

Skript

Generalisierte lineare Modelle

Prof. Dr. Bernd Rönz



Humboldt-Universität zu Berlin

Wirtschaftswissenschaftliche Fakultät

Institut für Statistik und Ökonometrie

2001

Inhaltsverzeichnis

1	Einführung	1
2	Verallgemeinerte lineare Modelle (generalized linear models, GLM)	21
2.1	Die exponentielle Familie von Verteilungen (exponential family of distributions)	22
2.2	Das Konzept der verallgemeinerten linearen Modelle	30
2.3	Schätzung verallgemeinerter linearer Modelle	33
2.4	Hypothesenprüfung in verallgemeinerten linearen Modellen	37
3	Modellierung binärer Daten	49
3.1	Das Logit Modell (logistische Regression)	49
3.2	Die Schätzung des Modells und Hypothesenprüfung	63
3.3	Zur Interpretation der Ergebnisse des logistischen Modells	73
3.4	Das Probit-Modell	108
4	Das multinomiale Logit Modell	115
5	Modellierung multinomialer Daten (log-lineare Modelle)	137
5.1	Die zweidimensionale Kontingenztafel	138
5.2	Stichprobenmodelle für multinomiale Daten	141
5.2.1	Multinomialer Stichprobenmodell	141
5.2.2	Produkt-multinomialer Stichprobenmodell	145

5.2.3	Poisson-Stichprobenmodell	146
5.3	Hypothesenformulierung	148
5.4	Das log-lineare Modell	151
5.5	Schätzung log-linearer Modelle	160
5.6	Hypothesenprüfung	168
5.7	Beispiel	171
5.7.1	Modell unter der Nullhypothese	174
5.7.2	Saturiertes Modell	186
5.8	Log-lineare Modelle für dreidimensionale Kontingenzta- bellen	198
5.8.1	Die Modelle	198
5.8.2	Beispiel	210
5.9	Weitere Probleme	228
Literaturverzeichnis		237
Index		257

1 Einführung

Die wohl am häufigsten gestellte Frage bei sozio-ökonomischen Untersuchungen ist die nach möglichen Zusammenhängen bzw. Abhängigkeiten von Variablen. Den Beitrag, den die Statistik zur Beantwortung dieser Frage leisten kann, besteht in der Überprüfung, ob sich aufgrund von Beobachtungen der Variablen an einer Vielzahl von statistischen Einheiten ein solcher Zusammenhang *quantitativ* nachweisen lässt. Im Ergebnis einer solchen Analyse wird somit die Aussage stehen: *Statistisch* lässt sich eine Abhängigkeit zwischen den Merkmalen nachweisen bzw. nicht nachweisen, wobei insbesondere die Stärke und die Form der Beziehung quantifiziert werden sowie die Richtung des Einflusses ermittelt wird. Dies impliziert sofort, dass festgestellt wird, ob die untersuchten Merkmale in einem bestimmten Ausmaß eine gemeinsame Variation aufweisen bzw. ob die gemeinsame Häufigkeitsverteilung der Variablen eine Beziehung erkennen lässt.

An jede statistische Modellierung sind jedoch Voraussetzungen gebunden, die es im konkreten Fall einzuhalten gilt, d.h., nicht jede beliebige sozio-ökonomische Problemstellung kann mit ein und demselben statistischen Modell bearbeitet werden. Aufgrund dieser Voraussetzungen ergeben sich spezielle Klassen von Modellen.

Für Jahrzehnte dominierend waren die bekannten linearen Modelle, zu denen als wichtigste

1 Einführung

- lineare Regressionsmodelle
- Modelle der Varianzanalyse
- Modelle der Kovarianzanalyse

gehören. Jedoch können sie aufgrund ihrer Voraussetzungen nicht das breite Spektrum von möglichen Abhängigkeitsuntersuchungen abdecken. Es ergibt sich somit die Notwendigkeit einer Generalisierung, was u.a. zu den verallgemeinerten linearen Modellen führt, wobei die linearen Modelle als Spezialfälle in diese verallgemeinerten Modelle eingeschlossen sind.

Verallgemeinerte lineare Modelle (generalized linear models, GLM) beinhalten Modelle auf der Basis eines einheitlichen theoretischen und konzeptionellen Rahmens der Modellbildung. Allen diesen Modellen ist gemeinsam, dass die Beziehungen zwischen zwei oder mehreren Variablen, die an Gruppen von Objekten (statistischen Einheiten) beobachtet wurden, untersucht werden, d.h., es liegt ein multivariater Datensatz zugrunde. Dabei wird weiterhin unterschieden zwischen

- unabhängigen oder erklärenden Variablen (independent, explanatory or predictor variables), deren Beobachtungswerte als fest vorgegeben (nicht zufällig) angesehen werden, d.h., sie sind keine Zufallsvariablen. Ist die unabhängige Variable eine qualitative Variable, wird sie oft als Faktor und ihre Ausprägungen als Faktorstufen (levels) bezeichnet. Ist sie dagegen eine metrisch skalierte (oder quantitative Variable), bezeichnet man sie als Kovariable (covariate).
- abhängigen oder erklärten Variablen (dependent variables, outcomes, explained or response variables), die von den erklärenden Variablen beeinflusst werden und darüber hinaus von einer zufälligen Störgröße abhängen, wodurch sie ebenfalls Zufallsvariablen sind.

Die Betrachtungen werden im weiteren in soweit eingeschränkt, dass stets eine abhängige Variable, bezeichnet mit Y , und eine oder mehrere erklärende Variablen, symbolisiert mit X_1, \dots, X_p , in die Modellierung einbezogen werden. Die verwendeten Variablen können

verschiedene Skalenniveaus (Nominalskala, Ordinalskala, metrische Skala mit der weiteren Unterteilung in Intervallskala, Verhältnisskala und Absolutskala) aufweisen, die bei der statistischen Modellbildung zu berücksichtigen sind.

Die statistische Modellierung läuft im allgemeinen in drei Schritten ab:

1. Modellspezifikation, d.h. ausgehend von der Problemstellung die Formulierung von Hypothesen sowie die Umsetzung dieser Hypothesen in Gleichungen, die die wesentlichen Charakteristika der abhängigen Variablen erfassen, und die Entscheidung über die Verteilungsfunktionen der Zufallsvariablen des Modells,
2. Schätzung der unbekannten Parameter des Modells,
3. Hypothesenprüfung mittels statistischer Tests, vor allem die Überprüfung der Anpassung des Modells an die gegebenen Beobachtungen.

Liegen für eine Variable Y eine Anzahl n von Beobachtungen¹ y_i ($i = 1, \dots, n$) vor, so sucht man zur Erklärung der Variabilität in dieser Variablen nach Mustern oder Strukturen in diesen Daten. Diese Muster oder Strukturen werden oftmals durch die Effekte oder die Abhängigkeit von anderen (erklärenden) Variablen hervorgerufen, was sich mittels eines auf (fach)theoretischen Überlegungen basierenden Modells erfassen lässt. Das Modell soll in soweit allgemeingültig sein, dass es nicht nur die gerade vorliegende Datenreihe beschreibt, sondern auch wiederholt erhobene Beobachtungen zu denselben Variablen.

Bei der Modellspezifikation ist jedoch generell davon auszugehen, dass nicht alle erklärenden Variablen, die einen Einfluß auf die abhängige Variable ausüben, erfaßt werden können. Die Gründe dafür liegen darin, dass

- nicht für alle erklärenden Variablen Beobachtungen zur Verfügung stehen;
- es spezifische Einflüsse bei den einzelnen statistischen Einheiten gibt, deren Erfassung im Modell kaum der Allgemeingültigkeit des Modells dienlich wäre;

¹Beobachtungen werden hier im weitesten Sinne verstanden, d.h., es können beobachtete oder gemessene Werte, Häufigkeiten (Anzahlen), Befragungsergebnisse usw. sein.

1 Einführung

- ein Modell die wesentlichen Charakteristika der Daten zusammenfassend beschreiben soll.

Aus diesem Grunde wird durch die einbezogenen erklärenden Variablen nur ein Teil der Variabilität der abhängigen Variablen erklärt.

Die mathematische Beschreibung dieser Beziehung wird als deterministisch (nicht zufällig) angesehen, als systematische Komponente bezeichnet und enthält in der Regel eine Reihe von unbekannten Konstanten (Parameter), die es aufgrund der Daten zu schätzen gilt. Liegen Schätzwerte für die Parameter vor, erhält man aufgrund dieser systematischen Komponente Werte, symbolisiert mit $\hat{y}_1, \dots, \hat{y}_n$ oder $\hat{\mu}_1, \dots, \hat{\mu}_n$, die als theoretische Werte oder angepasste Werte (fitted values) bezeichnet werden. Man geht davon aus, dass in der systematischen Komponente alle wesentlichen Einflüsse auf die Variable Y enthalten sind.

Der nicht erklärte Teil der Variabilität der abhängigen Variablen wird auf eine Vielzahl sich überlagernder (Zufalls-)Einflüsse zurückgeführt, für die keine Beobachtungen vorliegen und die nur wahrscheinlichkeitstheoretisch beschrieben werden können. Die Gesamtheit dieser Einflüsse werden bei jeder beobachteten Einheit in einer Zufallsvariablen zusammengefasst, über deren Verteilung Annahmen getroffen werden müssen. Ihre Realisationen ergeben sich nach der Schätzung der systematischen Komponente des Modells als Differenz zwischen beobachtetem Wert von Y und dem theoretischen Wert: $y_i - \hat{\mu}_i = e_i$ ($i = 1, \dots, n$). e_i wird als Residuum bezeichnet.

Das sich insgesamt ergebende statistische Modell besteht somit aus einer systematischen Komponente, einer Zufallskomponente und einer Wahrscheinlichkeitsverteilung zur Beschreibung der Form der zufälligen Variation. Durch die Zufallskomponente wird auch die abhängige Variable Y zu einer Zufallsvariablen.

Die statistische Modellierung wird oftmals mit der Übertragung und dem Empfang von Informationen verglichen, die eine Nachricht, das Signal, und Störungen (Rauschen, noise) umfasst. Statistische Daten stellen Informationen dar, so dass ein Modell dieser Daten das Signal und die Störung erfassen muss. Daraus leiten sich eine Reihe von Bezeichnungen ab, die vor allem in der englischsprachigen Fachliteratur verwendet werden.

Das Ziel statistischer Modellierung ist stets, so viel Informationen über die abhängige Variable wie möglich aus den Daten zu gewinnen. Ein Kriterium statistischer Modelle ist somit, einen möglichst großen Anteil der Gesamtvariabilität der abhängigen Variablen durch das Modell zu erklären und damit den Teil der Variabilität, der der Störung zuzuschreiben ist, klein zu halten. Daraus folgt, dass für alle $i = 1, \dots, n$ die theoretischen Werte $\hat{\mu}_i$ (und somit die Werte der unbekannten Parameter) so zu bestimmen sind, dass sie möglichst nahe an den Beobachtungen y_i liegen. Diese allgemeine Forderung muss durch die Festlegung eines Diskrepanzmaßes oder Distanzmaßes präzisiert werden. Zum Beispiel wird bei der bekannten Methode der kleinsten Quadrate als ein solches Distanzmaß die L_2 -Norm (Summe der quadrierten Abweichungen) verwendet:

$\sum (y_i - \hat{\mu}_i)^2$. Die Güte der Anpassung des Modells (model fitting) ist nach der Schätzung der Parameter zu überprüfen. Dafür stehen statistische Tests zur Verfügung.

Es ist jedoch zu beachten, dass durch die Aufnahme einer genügend großen Anzahl von erklärenden Variablen und damit von Parametern in das Modell eine beliebig gewünschte Übereinstimmung von y_i und $\hat{\mu}_i$ erreicht werden kann. Bei Verwendung von so viel Parametern, wie es Beobachtungen gibt, kann die Anpassung sogar perfekt werden. Ein solches Modell wird als gesättigtes Modell (saturiertes Modell), volles Modell (full model) oder ausführliches Modell bezeichnet. Dadurch wird jedoch keine Reduktion der Komplexität erreicht, kein einfacheres theoretisches Modell für die Daten erstellt. Eine sehr genaue Anpassung des Modells an eine gegebene Datenreihe schränkt zum anderen seinen Gültigkeitsbereich ein, d.h., das oben formulierte Ziel der Allgemeingültigkeit des Modells wird nicht erreicht, da für eine andere Beobachtungsreihe der gleichen Variablen Y Veränderungen des Modells unvermeidlich sein würden.

Hieraus ergibt sich weiteres Kriterium statistischer Modelle: die Einfachheit des Modells im Sinne einer geringen Anzahl von Parametern (parsimoniousness). Man wird in der Praxis stets ein einfacheres Modell vorziehen, wenn es die Daten adäquat beschreibt, d.h., im Vergleich zu komplizierten Modellen nur unwesentlich mehr an unerklärter Variabilität übrig lässt.

1 Einführung

In vielen Anwendungsfällen wird man deshalb, aufgrund unterschiedlicher Hypothesen über die Parameter des Modells, verschiedene statistische Modelle spezifizieren können. Die Überprüfung dieser Hypothesen erfolgt in Form des Vergleichs der Modellgüte der diesen Hypothesen zugeordneten Modelle.

