmittelwahl aufweisen, in das Modell aufgenommen werden.

Um eine zufällige Struktur im Modell adäquat zu berücksichtigen, wäre es von Interesse zufällige Effekte in das Modell aufzunehmen. In Kapitel 4 wurde bereits eine mögliche Schätzmethode, die penalisierte Quasi-Likelihood-Methode, vorgestellt. Die Praxis zeigt jedoch, dass diese Schätzmethode nicht immer geeignet ist. In diesem Zusammenhang ist es sicher sinnvoll die Untersuchung mit Markov Chain Monte Carlo Methoden fortzusetzen.

A Numerische Bestimmung der ML-Schätzung

Zur Bestimmung der Maximum-Likelihood-Schätzungen ist das Gleichungssystem

$$\mathbf{s}(\hat{\boldsymbol{\beta}}) = \frac{\partial l(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} = 0$$

zu lösen. Likelihoodgleichungen sind häufig nicht linear in $\boldsymbol{\beta}$, sodass es keine analytische Lösung für $\hat{\boldsymbol{\beta}}$ gibt. Die Gleichungen können numerisch mit Hilfe eines Iterationsalgorithmus bestimmt werden. Eine Methode ist das im folgenden kurz skizzierte *Newton-Raphson* Verfahren.

Initialisierung:

Setze einen Anfangswert $\hat{\boldsymbol{\beta}}^{(0)}$, beispielsweise $\hat{\boldsymbol{\beta}}^{(0)} = 0$.

Schritt 1: Taylorentwicklung

Linearisiere die Scorefunktion durch Taylorentwicklung an der Stelle $\hat{\boldsymbol{\beta}}^{(k)}$. In linearer Näherung gilt

$$\mathbf{s}(\boldsymbol{\beta}) \approx \mathbf{s}_l(\boldsymbol{\beta}) = \mathbf{s}(\hat{\boldsymbol{\beta}}^{(k)}) + \frac{\partial \mathbf{s}(\hat{\boldsymbol{\beta}}^{(k)})}{\partial \boldsymbol{\beta}} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}^{(k)}).$$

Schritt 2: Lösung der linearisierten Form

Löse das Gleichungssystem $\mathbf{s}_l(\boldsymbol{\beta}) = 0$. Daraus ergibt sich mit der Hessematrix $\mathbf{H}(\boldsymbol{\beta}) = \partial \mathbf{s} / \partial \boldsymbol{\beta} = \partial^2 l(\boldsymbol{\beta}) / \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T$

$$\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}^{(k)} - \left(\frac{\partial^2 l(\hat{\boldsymbol{\beta}}^{(k)})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right)^{-1} \mathbf{s}(\hat{\boldsymbol{\beta}}^{(k)}).$$

Mit $\mathbf{H}(\boldsymbol{\beta})$ ergibt sich als neuer Schätzer

$$\hat{\boldsymbol{\beta}}^{(k+1)} = \hat{\boldsymbol{\beta}}^{(k)} - \mathbf{H}(\hat{\boldsymbol{\beta}}^{(k)})^{-1} \mathbf{s}(\hat{\boldsymbol{\beta}}^{(k)})$$

Die Matrix $-\mathbf{H}(\hat{\boldsymbol{\beta}}^{(k)})$ wird häufig als *beobachtete Informationsmatrix* bezeichnet.

Schritt 3: Überprüfung der Veränderung

Wenn

$$\frac{\|\hat{\boldsymbol{\beta}}^{(k+1)} - \hat{\boldsymbol{\beta}}^{(k)}\|}{\|\hat{\boldsymbol{\beta}}^{(k)}\|}$$

kleiner als eine vorgegebene Schranke ε ist (mit $\|\cdot\|$: euklidische Norm), setze $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^{(k+1)}$. Andernfalls erhöhe k zu $k + 1$ und gehe zu Schritt 1.

B S-Plus Funktionen

B.1 S-Plus Funktion zur explorativen Datenanalyse von Haupteffekten

```

haupteffekt <- function(y1, y2, y3, y4, ref1 = "REF.KAT.IV", ref2 = "REF.KAT.IV",
  ref3 = "REF.KAT.IV", z = "VERKEHRSSBEREICH", x1, x2, x3, label)
{
  haupteinfluss <- function(y = Response, covar, label = "", conf = T) {
    temp <- table(y, covar)
    n <- apply(temp, 2, sum)
    logit <- log((temp[2, ] + 0.5)/(temp[1, ] + 0.5))
    lab <- dimnames(temp)[[2]]
    lab1 <- dimnames(temp)[[1]]
    var.hat <- ((temp[2, ] + 0.5) * (temp[1, ] + 0.5))/(n)
    upper <- logit + (qnorm(0.95) * sqrt(1/var.hat))
    lower <- logit - (qnorm(0.95) * sqrt(1/var.hat))
    if(conf == T) {
      ry <- range(c(upper, lower))
      plot(logit, ylim = ry, axes = F, pch = 3, xlab = "", ylab = "empiri-
        sche Logits", type = "b", lty = 1, col = 1, lwd = 2)
      axis(2)
      axis(1, at = 1:length(logit), labels = lab)
      lines(1:length(logit), upper, lwd = 2, lty = 2, col = 1)
      lines(1:length(logit), lower, lwd = 2, lty = 2, col = 1)
      par(new = T)
      plot(logit, ylim = ry, axes = F, pch = 3, xlab = "", ylab =
        "empirische Logits", type = "n", lty = 1, col = 1, lwd = 2)
    }
    else {
      plot(logit, axes = F, pch = 3, xlab = "", ylab = "empirische Logits",
        type = "b", lty = 1, col = 1, lwd = 2)
      axis(2)
      axis(1, at = 1:length(logit), labels = lab)
    }
    cat(paste("Referenzkategorie und Verkehrsmittel", lab1[2]))
    cat("\n")
    mat <- rbind(round(logit, digit = 2), n)
    dimnames(mat)[[1]] <- c("emp. Logits", "n")
    print(mat)
    cat("\n")
    invisible()
  }
  par(mfrow = c(3, 1))
  cat("\n")
  cat(paste("Haupteffekte f\374r", label))
  cat("\n")

```

```

cat(paste("=====\n"))
cat("\n")
cat(paste("Als Referenzkategorie wurde", y1, "gew\344hlt\n"))
cat("\n")
m1 <- mobilitaetsdaten [mobilitaetsdaten [,z]==y1 | mobilitaetsdaten[, z]==y2,]
attach(m1)
h1 <- haupteinfluss(ref1, x1, label, conf = T)
title(paste("Haupteffekte des Regressors", label, "mit Referenzkategorie", y1, "und
Verkehrsmittel", y2))
detach("m1")
m2 <- mobilitaetsdaten [mobilitaetsdaten [,z]==y1 | mobilitaetsdaten[, z]==y4,]
attach(m2)
h2 <- haupteinfluss(ref2, x2, label, conf = T)
title(paste("Referenzkategorie", y1, "und Verkehrsmittel", y4))
detach("m2")
m3 <- mobilitaetsdaten [mobilitaetsdaten [,z]==y1 | mobilitaetsdaten[, z]==y3,]
attach(m3)
h3 <- haupteinfluss(ref3, x3, label, conf = T)
title(paste("Referenzkategorie", y1, "und Verkehrsmittel", y3))
detach("m3")
invisible()
}

```

B.2 S-Plus Funktion zum Berechnen der relativen Häufigkeiten

```

prozent <- function(y = VERKEHRSSBEREICH, x = covar, label)
}
temp <- table(y, x)
n <- apply(temp, 2, sum)
temp <- table(x, y)
cat(paste("\nBeobachtungen:", label, "\n"))
cat(paste("—————\n"))
print(temp)
cat("\n")
p <- round(temp[, ]/n, digit = 3) * 100
cat(paste("Anteil in Prozent:\n"))
cat(paste("—————\n"))
print(p)
cat("\n")
invisible()
}

```

B.3 S-Plus Funktion zur explorativen Datenanalyse von Interaktionen

```

interaktion <- function(y1, y2, y3, y4, ref1 = "REF.KAT.IV", ref2 = "REF.KAT.IV",
  ref3 = "REF.KAT.IV", z = "VERKEHRSBEREICH", x11, x12, x13, x21, x22, x23,
  label1, label2)
{
  interaktion.plot <- function(y = Response, x1, x2, label1 = "",
    label2 = "", conf = T) {
    temp <- table(x1, x2, y)
    lab1 <- dimnames(temp)[[3]]
    d.temp <- dim(temp)
    if(d.temp[1] != length(levels(x1)) | d.temp[2] != length(levels(x2)) |
      d.temp[3] != 2) {
      print(temp)
      stop("leere Zellen")
    }
    n <- temp[, , 1] + temp[, , 2]
    temp1 <- as.character(interaction(label1, levels(x1)))
    temp2 <- as.character(interaction(label2, levels(x2)))
    dimnames(n) <- list(temp1, temp2)
    logit <- log((temp[, , 2] + 0.5)/(temp[, , 1] + 0.5))
    var.hat <- ((temp[, , 2] + 0.5) * (temp[, , 1] + 0.5))/n
    var.hat[var.hat == Inf] <- NA
    upper <- logit + (qnorm(0.95) * sqrt(1/var.hat))
    lower <- logit - (qnorm(0.95) * sqrt(1/var.hat))
    dimnames(logit) <- dimnames(n)
    r <- range(c(upper, lower, -2, 2), na.rm = T)
    lab <- levels(x2)
    if(conf == F) {
      r <- range(logit)
      plot(as.vector(logit[1, ]), axes = F, pch = 3, xlab = "", ylab
        = "empirische Logits", sub = label2, ylim = r, type = "n")
      axis(2)
      axis(1, at = 1:length(logit[1, ]), labels = lab)
      for(i in 1:length(levels(x1))) {
        par(new = T)
        plot(as.vector(logit[i, ]), axes = F, xlab = "", ylab = "", ylim = r,
          lty = i + 1, lwd = 2, col = i + 1, type = "l")
        text(1 + 0.25, logit[i, 1], paste(levels(x1)[i]), col = i + 1)
      }
    }
    else {
      plot(as.vector(logit[1, ]), axes = F, pch = 3, xlab = , ylab
        = "empirische Logits", sub = label2, ylim = r, type = "n")
      axis(2)
      axis(1, at = 1:length(logit[1, ]), labels = lab)
      for(i in 1:length(levels(x1))) {