Um das Grundprinzip der statistischen Modellierung, das auch für die verallgemeinerten linearen Modelle gültig ist, zu demonstrieren, sollen zwei einfache Beispiele angeführt werden, die auf bekannten linearen Modellen beruhen. Gleichzeitig sollen dabei zwei wichtige Schätzmethoden für die unbekannten Modellparameter in Erinnerung gerufen werden.

Beispiel 1.1

Es soll der Frage nachgegangen werden, ob Unterschiede im mittleren Einkommen bezüglich des Geschlechts existieren.

Die zu erklärende (abhängige) Variable Y ist das Einkommen, die metrisches Skalenniveau aufweist. Die (mögliche Unterschiede in den Mittelwerten) erklärende Variable X ist das Geschlecht, eine nominalskalierte Variable, die deshalb als Faktor bezeichnet wird, wobei die Faktorstufen (levels) männlich bzw. weiblich sind.

Die für diese Problemstellung zu testenden Hypothesen sind:

Nullhypothese: Es besteht kein Unterschied im mittleren Einkommen μ_1 und μ_2 bezüglich des Geschlechts;

$$H_0 : \mu_1 = \mu_2.$$

Alternativhypothese: Es besteht ein Unterschied im mittleren Einkommen μ_1 und μ_2 bezüglich des Geschlechts;

$$H_1 : \mu_1 \neq \mu_2.$$

Gleichbedeutende Hypothesenformulierungen sind:

$$H_0 : \mu_1 - \mu_2 = 0 \quad H_1 : \mu_1 - \mu_2 \neq 0.$$

Zur Überprüfung dieser Hypothesen werden aus der Grundgesamtheit der männlichen und der Grundgesamtheit der weiblichen Erwerbstätigen² unabhängige Zufallsstichpro-

²Die „männliche“ und „weibliche“ Grundgesamtheit müssen dabei nicht notwendig alle Erwerbstätigen

ben gleichen Umfangs (dies nur der Einfachheit halber) gezogen.

Zu jeder Hypothese ist ein adäquates Modell zu spezifizieren:

$$\text{Modell 0: } Y_{ji} = \mu + E_{ji} \quad (1.1)$$

$$\text{Modell 1a: } Y_{ji} = \mu_i + E_{ji} \quad (1.2)$$

Darin bedeuten:

Y_{ji} das Einkommen der i -ten Person in der j -ten Stichprobe ($i = 1, \dots, n$; $j = 1$ für männlich und $j = 2$ für weiblich);

μ ein Parameter, der das mittlere Einkommen repräsentiert, wenn beide Stichproben aus einer einheitlichen Grundgesamtheit stammen würden;

μ_j ein Parameter, der das mittlere Einkommen der j -ten Grundgesamtheit repräsentiert, d.h. ein mittleres Einkommen, das durch die j -te Faktorstufe determiniert ist. μ bzw. μ_j ist die systematische Komponente des Einkommens des jeweiligen Modells.

E_{ji} Zufallsvariable³, die für die i -te Person der j -ten Stichprobe alle individuellen Einflüsse auf die Höhe des Einkommens beinhaltet. Sie entspricht der Zufallskomponente des Modells und wird manchmal auch als Zufallsfehler (random error term) bezeichnet. Über die E_{ji} wird vorausgesetzt, dass sie unabhängig und identisch⁴ normalverteilt sind mit dem Erwartungswert Null und Varianz σ^2 :
 $E_{ji} \sim N(0; \sigma^2)$ für alle j und i .

Aus letzterer Annahme ergibt sich, dass auch die Response-Variablen Y_{ji} Zufallsvariablen sind, die unabhängig und identisch normalverteilt sind mit dem Erwartungswert μ im Modell 0 bzw. μ_j im Modell 1 und Varianz σ^2 :

$Y_{ji} \sim N(\mu; \sigma^2)$ für alle j und i im Modell 0;

erfassen, je nach Fragestellung kann auch ein spezieller Wirtschaftszweig, eine Region usw. zugrunde gelegt werden.

³Obwohl in der Literatur üblicherweise für diese Modellkomponente ein kleiner Buchstabe geschrieben wird, soll durch die hier verwendete Symbolik deutlich hervorgehoben werden, dass es sich um eine Zufallsvariable handelt.

⁴Die Kurzschreibweise für unabhängig und identisch verteilt ist: iid (independent and identically distributed).

1 Einführung

$Y_{ji} \sim N(\mu_j; \sigma^2)$ für alle j und i im Modell 1.

Ob die Stichproben wirklich aus normalverteilten Grundgesamtheiten stammen, ist im konkreten Fall stets zu prüfen, soll hier aber als gültig angenommen werden.

Um den Effekt auf das Einkommen, den die jeweilige Faktorstufe bewirkt, deutlicher hervorzuheben, ist folgende alternative Modellspezifikation möglich:

$$\text{Modell 1b : } Y_{ji} = \mu + \alpha_j + E_{ji} \quad (1.3)$$

α_j ($j = 1, 2$) sind hierin Parameter, die diesen unterschiedlichen Effekt repräsentieren, und es gilt: $\alpha_j = \mu_j - \mu$. Daraus ergibt sich auch eine weitere Formulierungsmöglichkeit der Hypothesen

$$H_0 : \alpha_1 = \alpha_2 = 0 \text{ und } H_1 : \alpha_1 \neq \alpha_2.$$

Bei dem hier formulierten Modell handelt es sich um das (lineare) Modell der (einfaktoriellen) Varianzanalyse.

Das Testen der Hypothesen führt nunmehr darauf hinaus, die Angemessenheit (Adäquatheit) der beiden Modelle zu vergleichen. Dazu werden die Stichproben gezogen, die zu den Realisationen y_{ji} ($j = 1, 2; i = 1, \dots, n$) führen, und die Parameter der Modelle geschätzt.

Für die Parameterschätzung wird die Maximum-Likelihood-Methode verwendet, wobei σ^2 wie eine bekannte Konstante behandelt wird.

Zur Maximum-Likelihood-Methode

Bei der ML-Methode wird von einer Zufallsvariablen Y ausgegangen, deren Dichtefunktion (falls Y stetig ist) bzw. Wahrscheinlichkeitsfunktion (falls Y diskret ist) von unbekannten Parametern $\boldsymbol{\theta}^T = (\theta_1, \dots, \theta_p)$ abhängt: $f(y|\theta_1, \dots, \theta_p) = f(y; \boldsymbol{\theta})$. Der Verteilungstyp der Zufallsvariablen muß bekannt sein. Aus der Grundgesamtheit wird eine einfache Zufallsstichprobe Y_1, \dots, Y_n vom Umfang n gezogen, bei der die einzelnen Stichprobenvariablen unabhängig und identisch verteilt sind. Die Wahrscheinlichkeit, dass die Stichprobenvariable Y_i die Realisation y_i annimmt, ist die bedingte Wahrscheinlichkeit (bedingt durch die Parameter)

$$P(Y_i = y_i | \theta_1, \dots, \theta_p) = f(y_i | \theta_1, \dots, \theta_p) = f(y_i; \boldsymbol{\theta}), \quad i = 1, \dots, n. \quad (1.4)$$

Aufgrund des Multiplikationssatzes für unabhängige Ereignisse folgt für die Wahrscheinlichkeit, die Realisation der Stichprobe (y_1, \dots, y_n) zu erhalten:

$$\begin{aligned} & P(\{Y_1 = y_1\} \cap \{Y_2 = y_2\} \cap \dots \cap \{Y_n = y_n\} | \boldsymbol{\theta}) \\ &= P(Y_1 = y_1 | \boldsymbol{\theta}) \cdot P(Y_2 = y_2 | \boldsymbol{\theta}) \cdot \dots \cdot P(Y_i = y_i | \boldsymbol{\theta}) \cdot \dots \cdot P(Y_n = y_n | \boldsymbol{\theta}) \end{aligned} \quad (1.5)$$

oder

$$f(y_1, \dots, y_n | \boldsymbol{\theta}) = f(\mathbf{y} | \boldsymbol{\theta}) = f(y_1 | \boldsymbol{\theta}) \cdot \dots \cdot f(y_n | \boldsymbol{\theta}) = \prod_i f(y_i | \boldsymbol{\theta}) \quad (1.6)$$

Die gemeinsame Wahrscheinlichkeitsfunktion bzw. Dichtefunktion ist eine Funktion der Zufallsvariablen Y_i ($i = 1, \dots, n$), bedingt durch die Parameter $\boldsymbol{\theta}$.

Wenn die Stichprobe gezogen wurde, d.h., die Stichprobenrealisation $\boldsymbol{\theta}$ nunmehr fest vorgegeben ist, dann hängt der Wert des Produktes der Wahrscheinlichkeiten bzw. der Wert der Dichtefunktion nur noch von den Parametern $\boldsymbol{\theta}$ ab, d.h. ist bei festen y_1, \dots, y_n genannt eine Funktion der unbekannten Parameter $\boldsymbol{\theta}$. Das Produkt wird dann Likelihood-Funktion zur Stichprobe y_1, \dots, y_n genannt und mit $L(\boldsymbol{\theta}; \mathbf{y})$ bezeichnet:

$$L(\boldsymbol{\theta}; \mathbf{y}) = L(\theta_1, \dots, \theta_p | y_1, \dots, y_n) = f(y_1 | \boldsymbol{\theta}) \cdot f(y_2 | \boldsymbol{\theta}) \cdot \dots \cdot f(y_n | \boldsymbol{\theta}) = \prod_i f(y_i | \boldsymbol{\theta}) \quad (1.7)$$

Wie man sieht, stimmen $f(\mathbf{y} | \boldsymbol{\theta})$ und $L(\boldsymbol{\theta}; \mathbf{y})$ algebraisch überein, doch die Notation kennzeichnet den unterschiedlichen Inhalt:

- $f(\mathbf{y} | \boldsymbol{\theta})$ als Funktion der Zufallsvariablen bei festen Parametern, die die Wahrscheinlichkeits- bzw. Dichtefunktion spezifiziert,
- $L(\boldsymbol{\theta}; \mathbf{y})$ als Funktion der Parameter bei festen Realisationen der Zufallsvariablen, die die relative Plausibilität der verschiedenen Werte der Parameter angibt.

Als Maximum-Likelihood-Schätzung (ML-Schätzer) für die unbekannten Parameter $\boldsymbol{\theta}$ wählt man diejenigen Werte $\hat{\boldsymbol{\theta}}$, die die Funktion $L(\boldsymbol{\theta}; \mathbf{y})$ maximieren:

$$L(\hat{\boldsymbol{\theta}}; \mathbf{y}) = \max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}; \mathbf{y}). \quad (1.8)$$

1 Einführung

Von allen möglichen Werten des Parameters θ_j wird also derjenige als Maximum-Likelihood-Schätzung $\hat{\theta}_j$ (für alle $j = 1, \dots, p$) verwendet, bei dem der gegebenen Stichprobenrealisation die größte Dichte bzw. Wahrscheinlichkeit zukommt. Aufgrund des konkreten Stichprobenergebnisses wird also der „plausibelste“ Schätzwert für den unbekannten Parameter gesucht. In diesem Sinne bedeutet Likelihood Mutmaßlichkeit oder Plausibilität.

Ist $L(\boldsymbol{\theta} ; \mathbf{y})$ eine in $\boldsymbol{\theta}$ differenzierbare Funktion, dann hat sie unter gewissen Voraussetzungen bei gegebenen Werten y_1, \dots, y_n genau ein Maximum.

Für die Maximierung der Likelihood-Funktion ist eine notwendige Bedingung, dass die ersten partiellen Ableitungen von $L(\boldsymbol{\theta} ; \mathbf{y})$ nach jedem Element θ_j von $\boldsymbol{\theta}$ Null sind, d.h., es sind die simultanen Gleichungen

$$\frac{\partial L(\boldsymbol{\theta} ; \mathbf{y})}{\partial \theta_j} = 0, \quad \text{für } j = 1, \dots, p \quad (1.9)$$

zu lösen. Weiterhin ist zu prüfen, ob diese Lösungen dem Maximum von $L(\boldsymbol{\theta} ; \mathbf{y})$ entsprechen, d.h., die Matrix der zweiten Ableitungen an der Stelle $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ muß negativ definit sein.

Um die Differentiation zu erleichtern, wird oftmals anstelle der Likelihood-Funktion die logarithmierte Likelihood-Funktion, symbolisiert mit $l(\boldsymbol{\theta} ; \mathbf{y})$, verwendet, die das Maximum genau an der Stelle wie die Likelihood-Funktion aufweist.

Unter bestimmten Bedingungen ist der ML-Schätzer asymptotisch erwartungstreu, konsistent, asymptotisch effizient und asymptotisch normalverteilt.

Eine weitere wichtige Eigenschaft der ML-Schätzung ist: Wenn $g(\boldsymbol{\theta})$ eine Funktion der Parameter $\boldsymbol{\theta}$ ist, dann ist der ML-Schätzer von $g(\boldsymbol{\theta})$ gleich $g(\hat{\boldsymbol{\theta}})$.

Modell 0 des Beispiels 1.1:

Die Grundgesamtheiten sind laut Voraussetzung des Modells 0 normalverteilt:

$$Y_j \sim N(\mu; \sigma^2).$$

Aus diesen Grundgesamtheiten wird je eine einfache Zufallsstichprobe

$$Y_{11}, \dots, Y_{1n} \text{ und } Y_{21}, \dots, Y_{2n} \quad \text{bzw. } Y_{ji} (j = 1, 2; i = 1, \dots, n)$$

gezogen. Dann sind die Stichprobenvariablen Y_{ji} iid $N(\mu; \sigma^2)$:

$$f_0(y_{ji}|\mu; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_{ji}-\mu)^2}{2\sigma^2}}. \quad (1.10)$$

Die Realisationen der Stichprobe sind $y_{11}, \dots, y_{1n}, y_{21}, \dots, y_{2n}$ bzw. y_{ji} ($j = 1, 2; i = 1, \dots, n$). Die gemeinsame Wahrscheinlichkeitsdichte aller Stichprobenvariablen bei Modell M_0 ist:

$$\begin{aligned} f_0(\mathbf{y}|\mu, \sigma^2) &= \prod_{j=1}^2 \prod_{i=1}^n f(y_{ji}|\mu, \sigma^2) = \prod_{j=1}^2 \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2} \cdot (y_{ji}-\mu)^2} \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^{2n} \prod_{j=1}^2 \prod_{i=1}^n \exp \left\{ -\frac{1}{2\sigma^2} \cdot (y_{ji} - \mu)^2 \right\} \\ &= (2\pi\sigma^2)^{-n} \prod_{j=1}^2 \prod_{i=1}^n \exp \left\{ -\frac{1}{2\sigma^2} \cdot (y_{ji} - \mu)^2 \right\} \end{aligned} \quad (1.11)$$

Die Likelihood-Funktion für das Modell M_0 lautet:

$$L_0(\mu|\mathbf{y}, \sigma^2) = (2\pi\sigma^2)^{-n} \prod_{j=1}^2 \prod_{i=1}^n \exp \left\{ -\frac{1}{2\sigma^2} \cdot (y_{ji} - \mu)^2 \right\} \quad (1.12)$$

und demzufolge die log-Likelihood-Funktion

$$l_0 = -n \cdot \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \cdot \sum_{j=1}^2 \sum_{i=1}^n (y_{ji} - \mu)^2. \quad (1.13)$$

Ableitung von l_0 nach μ und Nullsetzen führt zum ML-Schätzer $\hat{\mu}$:

$$\begin{aligned} \frac{\partial l_0}{\partial \mu} &= -\frac{1}{2\sigma^2} \cdot 2 \sum_{j=1}^2 \sum_{i=1}^n (y_{ji} - \mu)(-1) \stackrel{!}{=} 0 \\ \sum_{j=1}^2 \sum_{i=1}^n y_{ji} - 2n\hat{\mu} &= 0 \\ \hat{\mu} &= \frac{1}{2n} \sum_{j=1}^2 \sum_{i=1}^n y_{ji} = \bar{y}. \end{aligned} \quad (1.14)$$

Weiterhin ist:

$$\frac{\partial^2 l_0}{\partial \mu^2} = -2n. \quad (1.15)$$

1 Einführung

Abgesehen von der als bekannt vorausgesetzten Varianz σ^2 und dem konstanten Faktor $n \cdot \log(2\pi\sigma^2)$ ist $-2l_0$ identisch mit dem Kriterium der Summe der quadrierten Abweichungen:

$$-2l_0 = \frac{1}{\sigma^2} \cdot \sum_{j=1}^2 \sum_{i=1}^n (y_{ji} - \mu)^2 + 2n \cdot \log(2\pi\sigma^2). \quad (1.16)$$

Bei variierendem μ nimmt $-2l_0$ den minimalen Wert bei $\hat{\mu} = \bar{y}$ an.

Der maximale Wert \hat{l}_0 , den die l_0 an der Stelle $\hat{\mu}$ annimmt, ist

$$\hat{l}_0 = -n \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \cdot \sum_{j=1}^2 \sum_{i=1}^n (y_{ji} - \bar{y})^2 = -n \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \hat{S}_0, \quad (1.17)$$

worin

$$\hat{S} = \sum_{j=1}^2 \sum_{i=1}^n (y_{ji} - \bar{y})^2 \quad (1.18)$$

ist.

Modell 1 des Beispiels 1.1:

Die Grundgesamtheiten des Modells 1 sind laut Voraussetzung normalverteilt:

$Y_j \sim N(\mu_j; \sigma^2)$, $j = 1, 2$; $i = 1, \dots, n$.

Dann sind die Stichprobenvariablen Y_{ji} iid $N(\mu_j; \sigma^2)$:

$$f_1(y_{ji} | \mu_j; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_{ji} - \mu_j)^2}{2\sigma^2}}. \quad (1.19)$$

Die gemeinsame Wahrscheinlichkeitsdichte aller Stichprobenvariablen bei Modell M_1 ist:

$$\begin{aligned} f_1(\mathbf{y} | \mu_j, \sigma^2) &= \prod_{j=1}^2 \prod_{i=1}^n f(y_{ji} | \mu_j, \sigma^2) \\ &= (2\pi\sigma^2)^{-n} \prod_{j=1}^2 \prod_{i=1}^n \exp \left\{ -\frac{1}{2\sigma^2} \cdot (y_{ji} - \mu_j)^2 \right\} \end{aligned} \quad (1.20)$$

Die Likelihood-Funktion für Modell M_1 lautet:

$$L_1(\mu_j | \mathbf{y}, \sigma^2) = (2\pi\sigma^2)^{-n} \prod_{j=1}^2 \prod_{i=1}^n \exp \left\{ -\frac{1}{2\sigma^2} \cdot (y_{ji} - \mu_j)^2 \right\} \quad (1.21)$$

und demzufolge die log-Likelihood-Funktion

$$\begin{aligned}
l_1 &= -n \cdot \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \cdot \sum_{j=1}^2 \sum_{i=1}^n (y_{ji} - \mu_j)^2 \\
&= -n \cdot \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \left(\sum_{i=1}^n (y_{1i} - \mu_1)^2 + \sum_{i=1}^n (y_{2i} - \mu_2)^2 \right).
\end{aligned} \tag{1.22}$$

Partielle Ableitung von l_1 nach μ_j und Nullsetzen führt zu den ML-Schätzern $\hat{\mu}_j$:

$$\begin{aligned}
\frac{\partial l_1}{\partial \mu_j} &= -\frac{1}{2\sigma^2} \cdot 2 \sum_{j=1}^2 \sum_{i=1}^n (y_{ji} - \mu_j)(-1) \stackrel{!}{=} 0 \\
\sum_{j=1}^2 \sum_{i=1}^n y_{ji} - n\hat{\mu}_j &= 0 \\
\hat{\mu}_j &= \frac{1}{n} \sum_{j=1}^2 \sum_{i=1}^n y_{ji} = \bar{y}_j, \quad j = 1, 2
\end{aligned} \tag{1.23}$$

Weiterhin ist:

$$\frac{\partial^2 l_1}{\partial \mu_j^2} = -n. \tag{1.24}$$

Der maximale Wert \hat{l}_1 , den die l_1 annimmt, ist

$$\begin{aligned}
\hat{l}_1 &= -n \cdot \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \cdot \sum_{j=1}^2 \sum_{i=1}^n (y_{ji} - \bar{y}_j)^2 \\
&= -n \cdot \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \hat{S}_1
\end{aligned} \tag{1.25}$$

mit

$$\hat{S}_1 = \sum_{j=1}^2 \sum_{i=1}^n (y_{ji} - \bar{y}_j)^2. \tag{1.26}$$

Ein Vergleich von \hat{l}_0 und \hat{l}_1 zeigt, dass sie sich nur bezüglich \hat{S}_0 und \hat{S}_1 unterscheiden. Wenn die Nullhypothese $H_0 : \mu_1 = \mu_2$ wahr ist, dann stimmen die Modelle 0 und 1 überein, und \hat{l}_0 und \hat{l}_1 bzw. \hat{S}_0 und \hat{S}_1 sind annähernd gleich, d.h., sie weichen nur zufallsbedingt voneinander ab. In diesem Fall würde man das einfachere Modell 0 für die Anwendung vorziehen.

1 Einführung

Gilt jedoch die Alternativhypothese H_1 , so ist \hat{l}_0 kleiner als \hat{l}_1 bzw. \hat{S}_0 größer als \hat{S}_1 . Nunmehr sollte das Modell 1 verwendet werden. Dies läßt sich unter Verwendung des Verschiebungssatzes zeigen:

$$\begin{aligned}
 \hat{S}_0 &= \sum_{j=1}^2 \sum_{i=1}^n (y_{ji} - \bar{y})^2 = \sum_{j=1}^2 \sum_{i=1}^n [(y_{ji} - \bar{y}_j) + (\bar{y}_j - \bar{y})]^2 \\
 &= \sum_{j=1}^2 \left[\sum_{i=1}^n (y_{ji} - \bar{y}_j)^2 + 2(\bar{y}_j - \bar{y}) \sum_{i=1}^n (y_{ji} - \bar{y}_j) + n(\bar{y}_j - \bar{y})^2 \right] \\
 &\quad \quad \quad \hookrightarrow = 0 \\
 &= \sum_{j=1}^2 \sum_{i=1}^n (y_{ji} - \bar{y}_j)^2 + n \sum_{j=1}^2 (\bar{y}_j - \bar{y})^2 = \hat{S}_1 + n \sum_{j=1}^2 (\bar{y}_j - \bar{y})^2
 \end{aligned} \tag{1.27}$$

Um festzustellen, welche der Hypothesen aufgrund der Stichproben zu einem vorgegebenen Signifikanzniveau vorzuziehen ist, muss geprüft werden, ob \hat{S}_0 signifikant größer als \hat{S}_1 ist. Ein geeigneter Test ist der F-Test.

Dazu müssen die Stichprobenverteilungen der zu \hat{S}_1 und \hat{S}_0 gehörenden Zufallsvariablen betrachtet werden:

$$S_1 = \sum_{j=1}^2 \sum_{i=1}^n (Y_{ji} - \bar{Y}_j)^2, \quad S_0 = \sum_{j=1}^2 \sum_{i=1}^n (Y_{ji} - \bar{Y})^2. \tag{1.28}$$

Für das Modell M_1 gilt:

Analog zu oben kann gezeigt werden, dass unter Verwendung des Verschiebungssatzes gilt:

$$\begin{aligned}
 \frac{1}{\sigma^2} S_1 &= \frac{1}{\sigma^2} \sum_{j=1}^2 \sum_{i=1}^n (Y_{ji} - \bar{Y}_j)^2 = \frac{1}{\sigma^2} \sum_{j=1}^2 \sum_{i=1}^n (Y_{ji} - \mu_j)^2 - \frac{1}{\sigma^2} n \sum_{j=1}^2 (\bar{Y}_j - \mu_j)^2 \\
 &= \sum_{j=1}^2 \sum_{i=1}^n \left(\frac{Y_{ji} - \mu_j}{\sigma} \right)^2 - \sum_{j=1}^2 \left(\frac{\bar{Y}_j - \mu_j}{\sigma/\sqrt{n}} \right)^2
 \end{aligned} \tag{1.29}$$

Wenn die Y_{ji} 's unabhängig sind mit Verteilung $N(\mu_j; \sigma^2)$, dann sind auch die Gruppenmittelwerte \bar{Y}_j unabhängig mit Verteilung $N(\mu_j; \sigma^2/n)$.

S_1/σ^2 ist die Differenz zwischen der Summe der Quadrate von $2n$ unabhängigen Zufallsvariablen $(Y_{ji} - \mu_j)/\sigma$, die jede eine Verteilung $N(0;1)$ haben, und der Summe der Quadrate von zwei unabhängigen Zufallsvariablen $(\bar{Y}_j - \mu_j)/(\sigma^2/n)^{1/2}$, die auch je eine

Verteilung $N(0;1)$ aufweisen. Mit den Eigenschaften der Chi-Quadrat-Verteilung folgt:

$$S_1/\sigma^2 \sim \chi_{2n-2}^2.$$

Für das einfachere Modell M_0 gilt analog:

$$\begin{aligned} \frac{1}{\sigma^2} S_0 &= \frac{1}{\sigma^2} \sum_{j=1}^2 \sum_{i=1}^n (Y_{ji} - \bar{Y})^2 = \frac{1}{\sigma^2} \sum_{j=1}^2 \sum_{i=1}^n (Y_{ji} - \mu)^2 - \frac{1}{\sigma^2} 2n \sum_{j=1}^2 (\bar{Y} - \mu)^2 \\ &= \sum_{j=1}^2 \sum_{i=1}^n \left(\frac{Y_{ji} - \mu}{\sigma} \right)^2 - \sum_{j=1}^2 \left(\frac{\bar{Y} - \mu}{\sigma/\sqrt{2n}} \right)^2 \end{aligned} \quad (1.30)$$

Da angenommen wurde, dass die Y_{ji} 's unabhängig sind und eine Verteilung $N(\mu_j; \sigma^2)$ haben, ist ihr Mittelwert $\bar{Y} \sim N(\mu; \sigma^2/2n)$ verteilt. So ist der 2. Term in S_0/σ^2 das Quadrat einer Zufallsvariablen mit $N(0,1)$. Weiterhin, wenn $\mu_1 = \mu_2 = \mu$ (korrespondierend zu H_0), dann ist der 1. Term in S_0/σ^2 die Summe der Quadrate von $2n$ unabhängigen Zufallsvariablen $(Y_{ji} - \mu)/\sigma$, die jede $N(0;1)$ verteilt ist. Mit den Eigenschaften der Chi-Quadrat-Verteilung folgt: $S_0/\sigma^2 \sim \chi_{2n-1}^2$.