```

```

        par(new = T)
        plot(as.vector(logit[i, ]), axes = F, xlab = "", ylab = "", ylim = r,
             lty = i + 1, lwd = 2, col = i + 1, type = "l")
        lines(1:length(logit[i, ]), as.vector(upper[i, ]), col = i + 1, lty = 2)
        lines(1:length(logit[i, ]), as.vector(lower[i, ]), col = i + 1, lty = 2)
        text(1 + 0.25, logit[i, 1], paste(levels(x1)[i]), col = i + 1)
    }
}
cat("empirische Logits der Referenzkategorie und dem Verkehrsmittel",
    lab1[2], "\n")
cat("\n")
print(round(logit, digits = 2))
cat("\n")
cat("dazugeh\366rige Zellengr\366\337en\n")
cat("\n")
print(n)
cat("\n")
invisible()
}
par(mfrow=c(3,1))
cat("\n")
cat(paste("Interaktionen der Regressoren", label1, "und", label2))
cat("\n")
cat(paste("=====\n"))
cat("\n")
cat(paste("Als Referenzkategorie wurde", y1, "gew\344hlt\n"))
cat("\n")
m1 <- mobilitaetsdaten [mobilitaetsdaten [,z]==y1 | mobilitaetsdaten[, z]==y2,]
attach(m1)
i1 <- interaktion.plot(ref1, x11, x21, label1, label2, conf = T)
title(paste("Interaktionseffekte der Regressoren ", label1, "und",
            label2, "mit Referenzkategorie\n", y1, "und Verkehrsmittel", y2))
detach("m1")
m2 <- mobilitaetsdaten [mobilitaetsdaten [,z]==y1 | mobilitaetsdaten[, z]==y4,]
attach(m2)
i2 <- interaktion.plot (ref2, x12, x22, label1, label2, conf = T)
title(paste("Referenzkategorie\n", y1, "und Verkehrsmittel", y4))
detach("m2")
m3 <- mobilitaetsdaten [mobilitaetsdaten [,z]==y1 | mobilitaetsdaten[, z]==y3,]
attach(m3)
i3 <- interaktion.plot(ref3, x13, x23, label1, label2, conf = T)
title(paste("Referenzkategorie\n", y1, "und Verkehrsmittel", y3))
detach("m3")
invisible()
}

```

B.4 S-Plus Funktion zum Vergleich empirischer und geschätzter Logits

```

vergleich.plot <- function(y1, y2, y3, y4, ref1 = "REF.KAT.IV", ref2 = "REF.KAT.IV",
  ref3 = "REF.KAT.IV", z = "VERKEHRSBEREICH", x1, x2, x3, label, r)
{
  vergleich <- function(y = Response, covar, label, conf = T)
  {
    temp <- table(y, covar)
    lab <- dimnames(temp)[[1]]
    assign("lab", lab, frame = 0)
    n <- apply(temp, 2, sum)
    logit <- log((temp[2, ] + 0.5)/(temp[1, ] + 0.5))
    assign("logit", logit, frame = 0)
    mat <- rbind(round(logit, digit = 2), n)
    dimnames(mat)[[1]] <- c("emp. Logits", "Zellengr\B66\B37en")
    cat(paste("Referenzkategorie und Verkehrsmittel", lab[2]))
    cat("\n")
    print(mat)
    cat("\n")
    i <- rep(0, length(logit)-1)
    for (k in 2:length(logit)){
      i[k-1] <- (mat[2*k-1]-mat[1])
    }
    cat("\n")
    invisible()
    return(i)
  }
  cat("\n")
  cat("empirische Logits f\374r", label))
  cat("\n")
  cat(paste("=====\n"))
  cat("\n")
  cat(paste("Als Referenzkategorie wurde", y1, "gew\344hlt\n"))
  cat("\n")
  m1 <- mobilitaetsdaten[mobilitaetsdaten[,z]==y1 | mobilitaetsdaten[,z]==y2,]
  attach(m1)
  v1 <- vergleich(ref1, x1, label, conf = T)
  detach("m1")
  m2 <- mobilitaetsdaten[mobilitaet[,z]==y1 | mobilitaetsdaten[,z]==y4,]
  attach(m2)
  v2 <- vergleich(ref2, x2, label, conf = T)
  detach("m2")
  m3 <- mobilitaet[mobilitaetsdaten[,z]==y1 | mobilitaetsdaten[,z]==y3,]
  attach(m3)
  v3 <- vergleich(ref3, x3, label, conf = T)
  detach("m3")
  cat(paste("Differenzen der empirischen Logits\n"))
}

```

```

cat(paste("-----\n"))
cat("\n")
cat(paste("Referenzkategorie", y1, "und Verkehrsmittel", y2, " "))
print(v1)
cat(paste("Referenzkategorie", y1, "und Verkehrsmittel", y4, " "))
print(v2)
cat(paste("Referenzkategorie", y1, "und Verkehrsmittel", y3, " "))
print(v3)
cat("\n")
cat(paste("geschätzte Parameter \n"))
cat(paste("-----\n"))
print(round(coef(r), digit = 2))
a <- cbind(rep(1, (length(logit)-1)), rep(2, (length(logit)-1)),
           rep(3, (length(logit)-1)))
b <- c(v1, v2, v3)
e <- c()
for (d in 1:(length(logit)-1)) {
  e <- c(e, round(coef(r)[(d*3)+1], digit = 2))
}
e <- c(e)
for (d in 1:(length(logit)-1)) {
  e <- c(e, round(coef(r)[(d*3)+2], digit = 2))
}
e <- c(e)
for (d in 1:(length(logit)-1)) {
  e <- c(e, round(coef(r)[(d*3)+3], digit = 2))
}
x.achsenbeschriftung <- cbind(rep(c(y2, y4, y3), length(logit)-1))
plot(a, b, axes=F, xlab="Verkehrsmittel", ylab="emp. und gesch\344tzter Logits",
     pch="x", lty=2, col=2, lwd=2, ylim = c(min(v1, v2, v3,
        coef(r)[4:(length(logit)*3)]), max(v1, v2, v3, coef(r)[4:(length(logit)*3)])))
axis(2)
axis(1, at=1:length(a), labels=x.achsenbeschriftung)
title(paste("Vergleich der emp. Logits mit den gesch\344tzten Parametern"))
par(new=T)
plot(a, e, axes=F, xlab=" ", ylab=" ", pch="o", lty=2, col=3, lwd=2,
     ylim = c(min(v1, v2, v3, coef(r)[4:(length(logit)*3)]),
        max(v1, v2, v3, coef(r)[4:(length(logit)*3)])))
leg <- c("Diff. emp. Logits", "gesch\344tzte Parameter")
legend(c(1.1, 1.9), c(max(e - 0.25),max(e)), leg, col=c(1:2), pch="x\o")
}

```

C Datensatzbeschreibung

MB.TAG	Mobilitätstag (4.00 bis 3.59 Uhr)
	Value Label
	1 Mo
	2 Di
	3 Mi
	4 Do
	5 Fr
	6 Sa
	7 So
HH.GROESSE	Haushaltsgröße
	Value Label
	1 eine Person
	2 zwei Personen
	3 drei Personen
	4 vier und mehr Personen
HH.EINK.NEU	Haushaltseinkommen zusammengefasst
	Value Label
	2 niedriges HH-Einkommen (bis 3000 DM)
	3 mittleres HH-Einkommen (3000 bis 5000 DM)
	4 hohes HH-Einkommen (> 5000 DM)
HSSCHIEN	Anbindung an schienengeb. ÖV
	Value Label
	0 nein
	1 ja
S.KM.OEV.MIN	Entfernung zur nächsten Haltestelle (geskort)
ZUFR.ANB.OEV	Zufriedenheit mit der Anbindung an ÖV
	Value Label
	1 sehr zufrieden
	2 insgesamt zufrieden
	3 Verbesserung gewünscht