Die Statistik $S_0 - S_1$ repräsentiert die Differenz in der Anpassung zwischen den beiden Modellen. Wenn $H_0 : \mu_1 = \mu_2$ wahr ist, dann gilt

$$1/\sigma^2 (S_0 - S_1) \sim \chi_1^2, \quad f = (2n - 1) - (2n - 2) = 1.$$

Nun ist σ^2 jedoch unbekannt, so dass $S_0 - S_1$ nicht direkt mit kritischen Werten der Chi-Quadrat-Verteilung verglichen werden kann. σ^2 kann jedoch eliminiert werden, indem das Verhältnis von $(S_0 - S_1)/\sigma^2$ und der chi-quadrat-verteilten Zufallsvariablen S_1/σ^2 gebildet wird, wobei jede Zufallsvariable durch die Anzahl der Freiheitsgrade dividiert wird:

$$F = \frac{\frac{(S_0 - S_1)/\sigma^2}{1}}{\frac{S_1/\sigma^2}{2n - 2}} = \frac{S_0 - S_1}{\frac{S_1}{2n - 2}}. \quad (1.31)$$

Wenn H_0 korrekt ist, hat F eine F-Verteilung mit $f_1 = 1$ und $f_2 = 2n - 2$ Freiheitsgraden.

Der Ablehnungsbereich der H_0 ist $\{F | F > F_{f_1; f_2; 1-\alpha}\}$, wobei $F_{f_1; f_2; 1-\alpha}$ das $(1 - \alpha)$ -Quantil der F-Verteilung mit f_1 und f_2 Freiheitsgraden ist.

Beispiel 1.2

Für 2-Personen-Haushalte bzw. 4-Personen-Haushalte wird die Abhängigkeit der Haushaltsausgaben für Urlaub (Y) vom Haushaltsnettoeinkommen (X) untersucht, indem aus jeder Grundgesamtheit (Haushaltstyp) eine einfache Zufallsstichprobe vom Umfang n gezogen wird (gleicher Stichprobenumfang wird hier nur zur Vereinfachung angenommen). Beide Variablen weisen metrisches Skalenniveau auf; X wird deshalb auch als Kovariable bezeichnet. Es sei unterstellt, dass die Abhängigkeit mittels einer linearen Beziehung (Regressionsfunktion⁵) approximiert werden kann:

$$Y_{ji} = \beta_{0j} + \beta_{1j}x_{ji} + E_{ji}. \quad (1.32)$$

Darin bedeuten:

- Y_{ji} die Haushaltsausgaben für Urlaub des i-ten Haushalts des j-ten Haushaltstyps mit $i = 1, \dots, n$, $j = 1$ für 2-Personen-Haushalte und $j = 2$ für 4-Personen-Haushalte;
- β_{0j} Regressionskonstante der linearen Regressionsfunktion für den jeweiligen Haushaltstyp, $j = 1, 2$;
- β_{1j} Regressionskoeffizient, der den Anstieg der linearen Regressionsfunktion für den jeweiligen Haushaltstyp beinhaltet, $j = 1, 2$
- x_{ji} das Haushaltsnettoeinkommen des i-ten Haushalts des j-ten Haushaltstyps ($i = 1, \dots, n$ und $j = 1, 2$); diese Werte werden als gegeben betrachtet, so dass die erklärende Variable X keine Zufallsvariable ist;
- E_{ji} Zufallsvariable, die alle weiteren Einflüsse, die für die Festlegung der Haushaltsausgaben für den Urlaub im i-ten Haushalt des j-ten Haushaltstyps maßgebend sind, enthält. Sie entspricht der Zufallskomponente des Modells. Über die E_{ji} wird vorausgesetzt, dass sie unabhängig und identisch normalverteilt sind mit Erwartungswert Null und Varianz σ^2 : $E_{ji} \sim N(0; \sigma^2)$ für alle j und i.

⁵Vgl. zur Regressionsanalyse u.a. Rönz, B., Förster, E. (1992).

Die systematische Komponente des Modells (1.32) ist der bedingte Erwartungswert von Y_{ji} :

$$E(Y_{ji}|x_{ji}) = \beta_{0j} + \beta_{1j}x_{ji}. \quad (1.33)$$

Die interessierende Frage ist, ob in beiden Haushaltstypen die Veränderung der Haushaltsausgaben für Urlaub pro Einheit Veränderung des Haushaltsnettoeinkommens, d.h. die Regressionskoeffizienten β_{11} und β_{12} , gleich sind. Dies führt zu folgenden Hypothesen:

$$H_0 : \beta_{11} = \beta_{12} = \beta_1$$

$$H_1 : \beta_{11} \neq \beta_{12}.$$

Wenn H_1 gilt, ist das Modell (1.32) das angemessene Modell. Ist dagegen H_0 wahr, kann das folgende einfachere Modell verwendet werden:

$$Y_{ji} = \beta_{0j} + \beta_1 x_{ji} + E_{ji}. \quad (1.34)$$

Die Schätzung der unbekannten Parameter des Modells (1.32) bzw. (1.34) erfolgt mit der Methode der kleinsten Quadrate.

Zur Methode der kleinsten Quadrate (method of least squares)

Gegeben sind Zufallsvariablen Y_1, \dots, Y_n mit den Erwartungswerten $E(Y_i) = \mu_i$ ($i = 1, \dots, n$), die eine Funktion von unbekannten Parametern β_1, \dots, β_p ($p < n$) sind. Im Fall des linearen Regressionsmodells sind die μ_i Linearkombinationen der Parameter $\boldsymbol{\beta}$:

$$E(Y_i|\mathbf{x}_i) = \mu_i = g_i(\boldsymbol{\beta}) = \mathbf{x}_i^T \boldsymbol{\beta} \text{ bzw. } E(\mathbf{y}) = \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}. \quad (1.35)$$

Darin sind:

$$\mathbf{y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_i \\ \vdots \\ Y_n \end{pmatrix}; \quad \mathbf{X} = \begin{pmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1p} \\ \vdots & \dots & \vdots & \dots & \vdots \\ x_{i1} & \dots & x_{ij} & \dots & x_{ip} \\ \vdots & \dots & \vdots & \dots & \vdots \\ x_{n1} & \dots & x_{nj} & \dots & x_{np} \end{pmatrix}; \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_j \\ \vdots \\ \beta_p \end{pmatrix} \quad (1.36)$$

1 Einführung

Der MkQ-Schätzer \mathbf{b} für die unbekannten Parameter $\boldsymbol{\beta}$ ist dann so zu bestimmen, dass die Summe der quadrierten Abweichungen der Zufallsvariablen Y_i von ihrem Erwartungswert ein Minimum ergibt, d.h.

$$S(\boldsymbol{\beta}) = \sum_{i=1}^n (Y_i - \mu_i)^2 \quad (1.37)$$

bzw.

$$\begin{aligned} S(\boldsymbol{\beta}) &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{y}^T \mathbf{y} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} \end{aligned} \quad (1.38)$$

soll minimiert werden. Durch partielle Ableitung von $S(\boldsymbol{\beta})$ nach β_j ($j = 1, \dots, p$)

$$\partial S(\boldsymbol{\beta}) / \partial \boldsymbol{\beta} = -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = -2\mathbf{X}^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0} \quad (1.39)$$

erhält man die Normalgleichungen

$$\mathbf{X}^T \mathbf{X} \mathbf{b} = \mathbf{X}^T \mathbf{y}. \quad (1.40)$$

Vorausgesetzt, dass die Matrix $\mathbf{X}^T \mathbf{X}$ regulär und damit die inverse Matrix $(\mathbf{X}^T \mathbf{X})$ bestimmbar ist, erhält man die geschätzten Regressionsparameter als

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (1.41)$$

Es muss dann noch geprüft werden, ob $S(\boldsymbol{\beta})$ an der Stelle $\boldsymbol{\beta} = \mathbf{b}$ ein Minimum aufweist, d.h., die zweiten partiellen Ableitungen von $S(\boldsymbol{\beta})$ nach β_j für alle $j = 1, \dots, p$ müssen positiv sein. Der Vorteil der Methode der kleinsten Quadrate ist, dass keine Voraussetzungen an die Verteilung der Y_i gestellt werden, um den Schätzer zu erhalten. Allerdings sind solche Annahmen auch dann notwendig, wenn die Stichprobenverteilung des Schätzers bestimmt werden soll.

Modell unter H_1 des Beispiels 1.2: $E(Y_{ji}) = \mu_{ji} = \beta_{0j} + \beta_{1j}x_{ji}$

Die Forderung der MkQ lautet:

$$S(\beta_{0j}, \beta_{1j}) = \sum_{i=1}^n (y_{ji} - \beta_{0j} - \beta_{1j}x_{ji})^2 \rightarrow \min \quad (1.42)$$

Partielle Ableitung von $S(\beta_{0j}, \beta_{1j})$ nach den Parametern und Nullsetzung führt zu den Normalgleichungen:

$$\frac{\partial S(\beta_{0j}, \beta_{1j})}{\partial \beta_{0j}} = -2 \sum_{i=1}^n (y_{ji} - \beta_{0j} - \beta_{1j} x_{ji}) \doteq 0 \quad (1.43)$$

$$nb_{0j} + b_{1j} \sum_{i=1}^n x_{ji} = \sum_{i=1}^n y_{ji}$$

und

$$\frac{\partial S(\beta_{0j}, \beta_{1j})}{\partial \beta_{1j}} = -2 \sum_{i=1}^n (y_{ji} - \beta_{0j} - \beta_{1j} x_{ji}) x_{ji} \doteq 0 \quad (1.44)$$

$$b_{0j} \sum_{i=1}^n x_{ji} + b_{1j} \sum_{i=1}^n x_{ji}^2 = \sum_{i=1}^n x_{ji} y_{ji}$$

Als partielle Ableitung 2. Ordnung nach β_{0j} und β_{1j} erhält man:

$$\frac{\partial^2 S(\beta_{0j}, \beta_{1j})}{\partial \beta_{0j}^2} = 2n > 0, \quad \frac{\partial^2 S(\beta_{0j}, \beta_{1j})}{\partial \beta_{1j}^2} = 2 \sum_{i=1}^n x_{ji}^2 > 0 \quad (1.45)$$

Durch Lösung des Normalgleichungssystems ergeben sich die MkQ-Schätzer:

$$b_{1j} = \frac{n \sum_{i=1}^n x_{ji} y_{ji} - \sum_{i=1}^n x_{ji} \sum_{i=1}^n y_{ji}}{n \sum_{i=1}^n x_{ji}^2 - \left(\sum_{i=1}^n x_{ji} \right)^2} \quad (1.46)$$

$$b_{0j} = \frac{\sum_{i=1}^n y_{ji} \sum_{i=1}^n x_{ji}^2 - \sum_{i=1}^n x_{ji} \sum_{i=1}^n y_{ji} x_{ji}}{n \sum_{i=1}^n x_{ji}^2 - \left(\sum_{i=1}^n x_{ji} \right)^2}.$$

Das Minimum der Summe der Abweichungsquadrate der beobachteten Werte der abhängigen Variablen vom linearen Regressionsmodell unter H_1 , berechnet an der Stelle $\beta_{0j} = b_{0j}$ und $\beta_{1j} = b_{1j}$, ist dann

$$\hat{S}_1 = \sum_{j=1}^2 \sum_{i=1}^n (y_{ji} - b_{0j} - b_{1j} x_{ji})^2. \quad (1.47)$$

1 Einführung

Modell unter H_0 des Beispiels 1.2: $E(Y_{ji}) = \mu_{ji} = \beta_{0j} + \beta_1 x_{ji}$

Entsprechend erhält man die Kleinste-Quadrate-Schätzer für das Modell M_0 , wobei der Schätzer für b_{0j} identisch ist mit dem aus (1.46):

$$b_1 = \frac{n \sum_{j=1}^2 \sum_{i=1}^n x_{ji} y_{ji} - \sum_{j=1}^2 \left(\sum_{i=1}^n x_{ji} \sum_{i=1}^n y_{ji} \right)}{n \sum_{j=1}^2 \sum_{i=1}^n x_{ji}^2 - \sum_{j=1}^2 \left(\sum_{i=1}^n x_{ji} \right)^2}. \quad (1.48)$$

Für das Minimum der Summe der Abweichungsquadrate der beobachteten Werte der abhängigen Variablen vom linearen Regressionsmodell unter H_0 resultiert:

$$\hat{S}_0 = \sum_{j=1}^2 \sum_{i=1}^n (y_{ji} - b_{0j} - b_1 x_{ji})^2. \quad (1.49)$$

Analog zum Beispiel 1.1. wird zum Testen der Nullhypothese der F-Test verwendet. Die Stichprobenverteilung von S_1/σ^2 ist $\chi^2(2n-4)$ und von S_0/σ^2 ist $\chi^2(2n-3)$ unter der Voraussetzung, dass H_0 korrekt ist.

Die Anzahl der Freiheitsgrade ist jeweils die Anzahl der Beobachtungen minus der Anzahl zu schätzender Parameter.

Der Verbesserung in der Modellanpassung unter H_1 (1.32) gegenüber dem Modell unter H_0 (1.34) ist

$(S_0 - S_1)/\sigma^2$ mit einem Freiheitsgrad.