WETT.GUT	Gutes Wetter Value Label
	0 nein 1 ja
S.ENTF.KM	Entfernung (gescort)
STR.TYP.NEU	Streckentyp Value Label
	1 innerorts 2 inner- und/oder außerorts
WEG.MIT	Weg alleine oder mit mehreren Value Label
	0 mit mehreren 1 alleine
VORHER	Vorher damit unterwegs Value Label
	0 nein 1 ja
GEWOHN	Gewohnheit/bewährt Value Label
	0 nein 1 ja
OHNE.ALT	Keine andere Möglichkeit Value Label
	0 nein 1 ja
GUT.ENTF	Paßt zur Entfernung Value Label
	0 nein 1 ja

KURZ.T	Geringster Zeitaufwand Value Label
	0 nein 1 ja
BEQUEM	Bequem Value Label
	0 nein 1 ja
FLEXIBEL	Flexibel Value Label
	0 nein 1 ja
KOSTEN	Kostengünstig Value Label
	0 nein 1 ja
SPORT	Als Sport/zur Bewegung Value Label
	0 nein 1 ja
S.DAUERMIN	Dauer des Weges [min] (geskort)
EINDRUCK	Eindruck Value Label
	1 angenehm 2 durchschnittlich 3 unangenehm
S.ALTER.N	Alter (normiert und geskort)

NUTZUNG	Pkw-Nutzung Value Label
	1 Pkw-Hauptnutzer
	2 Pkw-Mitnutzer
	3 sonstige (hierunter fallen Nichtnutzer, kein Führerschein)
RAD.VERF	Rad-Verfügung Value Label
	0 nein
	1 ja
PERS.EINK.NEU	Persönliches Einkommen zusammengefasst Value Label
	1 niedriges Einkommen (bis 2000 DM)
	2 mittleres Einkommen (2000 bis 4000 DM)
	3 hohes Einkommen (> 4000 DM)
NETZKARTE	Besitz einer Netzkarte Value Label
	0 nein
	1 ja
BAHNCARD	Besitz einer Bahncard Value Label
	0 nein
	1 ja
EINPKW.N	Negative Einstellung zum Individualverkehr Value Label
	0 nein
	1 ja

D Ergebnisse der explorativen Datenanalyse

D.1 Plots einiger Haupteffekte

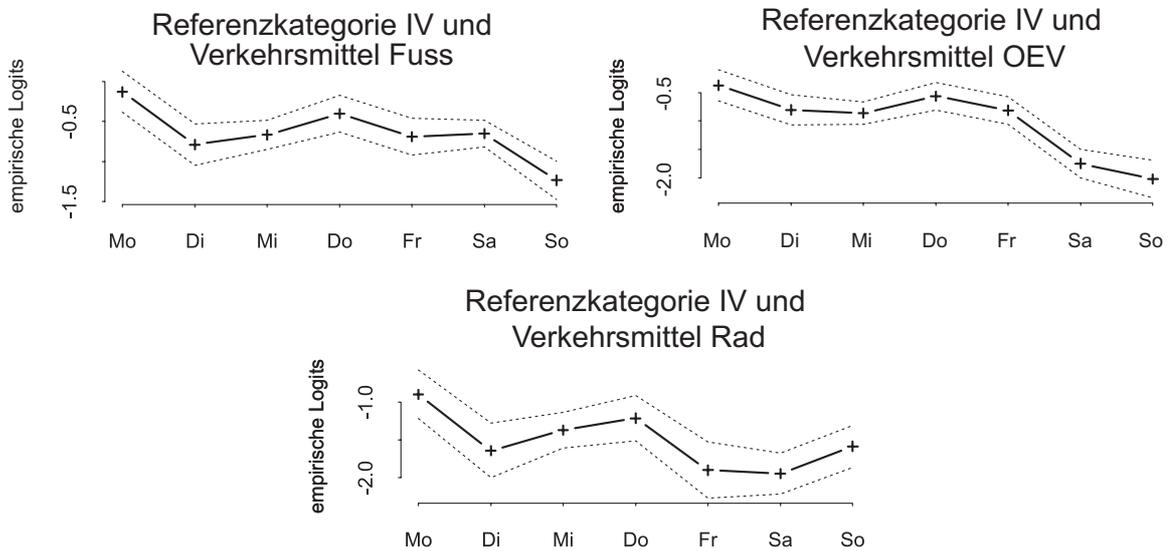


Abbildung 40: Haupteffekt des Regressors Mobilitätstag

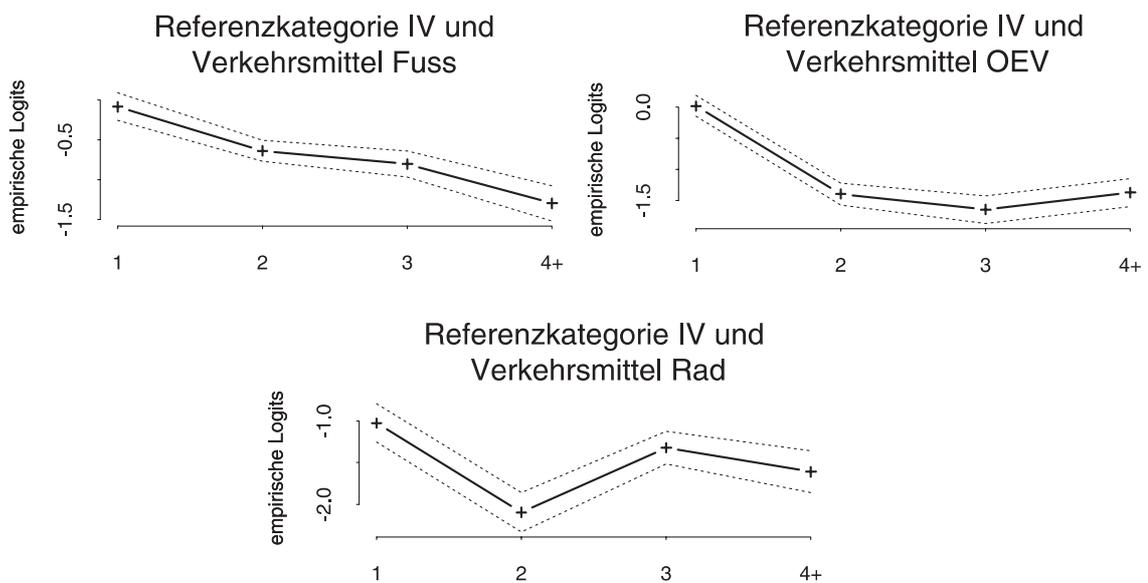


Abbildung 41: Haupteffekt des Regressors Haushaltsgröße

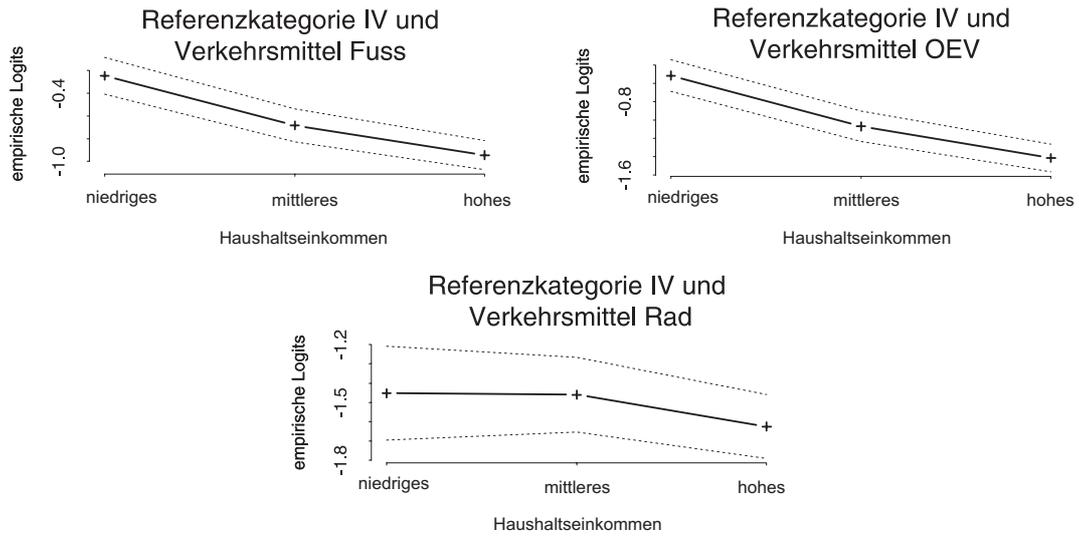


Abbildung 42: Haupteffekt des Regressors Haushaltseinkommen

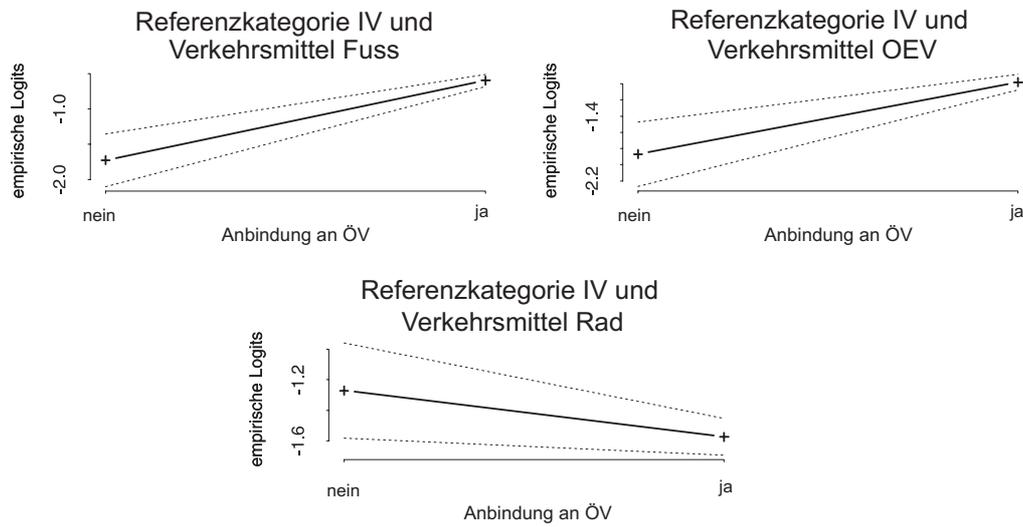


Abbildung 43: Haupteffekt des Regressors Anbindung an schienengeb. ÖV

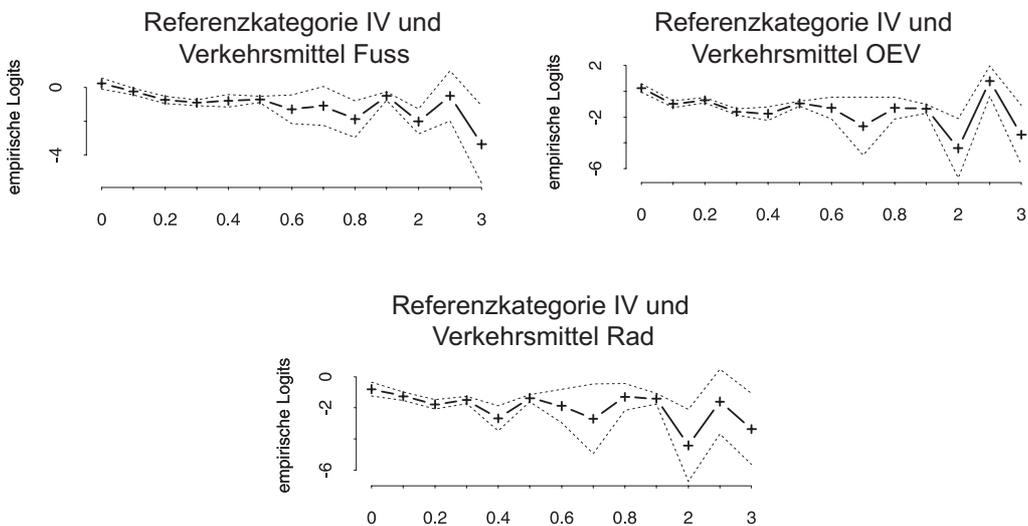


Abbildung 44: Haupteffekt des Regressors Entfernung zur nächstgelegenen Haltestelle

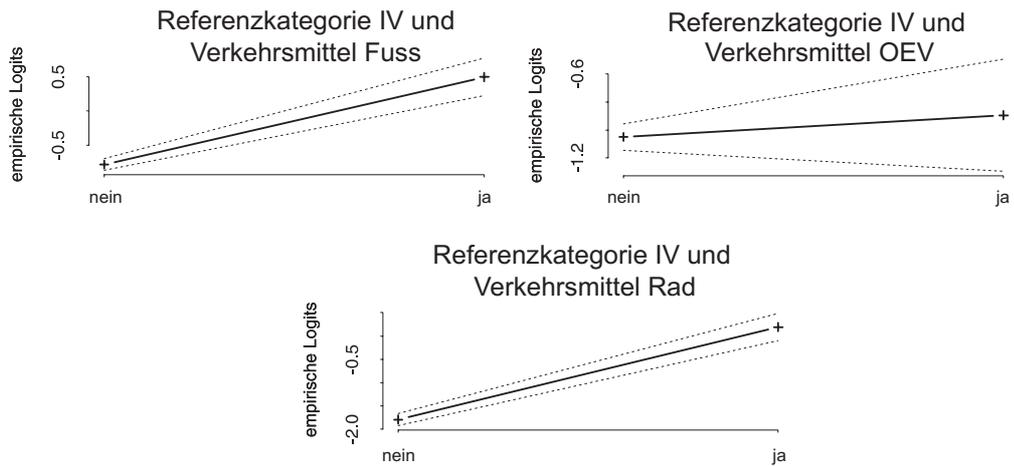


Abbildung 45: Haupteffekt des Regressors gutes Wetter

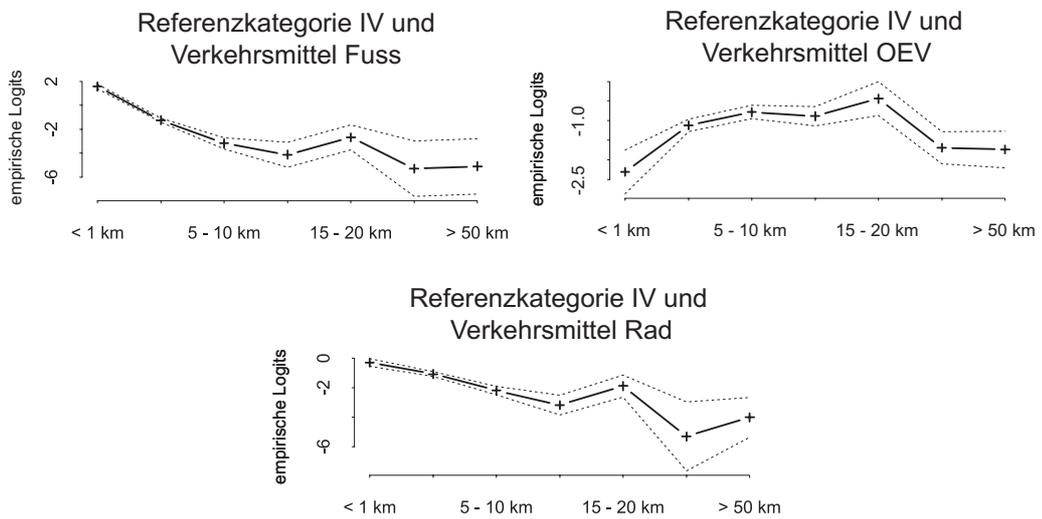


Abbildung 46: Haupteffekt des Regressors Entfernung [km]

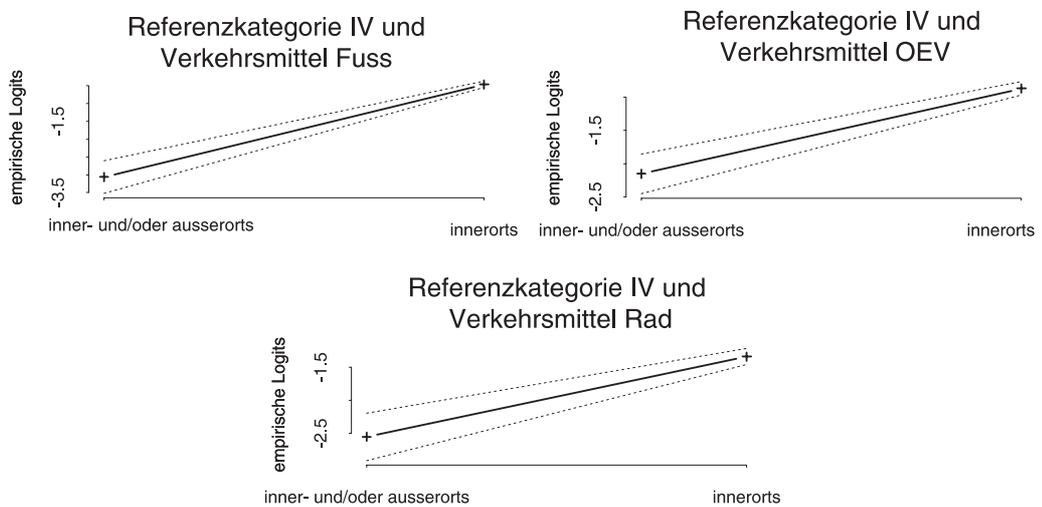


Abbildung 47: Haupteffekt des Regressors Streckentyp

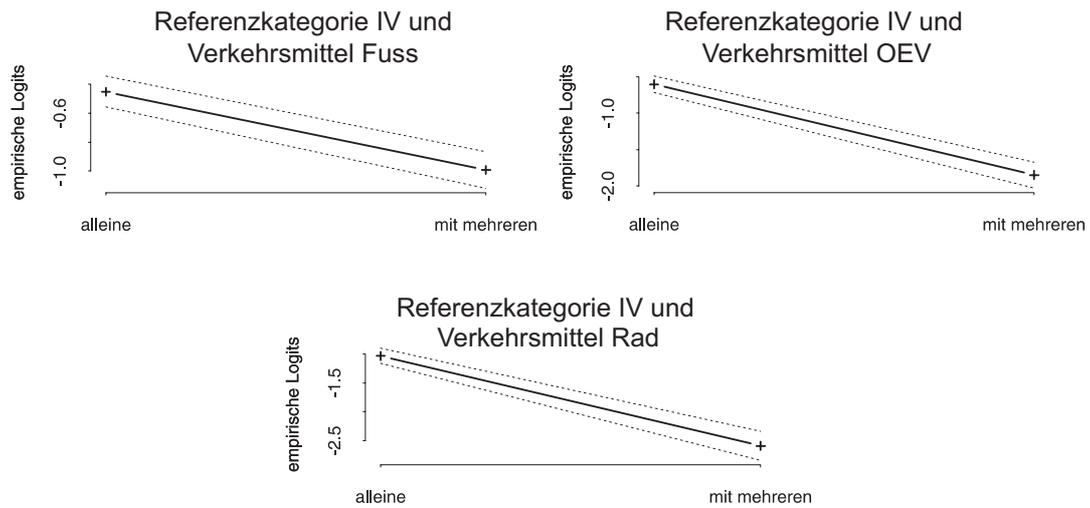


Abbildung 48: Haupteffekt des Regressors Weg alleine oder mit mehreren zurückgelegt