Diese Zufallsvariable ist $\chi^2(1)$ -verteilt. Somit folgt für die Teststatistik

$$F = \frac{\frac{S_0 - S_1}{S_1}}{\frac{2n - 4}{2n - 4}}. \quad (1.50)$$

Unter H_0 ist $F \sim F(1; 2n - 4)$.

2 Verallgemeinerte lineare Modelle (generalized linear models, GLM)

Lineare Modelle reichen für viele praktische Problemstellungen nicht aus, da ihre Annahmen

- direkte Abhängigkeit des Erwartungswertes $E(\mathbf{y}) = \boldsymbol{\mu}$ von der Linearkombination $\mathbf{X}\boldsymbol{\beta}$,
- Normalverteilung der Fehlerterme U_i ,
- eine stetige Verteilung von Y

zu restriktiv sind. Letztere Annahme trifft z.B nicht zu, wenn die response-Variable Y

- nur die Werte Null und Eins aufweist (z.B. nicht arbeitslos - 1, arbeitslos - 0) oder sonstigen nominalen Skalenniveaus ist,
- die Häufigkeit des Eintretens von Ereignissen beinhaltet (sie ist dann diskret),
- als Verhältnis definiert ist (z.B. Arbeitslosenquote).

Läßt man die Annahme der Normalverteilung der Fehlerterms fallen, wird es jedoch schwierig, wenn nicht gar unmöglich, Aussagen über die Verteilung der Schätzfunktionen für die Parameter zu treffen, die aber notwendig sind, wenn konfirmatorische Resultate

erzielt werden sollen. Für eine ganze Reihe spezieller Anwendungsfälle sind durch Abschwächung der Annahmen (zumindest approximative) Lösungsmöglichkeiten gefunden worden.

Nelder und Wedderburn (1972) haben gezeigt, dass viele statistische Modelle, die auf der Linearkombination der Parameter basieren, unter einem einheitlichen Modellrahmen zusammengefasst werden können, die sie als verallgemeinerte lineare Modelle bezeichneten.

Bevor das Konzept der verallgemeinerten linearen Modelle⁶ zusammenfassend definiert werden kann, sind einige Ausführungen zur Verteilung der Variablen Y in diesen Modellen erforderlich.

2.1 Die exponentielle Familie von Verteilungen (exponential family of distributions)

Gegeben ist eine Zufallsvariable Y , deren Dichtefunktion $f(y)$, wenn Y stetig ist, bzw. Wahrscheinlichkeitsfunktion $f(y) = P(Y = y)$, wenn Y diskret ist, von einem Parameter θ abhängt. Diese Verteilung gehört zur exponentiellen Familie von Verteilungen, wenn sie in der Form

$$f(y; \theta) = \exp[a(y)b(\theta) + c(\theta) + d(y)] \quad (2.1)$$

für irgendwelche Funktionen $a(\bullet)$, $b(\bullet)$, $c(\bullet)$ und $d(\bullet)$ geschrieben werden kann⁷.

Gilt $a(y) = y$, dann sagt man, dass die Funktion (2.1) in der kanonischen Form (canonical form) gegeben ist.

Wenn es neben dem Parameter θ , dem das Interesse gilt, weitere Parameter gibt, so werden sie als Störparameter (nuisance parameter) angesehen und als bekannt vorausgesetzt.

Für nachfolgende Anwendungen im Rahmen der verallgemeinerten linearen Modelle sind

⁶Vgl. u.a. Dobson, A.J. (1991); McCullagh, P., Nelder, J.A. (1991), Fahrmeir, L., Hamerle, A. (1984).

⁷Die hier gegebene Notation und die Herleitungen folgen Dobson, A.J. (1991).

2.1 Die exponentielle Familie von Verteilungen (exponential family of distributions)

folgende Herleitungen notwendig.

Die log-Likelihood-Funktion ist der Logarithmus von $f(y; \theta)$:

$$l(\theta; y) = \log f(y; \theta). \quad (2.2)$$

Die erste Ableitung von $l(\theta; y)$ nach θ ist

$$U = \frac{\partial l(\theta; y)}{\partial \theta} = \frac{\partial \log f(y; \theta)}{\partial \theta}, \quad (2.3)$$

wobei folgende Identität gilt, wie mittels der Anwendung der Kettenregel gezeigt werden kann:

$$\begin{aligned} \frac{\partial l(\theta; y)}{\partial \theta} &= \frac{\partial \log f(y; \theta)}{\partial f(y; \theta)} \cdot \frac{\partial f(y; \theta)}{\partial \theta} \\ &= \frac{1}{f(y; \theta)} \cdot \frac{\partial f(y; \theta)}{\partial \theta} \end{aligned} \quad (2.4)$$

U wird auch als score bezeichnet.

Für U werden nun Erwartungswert und Varianz bestimmt. Der Erwartungswert von (2.3) als Integration über den Wertebereich von Y führt unter Berücksichtigung von (2.4) zu

$$\begin{aligned} E(U) &= \int U f(y; \theta) dy = \int \frac{\partial \log f(y; \theta)}{\partial \theta} f(y; \theta) dy \\ &= \int \frac{1}{f(y; \theta)} \frac{\partial f(y; \theta)}{\partial \theta} f(y; \theta) dy \\ &= \int \frac{\partial f(y; \theta)}{\partial \theta} dy. \end{aligned} \quad (2.5)$$

Unter der Voraussetzung, dass die Operationen Differenziation nach θ und Erwartungswertbildung vertauscht werden können, erhält man für die rechte Seite von (2.5) wegen $\int f(y; \theta) dy = 1$:

$$\int \frac{\partial f(y; \theta)}{\partial \theta} dy = \frac{\partial}{\partial \theta} \int f(y; \theta) dy = \frac{\partial}{\partial \theta} 1 = 0 \quad (2.6)$$

und somit

$$E\left(\frac{\partial l(\theta; y)}{\partial \theta}\right) = E(U) = 0. \quad (2.7)$$

Bildet man die 2. Ableitung von $l(\theta; y)$ nach θ (d.h. die 1. Ableitung der U (2.3) nach θ) und wiederum den Erwartungswert, so ergibt sich unter der Voraussetzung, dass die Reihenfolge dieser Operationen Differenziation nach θ und Erwartungswertbildung vertauscht werden können:

$$\begin{aligned} \frac{\partial^2 l(\theta; y)}{\partial \theta^2} &= \frac{\partial^2 \log f(y; \theta)}{\partial \theta^2} \\ E\left(\frac{\partial^2 \log f(y; \theta)}{\partial \theta^2}\right) &= \int \frac{\partial^2 \log f(y; \theta)}{\partial \theta^2} f(y; \theta) dy \\ &= \frac{\partial}{\partial \theta} \int \frac{\partial \log f(y; \theta)}{\partial \theta} f(y; \theta) dy = \frac{\partial^2}{\partial \theta^2} \int f(y; \theta) dy \end{aligned} \quad (2.8)$$

Die rechte Seite ist wegen $\int f(y; \theta) dy = 1$ wiederum gleich Null.

Die linke Seite kann aufgrund der Produktregel in der folgenden Form geschrieben werden:

$$\int \frac{\partial^2 \log f(y; \theta)}{\partial \theta^2} f(y; \theta) dy + \int \frac{\partial \log f(y; \theta)}{\partial \theta} \frac{\partial f(y; \theta)}{\partial \theta} dy. \quad (2.9)$$

Nach Einsetzen des folgenden Ausdrucks, der aus der Identität (2.4) gewonnen wird,

$$\frac{\partial \log f(y; \theta)}{\partial \theta} f(y; \theta) = \frac{\partial f(y; \theta)}{\partial \theta}$$

in den zweiten Term von (2.9) folgt

$$\int \frac{\partial^2 \log f(y; \theta)}{\partial \theta^2} f(y; \theta) dy + \int \frac{\partial \log f(y; \theta)}{\partial \theta} \frac{\partial \log f(y; \theta)}{\partial \theta} f(y; \theta) dy = 0 \quad (2.10)$$

$$\int \frac{\partial^2 \log f(y; \theta)}{\partial \theta^2} f(y; \theta) dy + \int \left[\frac{\partial \log f(y; \theta)}{\partial \theta} \right]^2 f(y; \theta) dy = 0$$

Somit ist

$$E\left[-\frac{\partial^2 \log f(y; \theta)}{\partial \theta^2}\right] = E\left\{\left[\frac{\partial \log f(y; \theta)}{\partial \theta}\right]^2\right\} \quad (2.11)$$

$$E(-U') = E(U^2).$$

2.1 Die exponentielle Familie von Verteilungen (exponential family of distributions)

U' bezeichnet dabei die Ableitung von U bzw. die 2. Ableitung von $l(\theta; y)$ nach θ . Nach diesen Herleitungen ergibt sich für die Varianz⁸ von U unter Berücksichtigung, dass $E(U) = 0$ ist:

$$Var(U) = E[(U - E(U))^2] = E(U^2) - [E(U)]^2 = E(U^2) = E(-U'). \quad (2.12)$$

Die Varianz von U wird auch als Information bezeichnet.

Die Ergebnisse (2.7) und (2.12) gelten für jede Verteilung innerhalb der exponentiellen Familie.

Mit diesen Resultaten kann nun der Erwartungswert und die Varianz von $a(Y)$ in (2.1) bestimmt werden. Die log-Likelihood-Funktion von (2.1) ist

$$l(\theta; y) = a(y)b(\theta) + c(\theta) + d(y). \quad (2.13)$$

Damit sind

$$\begin{aligned} U &= \frac{\partial l(\boldsymbol{\theta}; y)}{\partial \theta} = a(y)b'(\theta) + c'(\theta) \\ U' &= \frac{\partial^2 l(\theta; y)}{\partial \theta^2} = a(y)b''(\theta) + c''(\theta) \end{aligned} \quad (2.14)$$

$$E(U) = b'(\theta)E[a(Y)] + c'(\theta)$$

$$E(-U') = -b''(\theta)E[a(Y)] - c''(\theta)$$

und es folgt unter Beachtung von $E(U) = 0$

$$E[a(Y)] = -\frac{c'(\theta)}{b'(\theta)}. \quad (2.15)$$

Für die Varianz von U ergibt sich (gemäß Fußnote 8):

$$\begin{aligned} Var(U) &= E[(U - E(U))^2] \\ &= E\{[a(Y)b'(\theta) + c'(\theta) - b'(\theta)E(a(Y)) - c'(\theta)]^2\} \end{aligned}$$

⁸Die Varianz einer Zufallsvariablen X ist definiert als $Var(X) = E[(X - E(X))^2] = E(X^2) - [E(X)]^2$.

$$\begin{aligned}
 &= [b'(\theta)]^2 E\{[a(Y) - E(a(Y))]^2\} \\
 &= [b'(\theta)]^2 \{E[(a(Y))^2] - [E(a(Y))]^2\} \\
 &= [b'(\theta)]^2 \text{Var}[a(Y)].
 \end{aligned} \tag{2.16}$$

Berücksichtigung von (2.12), $E(-U')$ aus (2.14) und (2.15) führt zur $\text{Var}[a(Y)]$:

$$\begin{aligned}
 [b'(\theta)]^2 \text{Var}[a(Y)] &= -b''(\theta)E[a(Y)] - c''(\theta) \\
 &= -b''(\theta) \left[\frac{-c'(\theta)}{b'(\theta)} \right] - c''(\theta) \\
 \text{Var}[a(Y)] &= \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{[b'(\theta)]^3}.
 \end{aligned} \tag{2.17}$$

Diese Ergebnisse können auf mehrere Zufallsvariablen verallgemeinert werden.

Gegeben sind nun n unabhängige Zufallsvariablen Y_1, \dots, Y_n , die alle dieselbe Verteilung aufweisen und deren Verteilung von den Parametern $\theta_1, \dots, \theta_p$ mit $p \leq n$ abhängen. Mit $\mathbf{y} = (Y_1, \dots, Y_n)^T$ und $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^T$ kann die gemeinsame log-Likelihood-Funktion der Y_i ($i = 1, \dots, n$) als

$$l(\boldsymbol{\theta} ; \mathbf{y}) = \sum_{i=1}^n l_i(\boldsymbol{\theta} ; y_i) \tag{2.18}$$

geschrieben werden. Die Ableitung nach θ_j ($j = 1, \dots, p$) ergibt die Scores

$$U_j = \frac{\partial l(\boldsymbol{\theta} ; \mathbf{y})}{\partial \theta_j} = \sum_{i=1}^n \frac{\partial l_i(\boldsymbol{\theta} ; y_i)}{\partial \theta_j} \tag{2.19}$$

Mit den gleichen Überlegungen wie vorher folgt

$$E \left[\frac{\partial l_i(\boldsymbol{\theta} ; \mathbf{y})}{\partial \theta_j} \right] = 0 \tag{2.20}$$

und

$$E(U_j) = 0 \quad \text{für alle } j. \tag{2.21}$$

Die Varianz-Kovarianz-Matrix (Informationsmatrix) der U_j ist $\mathfrak{S} = E(\mathbf{u}\mathbf{u}^T)$, wobei $\mathbf{u}^T = (U_1, \dots, U_p)$ der Vektor der ersten Ableitungen von (2.18) nach θ_j ist. Die Elemente der Informationsmatrix ergeben sich zu

$$\mathfrak{S}_{jk} = E(U_j U_k) = E \left[\frac{\partial l}{\partial \theta_j} \frac{\partial l}{\partial \theta_k} \right]. \tag{2.22}$$

2.1 Die exponentielle Familie von Verteilungen (exponential family of distributions)

Wie vorher kann gezeigt werden, dass

$$E \left[\frac{\partial l_i}{\partial \theta_j} \frac{\partial l_i}{\partial \theta_k} \right] = E \left[- \frac{\partial^2 l_i}{\partial \theta_j \partial \theta_k} \right] \quad (2.23)$$

und deshalb die Elemente der Informationsmatrix als

$$\mathfrak{I}_{jk} = E \left[- \frac{\partial^2 l_i}{\partial \theta_j \partial \theta_k} \right] \quad (2.24)$$

geschrieben werden können.