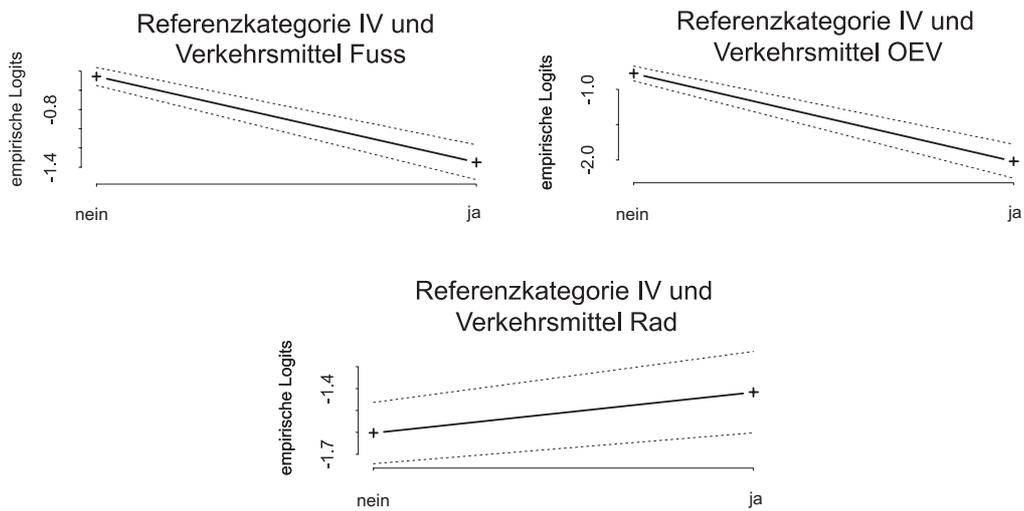


Abbildung 49: Haupteffekt des Regressors vorher mit diesem Verkehrsmittel unterwegs

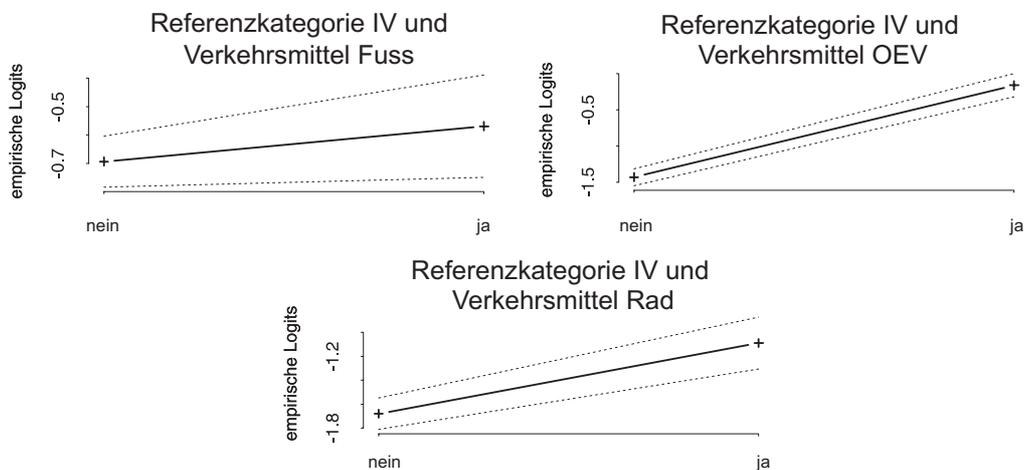


Abbildung 50: Haupteffekt Gewohnheit

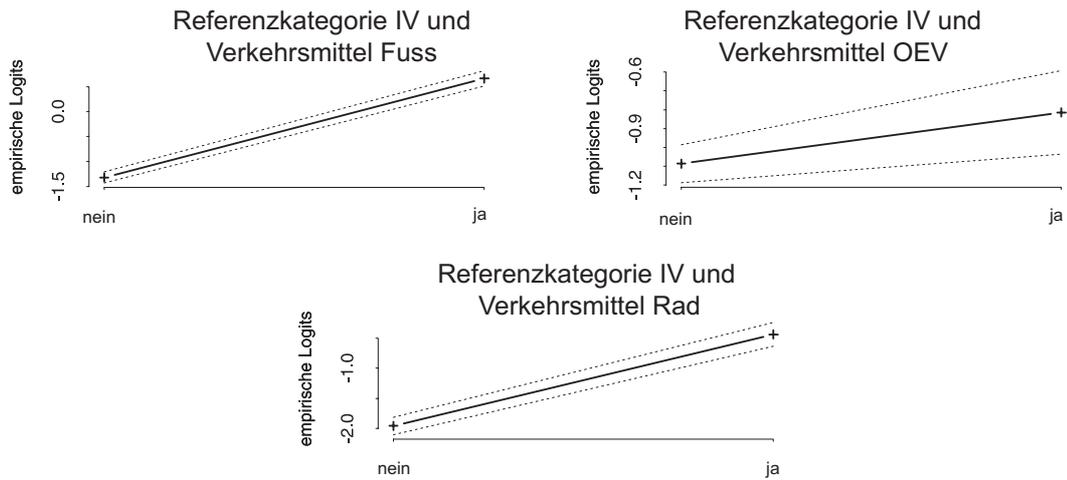


Abbildung 51: Haupteffekt des Regressors gute Entfernung

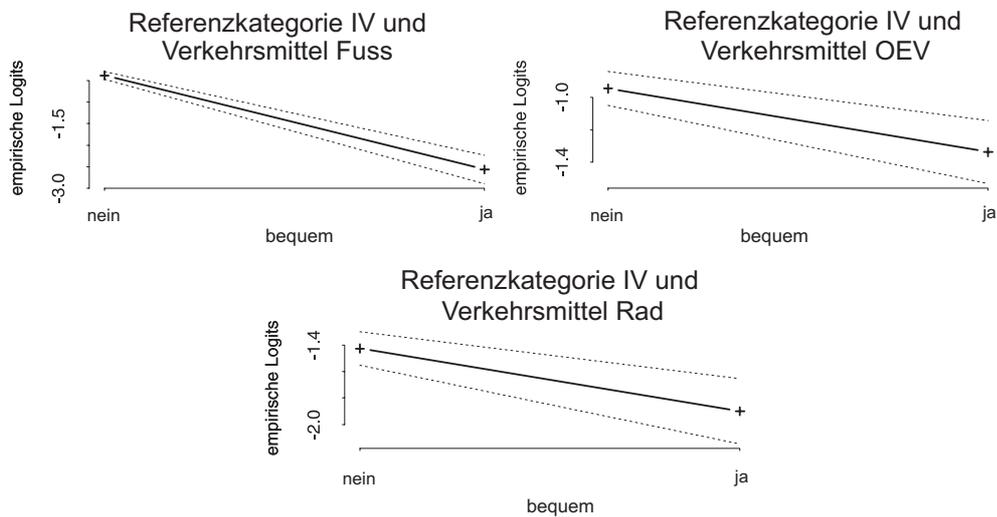


Abbildung 52: Haupteffekt des Regressors Bequemlichkeit

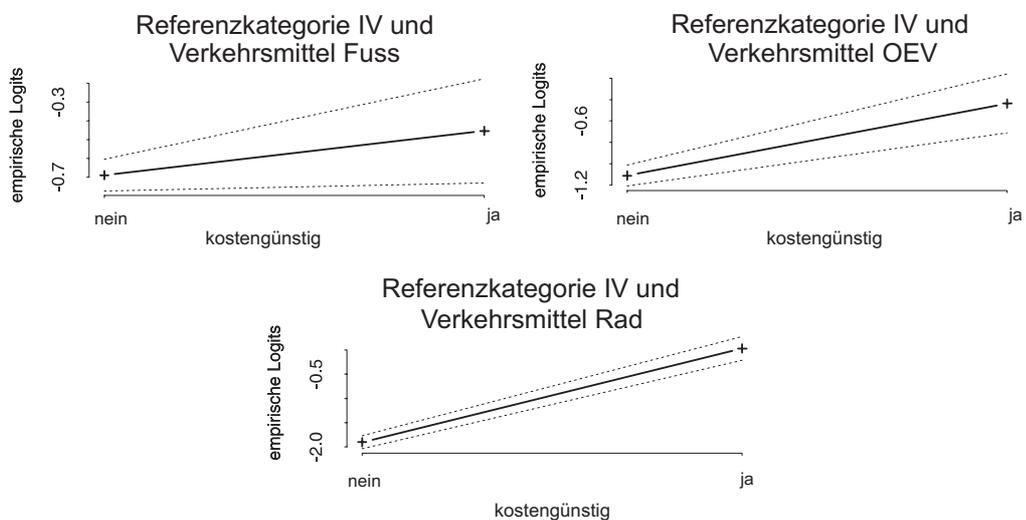


Abbildung 53: Haupteffekt des Regressors kostengünstig

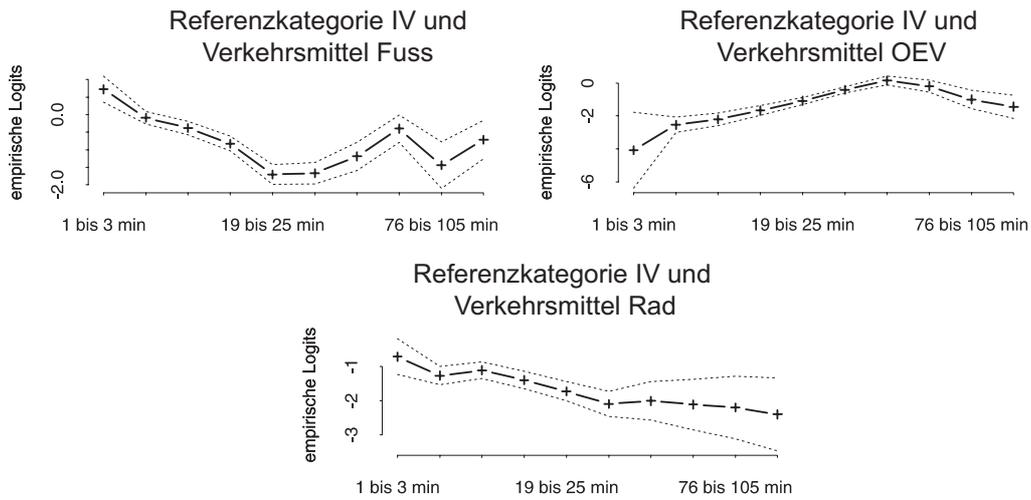


Abbildung 54: Haupteffekt des Regressors Wegedauer

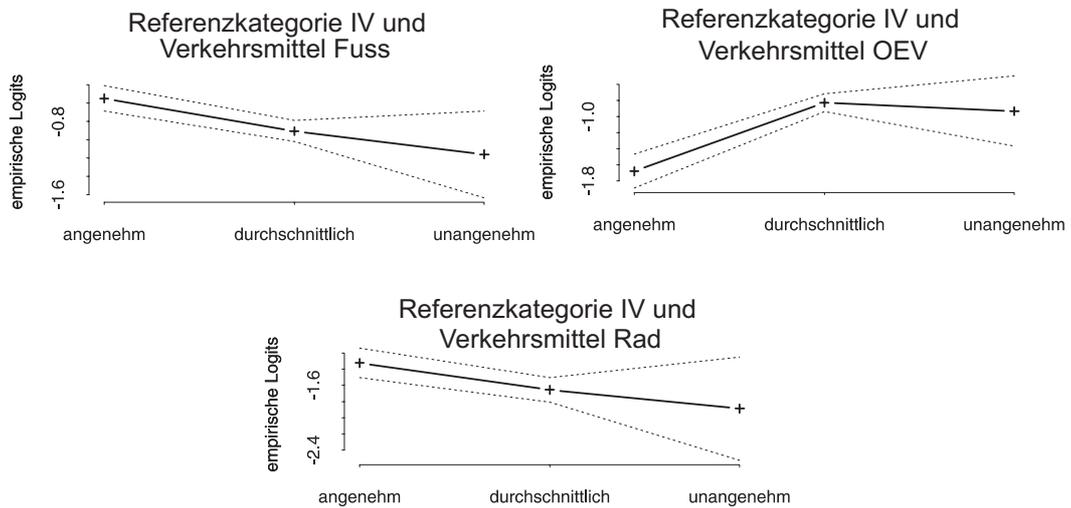


Abbildung 55: Haupteffekt des Regressors Eindruck des Weges / der Fahrt

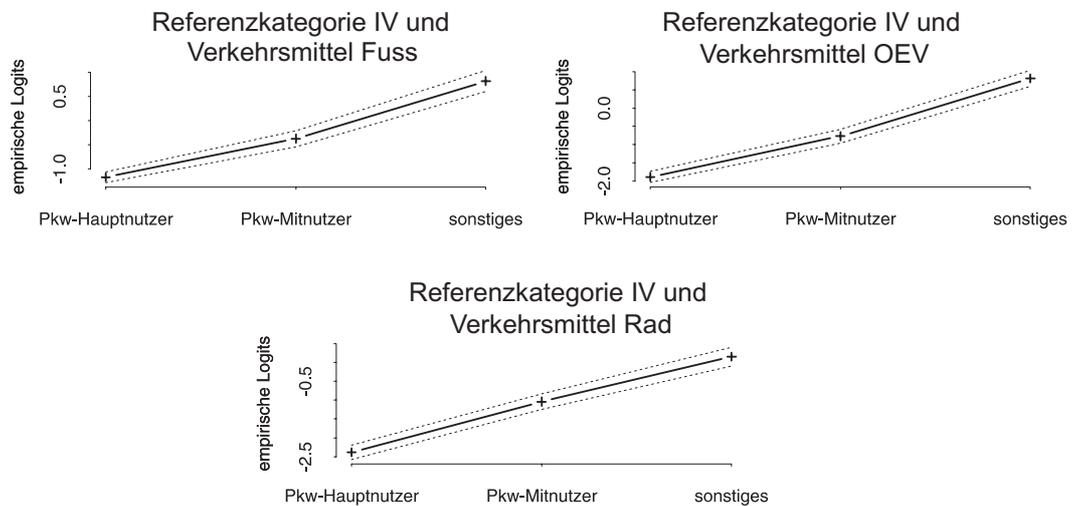


Abbildung 56: Haupteffekt des Regressors Nutzung

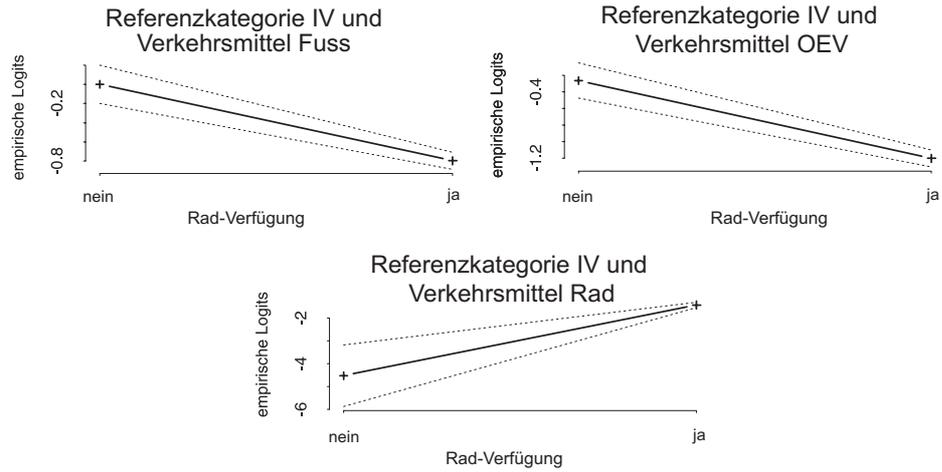


Abbildung 57: Haupteffekt des Regressors Rad-Verfügung

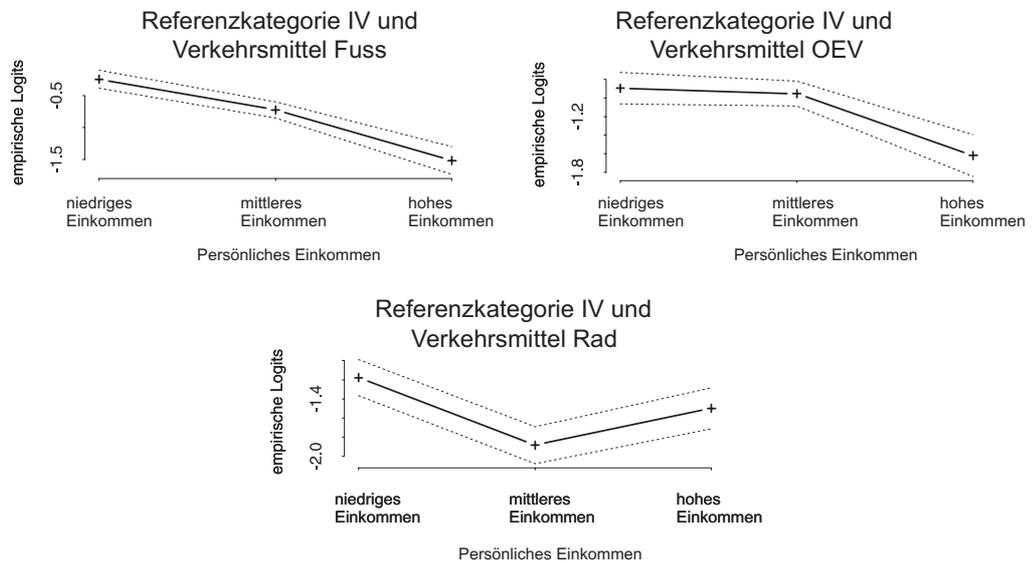


Abbildung 58: Haupteffekt des Regressors persönliches Einkommen

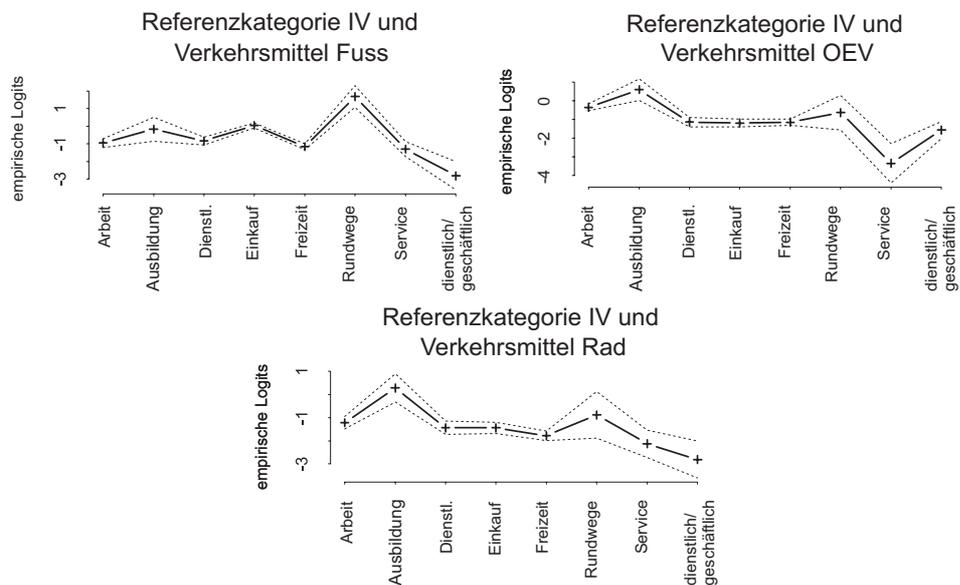


Abbildung 59: Haupteffekt des Regressors Wegzweck

D.2 Interaktionsplots

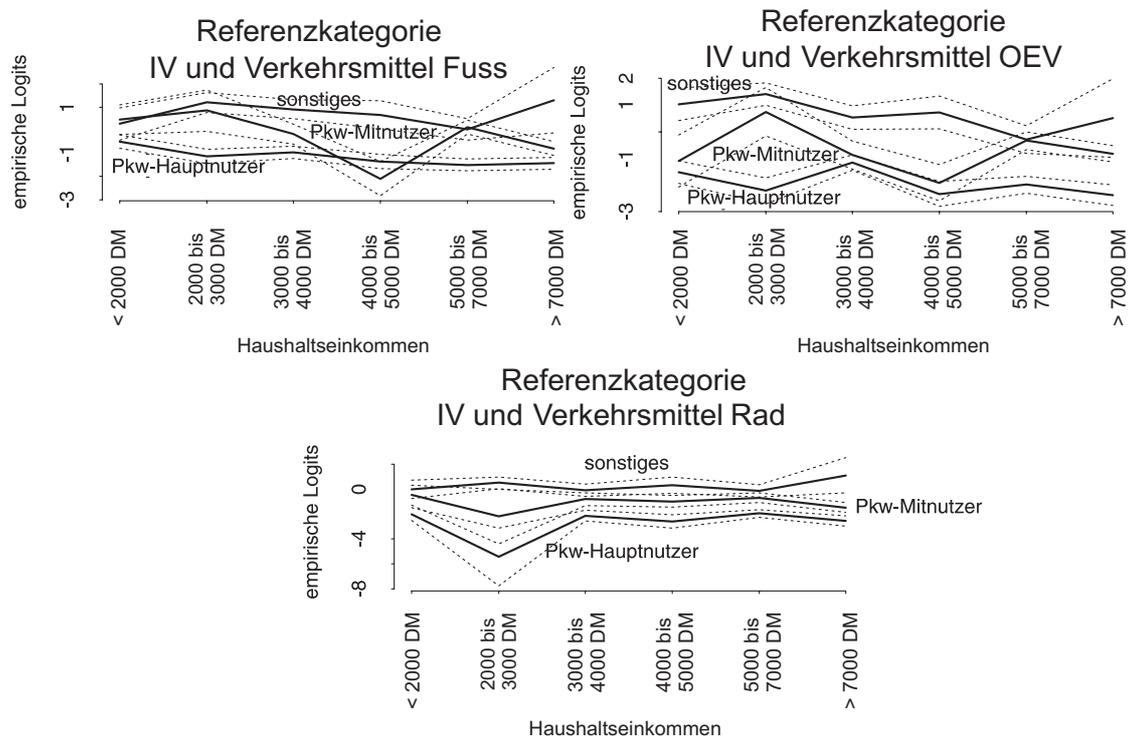


Abbildung 60: Interaktion der Regressoren Haushaltseinkommen und Pkw-Nutzung

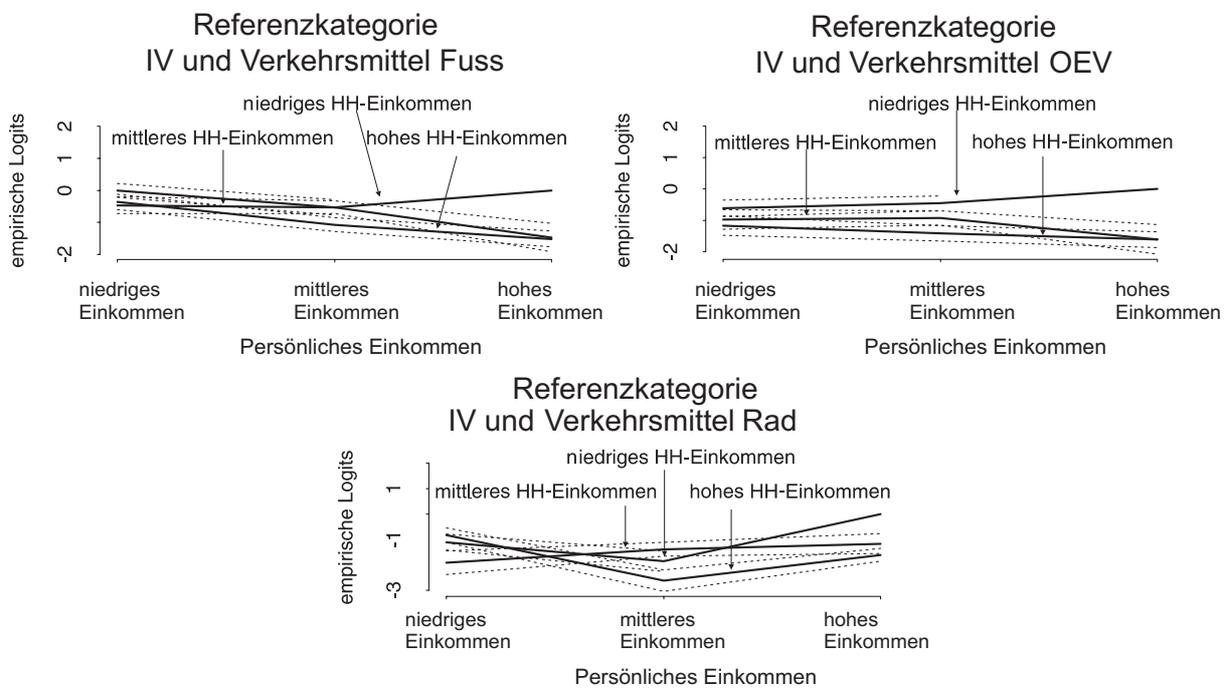


Abbildung 61: Interaktion der Regressoren Haushalts- und persönliches Einkommen

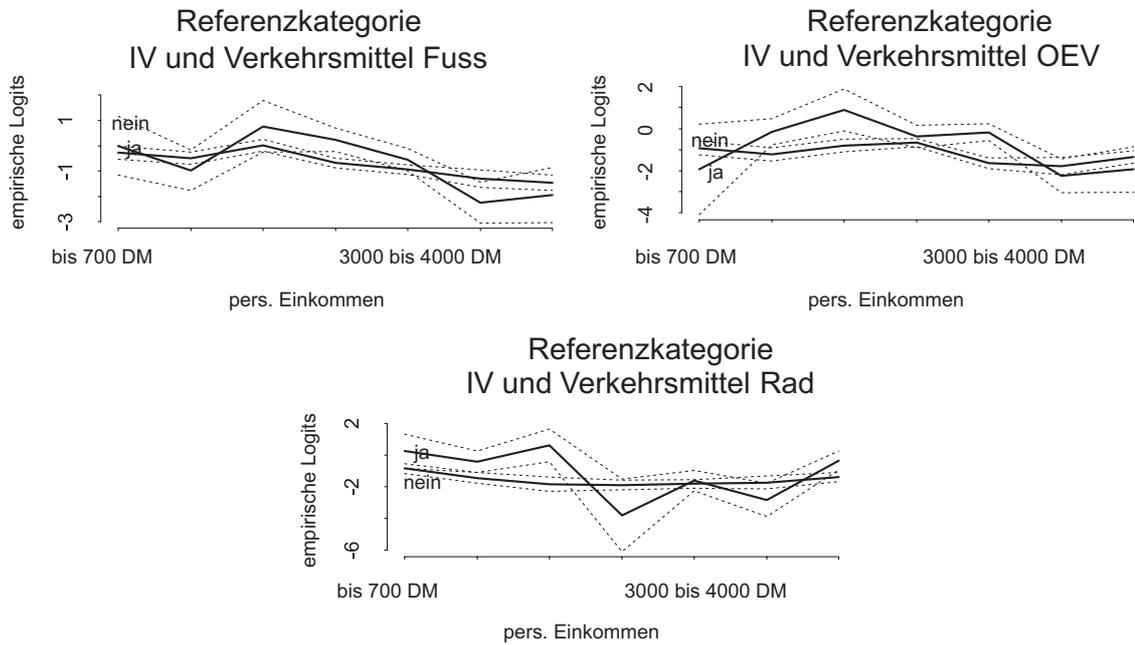


Abbildung 62: Interaktion der Regressoren Bahncard und persönliches Einkommen

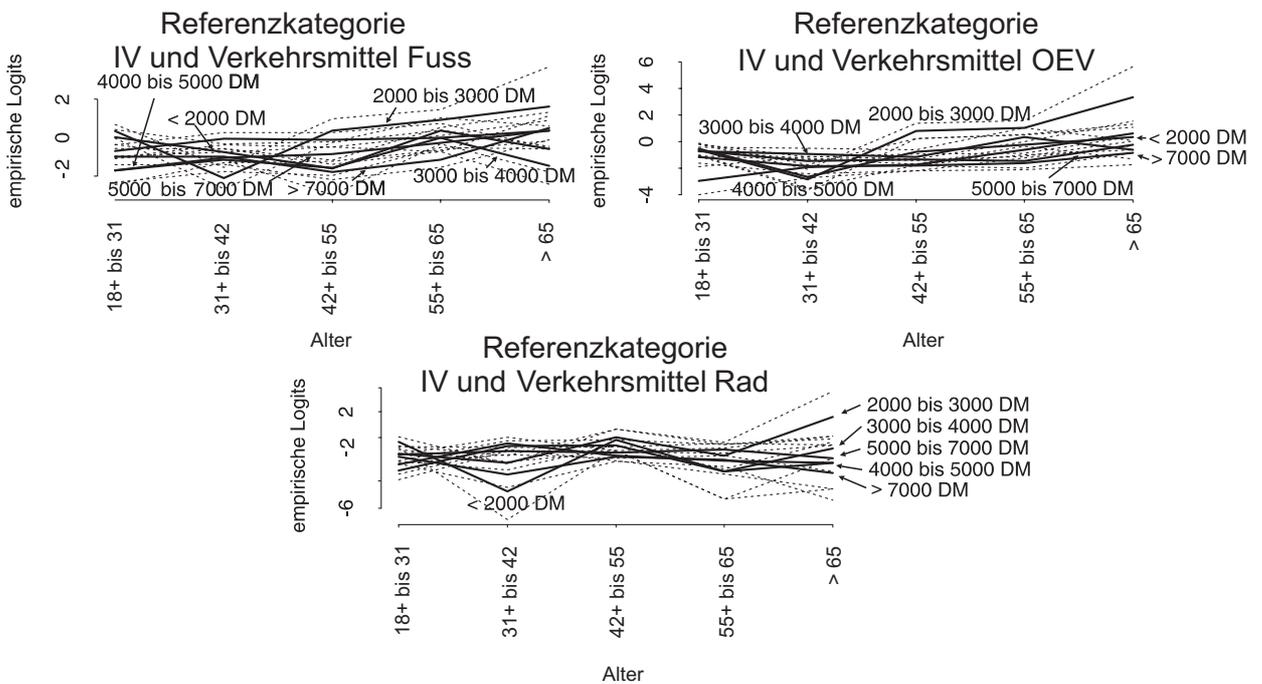


Abbildung 63: Interaktion der Regressoren Alter und Haushaltseinkommen

Literaturverzeichnis

- [1] Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transaction on Automatic Control* 19, 719 - 723.
- [2] Anscombe, F. J., 1956. On estimating binomial response relations. *Biometrika* 43, 461-464.
- [3] Axhausen, K. W., 1995. *Travel Diaries: An Annotated Catalogue* (2. Auflage). Innsbruck, Leopold-Franzens-Universität Innsbruck.