Es soll nun gezeigt werden, dass bekannte Verteilungen zur exponentiellen Familie gehören.

Normalverteilung:

Die Dichtefunktion der Normalverteilung lautet

$$\begin{aligned} f(y; \mu) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2\sigma^2} (y - \mu)^2 \right] \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp \left[-\frac{y^2}{2\sigma^2} + \frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} \right]. \end{aligned} \quad (2.25)$$

In dieser Funktion sind die beiden Parameter μ und σ^2 enthalten, wobei $\theta = \mu$ der interessierende und σ^2 der (als bekannt vorausgesetzte) nuisance Parameter ist.

Mit $(2\pi\sigma^2)^{-1/2} = \exp[-1/2 \cdot \log(2\pi\sigma^2)]$ folgt

$$f(y; \theta) = \exp \left[-\frac{y^2}{2\sigma^2} + \frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{1}{2} \log(2\pi\sigma^2) \right]. \quad (2.26)$$

Damit entspricht sie der Form (2.1)

$$f(y; \theta) = \exp[a(y)b(\theta) + c(\theta) + d(y)] \quad (2.1)$$

mit

$$\begin{aligned} a(y) &= y \\ b(\theta) &= b(\mu) = \frac{\mu}{\sigma^2} \end{aligned} \quad (2.27)$$

$$c(\theta) = c(\mu) = -\frac{\mu^2}{2\sigma^2} - \frac{1}{2} \log(2\pi\sigma^2)$$

$$d(y) = -\frac{y^2}{2\sigma^2}$$

2 Verallgemeinerte lineare Modelle (generalized linear models, GLM)

Wegen $a(y) = y$ liegt sie in der kanonischen Form vor.

Nach (2.15) bzw. (2.17) lässt sich leicht nachprüfen, dass $E(Y) = \mu$ und $Var(Y) = \sigma^2$ sind.

$$\begin{aligned} E[a(y)] = E(Y) &= -\frac{c'(\theta)}{b'(\theta)} = -\frac{c'(\mu)}{b'(\mu)}, \quad \theta = \mu \\ &= -\frac{-2\mu}{\frac{2\sigma^2}{1}} = \mu \end{aligned} \tag{2.28}$$

$$\begin{aligned} Var[a(Y)] &= \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{[b'(\theta)]^3} \\ &= \frac{-\left[-\frac{1}{\sigma^2}\right] \cdot \left[\frac{1}{\sigma^2}\right]}{\left[\frac{1}{\sigma^2}\right]^3} = \sigma^2 \end{aligned} \tag{2.29}$$

Binomialverteilung:

Die Wahrscheinlichkeitsfunktion der Binomialverteilung ist mit

$$f(y; \pi) = \binom{n}{y} \pi^y (1 - \pi)^{n-y} \tag{2.30}$$

gegeben. Hier ist $\theta = \pi$ der interessierende Parameter und n als Anzahl der durchgeführten Experimente bekannt. Als Verteilung der exponentiellen Familie geschrieben, geht (2.30) über in

$$\begin{aligned} f(y; \pi) &= \exp \left[y \log \pi + (n - y) \log(1 - \pi) + \log \binom{n}{y} \right] \\ &= \exp \left[y \log \pi - y \log(1 - \pi) + n \log(1 - \pi) + \log \binom{n}{y} \right] \\ &= \exp \left[y \log \left(\frac{\pi}{1 - \pi} \right) + n \log(1 - \pi) + \log \binom{n}{y} \right]. \end{aligned} \tag{2.31}$$

Es sind

$$\begin{aligned} a(y) &= y \\ b(\theta) &= b(\pi) = \log \left(\frac{\pi}{1 - \pi} \right) \end{aligned}$$

2.1 Die exponentielle Familie von Verteilungen (exponential family of distributions)

$$\begin{aligned} c(\theta) &= c(\pi) = n \log(1 - \pi) \\ d(y) &= \log \binom{n}{y}. \end{aligned} \quad (2.32)$$

Wegen $a(y) = y$ ist sie in der kanonischen Form gegeben.

Nach (2.15) bzw. (2.17) lässt sich ebenfalls nachprüfen, dass $E(Y) = n\pi$ und $Var(Y) = n\pi(1 - \pi)$ sind.

Die 1. Ableitung von $b(\pi) = \log[\pi/(1 - \pi)]$ unter Verwendung der Kettenregel und Quotientenregel ergibt

$$b'(\pi) = \frac{1}{\frac{\pi}{1 - \pi}} \cdot \frac{1}{(1 - \pi)^2} = \frac{1 - \pi}{\pi} \cdot \frac{1}{(1 - \pi)^2} = \frac{1}{\pi \cdot (1 - \pi)} = (\pi - \pi^2)^{-1}$$

und die 1. Ableitung von $c(\pi) = n \cdot \log(1 - \pi)$

$$c'(\pi) = n \frac{1}{1 - \pi} (-1) = -\frac{n}{1 - \pi} = -n(1 - \pi)^{-1}$$

Nach Einsetzen in (2.15) resultiert:

$$E(Y) = -\frac{c'(\pi)}{b'(\pi)} = -\frac{-\frac{n}{1 - \pi}}{\frac{1}{\pi(1 - \pi)}} = \frac{n}{1 - \pi} \cdot \pi(1 - \pi) = n\pi \quad (2.33)$$

Für die 2. Ableitung von $b(\pi)$ und $c(\pi)$ erhält man

$$\begin{aligned} b''(\pi) &= (-1)(\pi - \pi^2)^{-2}(1 - 2\pi) = \frac{-1 + 2\pi}{\pi^2(1 - \pi)^2} \\ c''(\pi) &= -n(-1)(1 - \pi)^{-2}(-1) = -\frac{n}{(1 - \pi)^2} \end{aligned}$$

Nach Einsetzen in (2.17) resultiert:

$$\begin{aligned} Var(Y) &= \frac{b''(\pi)c'(\pi) - c''(\pi)b'(\pi)}{[b'(\pi)]^3} \\ &= \frac{\frac{-1 + 2\pi}{\pi^2(1 - \pi)^2} \cdot \frac{-n}{1 - \pi} - \frac{-n}{(1 - \pi)^2} \cdot \frac{1}{\pi(1 - \pi)}}{\frac{1}{[\pi(1 - \pi)]^3}} \\ &= \frac{\frac{n - 2n\pi}{\pi^2(1 - \pi)^3} + \frac{n}{\pi(1 - \pi)^3}}{\frac{1}{\pi^3(1 - \pi)^3}} = \frac{n - 2n\pi + n\pi}{\pi^2(1 - \pi)^3} \pi^3(1 - \pi)^3 \\ &= n\pi(1 - \pi) \end{aligned} \quad (2.34)$$

Poisson-Verteilung:

Die Wahrscheinlichkeitsfunktion der Poisson-Verteilung lautet

$$f(y; \lambda) = \frac{\lambda^y e^{-\lambda}}{y!} \quad (2.35)$$

bzw. entsprechend (2.1) geschrieben

$$f(y, \lambda) = \exp[y \log \lambda - \lambda - \log y!] \quad (2.36)$$

mit

$$a(y) = y, \quad b(\theta) = b(\lambda) = \log \lambda, \quad c(\theta) = c(\lambda) = -\lambda \quad \text{und} \quad d(y) = -\log y!$$

Für die 1. und 2. Ableitung von $b(\lambda)$ und $c(\lambda)$ ergibt sich

$$b'(\lambda) = 1/\lambda = \lambda^{-1} \quad \text{und} \quad b''(\lambda) = (-1)\lambda^{-2}(1) = -1/\lambda^2;$$

$$c'(\lambda) = -1 \quad \text{und} \quad c''(\lambda) = 0,$$

so dass der Erwartungswert und die Varianz von Y resultiert

$$E(Y) = -c'(\lambda)/b'(\lambda) = -(-1)/(1/\lambda) = \lambda, \quad (2.37)$$

$$Var(Y) = \frac{b''(\lambda)c'(\lambda) - c''(\lambda)b'(\lambda)}{[b'(\lambda)]^3} = \frac{-\frac{1}{\lambda^2}(-1) - 0}{\frac{1}{\lambda^3}} = \lambda. \quad (2.38)$$

2.2 Das Konzept der verallgemeinerten linearen Modelle

Das Konzept der verallgemeinerten linearen Modelle umfasst folgende Komponenten:

1. Die response-Variablen Y_1, \dots, Y_n sind unabhängige Zufallsvariablen mit einer Verteilung aus der exponentiellen Familie, die folgende Eigenschaften hat:
 - a) Die Verteilung jeder Variablen Y_i ($i = 1, \dots, n$) ist in der kanonischen Form gegeben, d.h. $a(y_i) = y_i$, und hängt nur von einem Parameter θ_i ab, wobei die θ_i nicht alle gleich sein müssen. Liegt eine mehrparametrische Verteilung vor, wie z.B. die Normalverteilung, so werden die weiteren Parameter als bekannte Konstante für alle Y_i vorausgesetzt.

- b) Der Verteilungstyp ist für alle Variablen Y_i gleich, so dass die Indizierung für die Funktionen b, c und d entfallen kann.

Somit folgt für die Verteilung von Y_i

$$f(y_i; \theta_i) = \exp[y_i b(\theta_i) + c(\theta_i) + d(y_i)] \quad (2.39)$$

und für die gemeinsame Verteilung aller Y_i (wegen der Unabhängigkeit der Y_i)

$$f(y_1, \dots, y_n; \theta_1, \dots, \theta_n) = \exp \left[\sum_{i=1}^n y_i b(\theta_i) + \sum_{i=1}^n c(\theta_i) + \sum_{i=1}^n d(y_i) \right]. \quad (2.40)$$

Im Falle, dass $\theta_i = \theta$ für alle i gilt, ergibt sich für die gemeinsame Verteilung aller Y_i :

$$f(y_1, \dots, y_n; \theta) = \exp \left[b(\theta) \sum_{i=1}^n y_i + n c(\theta) + \sum_{i=1}^n d(y_i) \right]. \quad (2.41)$$

Die für lineare Modelle charakteristische Unabhängigkeit der response Variablen Y_i ($i = 1, \dots, n$) wird also ohne Modifizierung auf die verallgemeinerten linearen Modelle übernommen. Die Verteilung der Y_i muss jedoch nicht mehr zwangsläufig die Normalverteilung sein.

2. Ein weiterer Aspekt der Verallgemeinerung betrifft den Erwartungswert der response Variablen \mathbf{y} . Im Rahmen der Modellspezifikation interessiert die Abhängigkeit der response Variablen von bestimmten erklärenden Variablen X_j ($j = 1, \dots, p$) mit unbekannten Parametern β_1, \dots, β_p . Die Verknüpfung (link) des Erwartungswertes der response Variablen $\mu_i = E(Y_i)$ mit der Linearkombination der β 's, d.h. $\mathbf{X}\boldsymbol{\beta}$, erfolgt über eine Funktion $\eta_i = g(\mu_i)$, die eine monotone, differenzierbare Funktion ist und als link Funktion (**link function**) bezeichnet wird:

$$\eta_i = g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} = \sum_{j=1}^p x_{ij} \beta_j. \quad (2.42)$$

Für alle i zusammengefasst lässt sich (2.42) schreiben als

$$\boldsymbol{\eta} = g(\boldsymbol{\mu}) = \sum_j \mathbf{x}_j \beta_j = \mathbf{X} \boldsymbol{\beta} \quad (2.43)$$

2 Verallgemeinerte lineare Modelle (generalized linear models, GLM)

In (2.42) bzw. (2.43) sind

\mathbf{x}_i ein $p \times 1$ Vektor fest vorgegebener Werte x_{i1}, \dots, x_{ip} der i -ten statistischen Einheit für die erklärenden Variablen X_1, \dots, X_p ,

\mathbf{x}_j ein $n \times 1$ Vektor fest vorgegebener Werte x_{1j}, \dots, x_{nj} der erklärenden Variablen X_j über alle statistischen Einheiten,

\mathbf{X} eine $n \times p$ Matrix der Werte der erklärenden Variablen,

$\boldsymbol{\beta}$ ein $p \times 1$ Vektor der unbekannten, zu schätzenden Parameter,

$\boldsymbol{\eta}$ ein $n \times 1$ Vektor von link Funktionen und

$\boldsymbol{\mu}$ ein $n \times 1$ Vektor der Erwartungswerte $E(Y_i) = \mu_i$

$$\mathbf{x}_i^T = (x_{i1} \dots x_{ij} \dots x_{ip}); \mathbf{x}_j = \begin{pmatrix} x_{1j} \\ \vdots \\ x_{ij} \\ \vdots \\ x_{nj} \end{pmatrix}; \mathbf{X} = \begin{pmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i1} & \dots & x_{ij} & \dots & x_{ip} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & \dots & x_{nj} & \dots & x_{np} \end{pmatrix} \quad (2.44)$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_j \\ \vdots \\ \beta_p \end{pmatrix}; \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_j \\ \vdots \\ \mu_p \end{pmatrix}; \boldsymbol{\eta} = \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_j \\ \vdots \\ \eta_p \end{pmatrix}.$$

Die Parameter θ_i in (2.40), deren Anzahl gleich der Anzahl der Beobachtungen n sein kann, ist durch die link Funktion (2.42) auf eine geringere Anzahl p von Parametern der Abhängigkeit von den erklärenden Variablen zurückgeführt worden. Statt der Parameter θ_i ($i = 1, \dots, n$) sind nunmehr die Parameter β_j ($j = 1, \dots, p$; $p < n$) zu schätzen. Bei linearen Modellen $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ besteht eine direkte Abhängigkeit der Erwartungswerte der response Variablen \mathbf{y} von der Linearkombination $\mathbf{X}\boldsymbol{\beta}$: $E(\mathbf{y}) = \boldsymbol{\eta} = \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, d.h., $\boldsymbol{\mu}$ und $\boldsymbol{\eta}$ sind identisch. In diesem Sinne sind lineare Modelle in die Spezifikation verallgemeinerter linearer Modelle eingeschlossen.