- [4] Barndorff-Nielsen, O. E., Cox, D. R., 1989. *Asymptotic Techniques for Use in Statistics*, London, Chapman and Hall.
- [5] Bickel, P. J., Doksum K. A., 1977. *Mathematical Statistics: Basic Ideas and Selected Topics*. New Jersey, Prentice Hall.
- [6] Breslow, N. E., Clayton, D. G., 1993. Approximate Inference in Generalized Linear Mixed Models. *Journal of the American Statistical Association* 88, 9-25.
- [7] Burg, L., Haf, H., Wille, F., 1985. *Höhere Mathematik für Ingenieure*, Bd. 1, Analysis. B. G. Teubner, Stuttgart.
- [8] Casella, G., Berger, R. L., 1990. *Statistical Inference*. Wadsworth and Brooks/Cole, Pacific Grove, California.
- [9] Engel, B., Buist, W., 1993. Analysis of a generalized linear mixed model by iterative least squares with an application to logistic regression with components of variance. *Groep Landbouwwiskunde, Dienst Landbouwkundig Onderzoek*.
- [10] Engel, B., Keen, A., 1994. A simple approach for the analysis of generalzied linear mixed models. *Statistica Neerlandica*, Vol. 48, Nr. 1, 1-22.
- [11] Fahrmeir, L., Hamerle, A., Tutz, G., 1996. *Multivariate statistische Verfahren* (2. Auflage). Berlin, New York, de Gruyter Verlag.
- [12] Fahrmeir, L., Kaufmann, H., 1985. Consistency and asymptotic normality of the maximum likelihood estimator in generalzed linear models. *The Annals of Statistics*, Vol. 13, No. 1, 342-368.
- [13] Fahrmeir, L., Kaufmann, H., 1986. Asymptotic Inference in Discrete Response Models. *Statistical Papers*, 27, 179-205.
- [14] Fahrmeir, L., Kaufmann, H., 1987. Regression Model for Nonstationary Categorical Time Series. *Journal of Time Series Analysis*, 8, 147-160.

-
- [15] Fahrmeir, L., Tutz, G., 2001. *Multivariate Statistical Modelling Based on Generalized Linear Models*, New-York, Springer-Verlag.
- [16] Galton, F., 1889. *Natural inheritance*. Macmillan, London.
- [17] Haldane, J. B. S., 1955. The estimation and significance of the logarithm of a ratio of frequencies. *Ann. Hum Genet* 20, 309-311.