2.3 Schätzung verallgemeinerter linearer Modelle

Die zur Schätzung der unbekannten Parameter β in verallgemeinerten linearen Modellen verwendete Methode ist die Maximum-Likelihood-Methode, deren Prinzip in Kapitel 1 behandelt wurde.

Die log-Likelihood-Funktion der Verteilung von Y_i in der kanonischen Form (2.39) ist

$$l(\theta_i; y_i) = y_i b(\theta_i) + c(\theta_i) + d(y_i) \quad (2.45)$$

und die log-Likelihood-Funktion der gemeinsamen Verteilung der unabhängigen Zufallsvariablen Y_i (2.40) unter Verwendung der Vektoren $\mathbf{y}^T = (y_1, \dots, y_n)$ und $\boldsymbol{\theta}^T = (\theta_1, \dots, \theta_n)$:

$$l(\boldsymbol{\theta}; \mathbf{y}) = \sum_{i=1}^n l(\theta_i; y_i) = \sum_{i=1}^n y_i b(\theta_i) + \sum_{i=1}^n c(\theta_i) + \sum_{i=1}^n d(y_i). \quad (2.46)$$

Der Erwartungswert der Y_i ergibt sich gemäß (2.15):

$$E(Y_i) = \mu_i = -c'(\theta_i)/b'(\theta_i) \quad (2.15)$$

und die Varianz der Y_i gemäß (2.17):

$$Var(Y_i) = \frac{b''(\theta_i)c'(\theta_i) - c''(\theta_i)b'(\theta_i)}{[b'(\theta_i)]^3} \quad (2.17)$$

Da entsprechend der link Funktion (2.42) der Erwartungswert μ_i der Zufallsvariablen Y_i von den unbekannten Parametern β_j abhängt, erfolgt eine partielle Ableitung von (2.46) nach β_j

$$U_j = \frac{\partial l(\boldsymbol{\theta}; \mathbf{y})}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial l_i}{\partial \beta_j}, \quad j = 1, \dots, p. \quad (2.47)$$

Für die Ableitungen der l_i nach den β_j erhält man durch Anwendung der Kettenregel:

$$\frac{\partial l_i}{\partial \beta_j} = \frac{\partial l_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \beta_j}. \quad (2.48)$$

Den ersten Term auf der rechten Seite erhält man durch Differenzierung von (2.45) nach θ_i und anschließender Berücksichtigung von $c'(\theta_i) = -\mu_i b'(\theta_i)$ aus (2.15)

$$\frac{\partial l_i}{\partial \theta_i} = y_i b'(\theta_i) + c'(\theta_i) = b'(\theta_i)(y_i - \mu_i). \quad (2.49)$$

Differenzierung von (2.15) nach θ_i gemäß der Quotientenregel unter Verwendung von (2.17) ergibt:

$$\begin{aligned}\frac{\partial \mu_i}{\partial \theta_i} &= -\frac{c''(\theta_i)b'(\theta_i) - c'(\theta_i)b''(\theta_i)}{[b'(\theta_i)]^2} = \frac{b''(\theta_i)c'(\theta_i) - c''(\theta_i)b'(\theta_i)}{[b'(\theta_i)]^2} \\ &= b'(\theta_i) \frac{b''(\theta_i)c'(\theta_i) - c''(\theta_i)b'(\theta_i)}{[b'(\theta_i)]^3} = b'(\theta_i) \text{Var}(Y_i),\end{aligned}\tag{2.50}$$

womit der reziproke Ausdruck für den 2. Term in (2.48) gegeben ist.

Differenziert man schließlich noch die link Funktion (2.42) nach β_j , so folgt für den 3. Term in (2.48):

$$\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = x_{ij} \frac{\partial \mu_i}{\partial \eta_i},\tag{2.51}$$

worin x_{ij} das j-te Element des Vektor \mathbf{x}_i^T in (2.44) ist.

Setzt man nun (2.49) bis (2.51) in (2.48) ein, so erhält man

$$\frac{\partial l_i}{\partial \beta_j} = \frac{b'(\theta_i)(y_i - \mu_i) \cdot x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)}{b'(\theta_i) \text{Var}(Y_i)} = \frac{(y_i - \mu_i)x_{ij}}{\text{Var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right),\tag{2.52}$$

was wiederum in (2.47) verwendet wird:

$$U_j = \sum_{i=1}^n \frac{(y_i - \mu_i)x_{ij}}{\text{Var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right), \quad j = 1, \dots, p.\tag{2.53}$$

Dies ist das Resultat der ersten partiellen Ableitung der log-Likelihood-Funktion $l(\boldsymbol{\theta}; \mathbf{y})$ (2.47) nach den Parametern β_j .

Die Elemente der Varianz-Kovarianz-Matrix der Scores U_j (Informationsmatrix) sind gemäß (2.22) definiert als

$$\mathfrak{S}_{jk} = E(U_j U_k) = E \left[\frac{\partial l}{\partial \theta_j} \frac{\partial l}{\partial \theta_k} \right]\tag{2.22}$$

Den Beitrag, den die Variable Y_i zur Informationsmatrix leistet, kann mittels (2.52) bestimmt werden:

$$E \left[\frac{\partial l_i}{\partial \beta_j} \frac{\partial l_i}{\partial \beta_k} \right] = E \left[\left(\frac{(Y_i - \mu_i)x_{ij}}{\text{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} \right) \left(\frac{(Y_i - \mu_i)x_{ik}}{\text{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} \right) \right]$$

$$\begin{aligned}
 &= E \left[\frac{(Y_i - \mu_i)^2 x_{ij} x_{ik}}{[Var(Y_i)]^2} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \right] = E[(Y_i - \mu_i)^2] \frac{x_{ij} \cdot x_{ik}}{[Var(Y_i)]^2} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \quad (2.54) \\
 &= Var(Y_i) \frac{x_{ij} \cdot x_{ik}}{[Var(Y_i)]^2} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 = \frac{x_{ij} x_{ik}}{Var(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2.
 \end{aligned}$$

Setzt man

$$w_{ii} = \frac{1}{Var(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2, \quad (2.55)$$

so ergibt sich für das (j, k)-te Element der Informationsmatrix

$$\mathfrak{S}_{jk} = \sum_{i=1}^n \frac{x_{ij} x_{ik}}{Var(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 = \sum_{i=1}^n x_{ij} x_{ik} w_{ii}. \quad (2.56)$$

Für die gesamte Informationsmatrix \mathfrak{S} resultiert

$$\mathfrak{S} = \mathbf{X}^T \mathbf{W} \mathbf{X}, \quad (2.57)$$

worin \mathbf{X} eine $n \times p$ Matrix und \mathbf{W} eine $n \times n$ Diagonalmatrix sind.

Die ersten partiellen Ableitungen (2.53) werden gleich Null gesetzt. Die Gleichungen $U_j = 0$ ($j = 1, \dots, p$) sind im allgemeinen nichtlinear und müssen iterativ gelöst werden, was unter Verwendung eines Computers realisiert werden kann. Als Lösungsalgorithmen stehen die Newton-Raphson-Methode und die Scoring-Methode (method of scoring) zur Verfügung⁹, wobei eine Anfangslösung $\mathbf{b}_{(0)}$ vorgegeben wird.

Bei der Newton-Raphson-Methode erhält man im m-ten Iterationsschritt

$$\mathbf{b}_{(m)} = \mathbf{b}_{(m-1)} - \left(\frac{\partial^2 l}{\partial \beta_j \partial \beta_k} \right)_{(m-1)}^{-1} \cdot \mathbf{U}_{(m-1)}, \quad (2.58)$$

wobei als Ergebnisse aus dem (m-1)-ten Iterationsschritt sind: $\mathbf{b}_{(m-1)}$ der Vektor der geschätzten Parameter,

$$\left(\frac{\partial^2 l}{\partial \beta_j \partial \beta_k} \right)_{(m-1)} \quad (2.59)$$

die Matrix der zweiten partiellen Ableitungen der log-Likelihood-Funktion $l(\boldsymbol{\theta}; \mathbf{y})$ und $\mathbf{U}_{(m-1)}$ der Vektor der ersten partiellen Ableitungen von $l(\boldsymbol{\theta}; \mathbf{y})$.

⁹Siehe Dobson, A.J. (1991), S. 40 ff.; Judge, G.G. et al. (1988), Introduction of the Theory and Practice of Econometrics, Second Edition, John Wiley, New York, S. 515 ff., 792.

2 Verallgemeinerte lineare Modelle (generalized linear models, GLM)

Bei der Scoring-Methode wird die Matrix der zweiten partiellen Ableitungen der log-Likelihood-Funktion $l(\boldsymbol{\theta}; \mathbf{y})$ durch die Matrix der Erwartungswerte der zweiten partiellen Ableitungen

$$E \left[\frac{\partial^2 l}{\partial \beta_j \partial \beta_k} \right]_{(m-1)} \quad (2.60)$$

ersetzt wird. Wegen

$$\mathfrak{S}_{jk} = E(U_j U_k) = E \left[\frac{\partial l}{\partial \theta_j} \frac{\partial l}{\partial \theta_k} \right] = E \left[- \frac{\partial^2 l}{\partial \theta_j \partial \theta_k} \right] \quad (2.23)$$

und damit $\mathfrak{S} = E[\mathbf{U}\mathbf{U}^T]$ erhält man bei der Scoring-Methode im m-ten Iterationsschritt:

$$\mathbf{b}_{(m)} = \mathbf{b}_{(m-1)} - E \left(\frac{\partial^2 l}{\partial \beta_j \partial \beta_k} \right)_{(m-1)}^{-1} \cdot \mathbf{U}_{(m-1)} = \mathbf{b}_{(m-1)} - \mathfrak{S}_{(m-1)}^{-1} \cdot \mathbf{U}_{(m-1)} \quad (2.61)$$

Multiplikation beider Seiten mit $\mathfrak{S}_{(m-1)}$ führt zu:

$$\mathfrak{S}_{(m-1)} \mathbf{b}_{(m)} = \mathfrak{S}_{(m-1)} \mathbf{b}_{(m-1)} + \mathbf{U}_{(m-1)}. \quad (2.62)$$

Berücksichtigt man (2.56) und (2.53), dann steht auf der rechten Seite von (2.62) ein Vektor mit folgenden Elementen, die sich im (m - 1)-ten Iterationsschritt ergaben:

$$\sum_{k=1}^p \sum_{i=1}^n \frac{x_{ij} x_{ik}}{Var(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 b_{k;(m-1)} + \sum_{i=1}^n \frac{(y_i - \mu_i) x_{ij}}{Var(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right). \quad (2.63)$$

Unter Verwendung von (2.57) und (2.55) resultiert schließlich für (2.62) die Iterationsgleichung der Scoring-Methode:

$$\mathbf{X}^T \mathbf{W} \mathbf{X} \mathbf{b}_{(m)} = \mathbf{X}^T \mathbf{W} \mathbf{z}, \quad (2.64)$$

worin \mathbf{z} ein $n \times 1$ Vektor mit den Elementen

$$z_i = \sum_{k=1}^p x_{ik} b_{k;(m-1)} + (y_i - \mu_i) \left(\frac{\partial \eta_i}{\partial \mu_i} \right) \quad (2.65)$$

ist und $\mathfrak{S} = \mathbf{X}^T \mathbf{W} \mathbf{X}$ als auch $\mathbf{X}^T \mathbf{W} \mathbf{z}$ im (m - 1)-ten Iterationsschritt berechnet wurden. $\mathbf{b}_{(m)}$ ist der $p \times 1$ Vektor der Schätzungen für $\boldsymbol{\beta}$, der sich im m-ten Iterationsschritt durch Lösung von (2.64) ergibt. Die Scoring-Methode für die verallgemeinerten linearen Modelle ist somit eine iterative gewichtete Kleinst-Quadrate-Schätzung, denn (2.64) entspricht den Normalgleichungen eines linearen Modells im Ergebnis der Schätzung nach

der gewichteten Kleinst-Quadrate-Methode. (2.64) muss jedoch iterativ gelöst werden, da \mathbf{W} und \mathbf{z} von der Lösung $\mathbf{b}_{(m-1)}$ des vorherigen Iterationsschrittes abhängen. Der Iterationsprozess bricht ab, wenn eine vorgegebene Konvergenzschranke erreicht ist, d.h., wenn die Differenz zwischen $\mathbf{b}^{(m)}$ und $\mathbf{b}^{(m-1)}$ ausreichend klein ist.