- [18] Holz-Rau, H. C., 1990, *Bestimmungsgrößen des Verkehrsverhaltens: Analyse bundesweiter Haushaltsbefragungen und modellierende Hochrechnung*. Berlin, Institut für Verkehrsplanung und Verkehrswegebau der Technischen Universität Berlin.
- [19] Hosmer, D. H., Lemeshow, S., 1989. *Applied Logistic Regression*. New York, Wiley.
- [20] Jansen, J., 1990. On the statistical analysis of ordinal data when extravariation is present. *Applied Statistics* 39, 75-84.
- [21] McCullagh, P., Nelder, J.A., 1989. *Generalized Linear Models* (2. Auflage). New York, Chapman and Hall.
- [22] McCulloch, C. E., Searle, S. R., 2001. *Generalized, Linear, and Mixed Models*. New York, Wiley.
- [23] McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Hrsg.), *Frontiers in Econometrics*, Seite 105 - 142. New York, Academic Press.
- [24] Nelder, J. A., Wedderburn, R.W.M., 1972. Generalized Linear Models. *Journal of Royal Statistical Society, A* 135, 370-384.
- [25] Powers A. D., Xie Y., 2000. *Statistical Methods for categorical data analysis*. Academic Press.
- [26] Tutz, G., 1990. *Modelle für kategoriale Daten mit ordinalem Skalenniveau, parametrische und nonparametrische Ansätze*. Göttingen: Vandenhoeck & Ruprecht Verlag.
- [27] Tutz, G., 2000. *Die Analyse kategorialer Daten*. Oldenburg, Wissenschaftsverlag GmbH.
- [28] Urban, D., 1993. *Logit-Analyse: Statistische Verfahren zur Analyse von Modellen mit qualitativen Response-Variablen*, Stuttgart, Jena, New York.
- [29] Zängler, T. W., 1999. *Mikroanalyse des Mobilitätsverhaltens in Alltag und Freizeit*. Freising-Weihenstephan: TU München, Lehrstuhl für Wirtschaftslehre des Haushalts, Doktorarbeit im Auftrag der BMW AG München.

-
- [30] Zeger, S. L., Karim, M. R., 1991. Generalized linear models with random effects; A Gibbs sampling approach. *Journal of the American Statistical Association*, Vol. 86, No. 413, 79-86.