2.4 Hypothesenprüfung in verallgemeinerten linearen Modellen

In diesem Abschnitt soll ein Überblick¹⁰ über die Prüfung von Hypothesen in verallgemeinerten linearen Modellen, vor allem bezüglich der Güte der Anpassung, gegeben werden. Dafür ist es notwendig, dass die Verteilung unter der Nullhypothese zumindest approximativ bekannt ist. Es müssen deshalb einige Stichprobenverteilungen betrachtet werden, was unter der Voraussetzung großer Stichproben geschieht, so dass alle Ergebnisse asymptotisch gelten.

Die Stichprobenverteilung der Scores

Die Scores U_j

$$U_j = \frac{\partial l(\theta; y)}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial l_i}{\partial \beta_j}, \quad j = 1, \dots, p. \quad (2.47)$$

haben gemäß den Ausführungen in Abschnitt 2.1 den Erwartungswert :

$$E(U_j) = 0 \quad \text{für alle } j. \quad (2.7)$$

Die Varianz der Scores U_j ist gemäß (2.12):

$$\text{Var}(U_j) = E(U_j^2). \quad (2.66)$$

¹⁰Für detaillierte Ausführungen siehe u.a. Cox, D.R., Hinkley, D.V. (1974); Fahrmeir, L., Kaufmann, H. (1985).

2 Verallgemeinerte lineare Modelle (generalized linear models, GLM)

Die Kovarianz¹¹ zwischen den Scores U_j und U_k mit $j \neq k$ ergibt sich unter Beachtung von $E(U_j) = E(U_k) = 0$ zu:

$$\text{Cov}(U_j, U_k) = E[(U_j - E(U_j))(U_k - E(U_k))] = E(U_j, U_k). \quad (2.67)$$

Somit folgt für die Varianz-Kovarianz-Matrix der p Scores:

$$E(\mathbf{u}\mathbf{u}^T) = E \begin{pmatrix} U_1^2 & \dots & U_1U_j & \dots & U_1U_p \\ \vdots & & \vdots & & \vdots \\ U_jU_1 & \dots & U_j^2 & \dots & U_jU_p \\ \vdots & & \vdots & & \vdots \\ U_pU_1 & \dots & U_pU_j & \dots & U_p^2 \end{pmatrix}, \quad (2.68)$$

worin $\mathbf{u}^T = (U_1 \dots U_p)$ ist. Nach (2.22) und (2.56) ist aber

$$E(U_jU_k) = \mathfrak{S}_{jk} = E \left[\frac{\partial l}{\partial \beta_j} \frac{\partial l}{\partial \beta_k} \right] = \sum_{i=1}^n \frac{x_{ij}x_{ik}}{\text{Var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \quad (2.56)$$

das j,k-te Element der Informationsmatrix, so dass

$$E(\mathbf{u}\mathbf{u}^T) = \mathfrak{S} \quad (2.69)$$

die Informationsmatrix ist.

Nach dem Zentralen Grenzwertsatz folgt \mathbf{u} (zumindest asymptotisch) einer multivariaten Normalverteilung : $\mathbf{u} \approx N(\mathbf{0}, \mathfrak{S})$.

Vorausgesetzt, dass \mathfrak{S} nicht singular ist und die Inverse \mathfrak{S}^{-1} existiert, kann die Stichprobenfunktion $\mathbf{u}^T \mathfrak{S}^{-1} \mathbf{u}$ gebildet werden, die (zumindest asymptotisch) χ_p^2 -verteilt ist:

$$\mathbf{u}^T \mathfrak{S}^{-1} \mathbf{u} \sim \chi_p^2. \quad (2.70)$$

Dieses Ergebnis soll anhand zweier einfacher Beispiele demonstriert werden.

Beispiel 2.1:

Y_1, \dots, Y_n seien unabhängige und identisch $N(\mu; \sigma^2)$ -verteilte Zufallsvariablen, wobei σ^2

¹¹Sind X und Y zwei Zufallsvariable mit den Erwartungswerten E(X) und E(Y), dann ist die Kovarianz zwischen diesen beiden Variablen definiert als $\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$.

als bekannt vorausgesetzt wird. Als verallgemeinertes lineares Modell konzipiert gibt es hier nur einen Parameter $\theta = \mu$, der von Interesse ist, in diesem Fall jedoch keine erklärenden X-Variablen und die link Funktion ist die Identität $g(\mu) = \mu$.

Die Likelihood-Funktion ist

$$\begin{aligned} L(\mu, \sigma^2; y) &= \prod_{i=1}^n f(y_i | \mu, \sigma) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\frac{1}{2\sigma^2} \cdot \sum_{i=1}^n (y_i - \mu)^2} \end{aligned} \quad (2.71)$$

und die log-Likelihood-Funktion

$$l(\mu; y_1, \dots, y_n) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 - n \log[\sigma\sqrt{2\pi}]. \quad (2.72)$$

Die Ableitung nach μ ergibt

$$U = \frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu) \quad (2.73)$$

und damit resultiert die Zufallsvariable U (Score)

$$U = \frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - \mu) = \frac{n}{\sigma^2} (\bar{Y} - \mu). \quad (2.74)$$

Für den Erwartungswert von U folgt

$$E(U) = E \left[\frac{n}{\sigma^2} (\bar{Y} - \mu) \right] = \frac{n}{\sigma^2} E(\bar{Y}) - \frac{n}{\sigma^2} \mu = \frac{n}{\sigma^2} \mu - \frac{n}{\sigma^2} \mu = 0. \quad (2.75)$$

Die Informationsmatrix \mathfrak{S} besteht nur aus einem Element, d.h.

$$\begin{aligned} \mathfrak{S} &= E(U^2) = Var(U) = E \left[\left(\frac{n}{\sigma^2} (\bar{Y} - \mu) \right)^2 \right] = \frac{n^2}{\sigma^4} E[(\bar{Y} - \mu)^2] \\ &= \frac{n^2}{\sigma^4} Var(\bar{Y}) = \frac{n^2}{\sigma^4} \cdot \frac{\sigma^2}{n} = \frac{n}{\sigma^2} \end{aligned} \quad (2.76)$$

wegen $Var(\bar{Y}) = \sigma^2/n$. Die Stichprobenfunktion $\mathbf{u}^T \mathfrak{S}^{-1} \mathbf{u}$ ergibt in diesem Fall konkret zu

$$\mathbf{u}^T \mathfrak{S}^{-1} \mathbf{u} = \left[\frac{n(\bar{Y} - \mu)}{\sigma^2} \right]^2 \frac{\sigma^2}{n} = \frac{(\bar{Y} - \mu)^2}{\sigma^2/n} = \frac{(\bar{Y} - \mu)^2}{Var(\bar{Y})}. \quad (2.77)$$

2 Verallgemeinerte lineare Modelle (generalized linear models, GLM)

Wie bekannt gilt, da eine Normalerteilung in der Grundgesamtheit angenommen wurde:

$$\bar{Y} \sim N(\mu; \sigma^2/n) \text{ bzw. } (\bar{Y} - \mu)/(\sigma/\sqrt{n}) \sim N(0; 1)$$

und damit

$$(\bar{Y} - \mu)^2/(\sigma^2/n) \sim \chi_1^2,$$

d.h., in diesem Fall gilt $\mathbf{u}^T \mathfrak{S}^{-1} \mathbf{u} \sim \chi_1^2$ nicht nur approximativ, sondern exakt.

Beispiel 2.2:

Wenn Y eine $(n; \pi)$ -binomialverteilte Zufallsvariable ist, so sind

- die Likelihood-Funktion

$$L(\pi; y) = \binom{n}{y} \cdot \pi^y \cdot (1 - \pi)^{n-y} \quad (2.78)$$

- die log-Likelihood-Funktion

$$l(\pi; y) = y \log \pi + (n - y) \log(1 - \pi) + \log \binom{n}{y} \quad (2.79)$$

- die Ableitung nach π (Score)

$$U = \frac{Y}{\pi} - \frac{n - Y}{1 - \pi} = \frac{Y - n\pi}{\pi(1 - \pi)}. \quad (2.80)$$

Da der Erwartungswert der binomialverteilten Zufallsvariablen Y gleich $E(Y) = n\pi$ ist, folgt sofort

$$E(U) = E\left(\frac{Y - n\pi}{\pi(1 - \pi)}\right) = 0. \quad (2.81)$$

Die Informationsmatrix \mathfrak{S} besteht wiederum nur aus einem Element, d.h.

$$\mathfrak{S} = E(U^2) = \text{Var}(U) = \frac{1}{\pi^2(1 - \pi)^2} E[(Y - n\pi)^2] = \frac{\text{Var}(Y)}{\pi^2(1 - \pi)^2} = \frac{n}{\pi(1 - \pi)} \quad (2.82)$$

wegen $\text{Var}(Y) = n\pi(1 - \pi)$.

Für die Stichprobenfunktion $\mathbf{u}^T \mathfrak{S}^{-1} \mathbf{u}$ ergibt sich in diesem Beispiel

$$\mathbf{u}^T \mathfrak{S}^{-1} \mathbf{u} = \frac{(Y - n\pi)^2}{\pi^2(1 - \pi)^2} \frac{\pi(1 - \pi)}{n} = \frac{(Y - n\pi)^2}{n\pi(1 - \pi)}. \quad (2.83)$$

$\mathbf{u}^T \mathfrak{S}^{-1} \mathbf{u} \sim \chi_1^2$ ist jedoch äquivalent zu der Approximation der Binomialverteilung durch die Standardnormalverteilung bei sehr großen Stichproben, d.h.

$$\frac{Y - n\pi}{\sqrt{n\pi(1-\pi)}} \approx N(0; 1). \quad (2.84)$$

Die Stichprobenverteilung der ML-Schätzer

Der ML-Schätzer für die unbekannten Parameter $\boldsymbol{\beta}$ des verallgemeinerten linearen Modells wurde mit \mathbf{b} symbolisiert. Zur Ermittlung der Stichprobenverteilung von \mathbf{b} soll hier mit einem Beispiel begonnen werden.

Beispiel 2.3:

Y_i ($i = 1, \dots, n$) seien unabhängige und $N(\mu_i; \sigma^2)$ -verteilte Zufallsvariablen. $Var(Y_i) = \sigma^2$ für alle i sei bekannt. Der Erwartungswert der Y_i hänge linear von p erklärenden Variablen X_1, \dots, X_p ab, d.h. $\mu_i = E(Y_i) = \mathbf{x}_i^T \boldsymbol{\beta}$ bzw. $\boldsymbol{\mu} = E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$. Dann ist die link Funktion die Identität $g(\mu_i) = \mu_i$ und somit $\mu_i = \eta_i$ in (2.42) sowie $\partial\mu_i/\partial\eta_i = 1$ in (2.52). Die Konsequenzen sind:

- \mathbf{W} in (2.57) ist eine Diagonalmatrix mit den gleichen Elementen $w_{ii} = 1/\sigma^2$,
- die Elemente der Informationsmatrix gemäß (2.56) sind

$$\mathfrak{S}_{jk} = \frac{1}{\sigma^2} \sum_{i=1}^n x_{ij} x_{ik},$$

- die Informationsmatrix (2.57) ist gegeben mit $\mathfrak{S} = (1/\sigma^2) \mathbf{X}^T \mathbf{X}$,
- der Vektor \mathbf{z} mit Elementen gemäß (2.65) lässt sich als $\mathbf{z} = \mathbf{X}\mathbf{b} + \mathbf{y} - \mathbf{X}\mathbf{b} = \mathbf{y}$ schreiben,
- (2.64) ist dann $\mathbf{X}^T \mathbf{X} \mathbf{b} = \mathbf{X}^T \mathbf{y}$.

Wenn $\mathbf{X}^T \mathbf{X}$ nicht singulär ist, ergibt sich der ML-Schätzer zu

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}, \quad (2.85)$$

was ein wohl bekanntes Resultat aus der linearen Regressionsanalyse ist.

Für den Erwartungswert von \mathbf{b} ergibt sich

$$E(\mathbf{b}) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E(\mathbf{y}) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = \boldsymbol{\beta}, \quad (2.86)$$

d.h., \mathbf{b} ist ein erwartungstreuer Schätzer von $\boldsymbol{\beta}$. Für den Erwartungswert von $\mathbf{b} - \boldsymbol{\beta}$ folgt unmittelbar

$$E(\mathbf{b} - \boldsymbol{\beta}) = E(\mathbf{b}) - \boldsymbol{\beta} = \mathbf{0}. \quad (2.87)$$

Unter Verwendung von (2.85) und (2.86) folgt

$$\mathbf{b} - \boldsymbol{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \quad (2.88)$$

und für die Varianz-Kovarianz-Matrix von \mathbf{b} :

$$\begin{aligned} E[(\mathbf{b} - \boldsymbol{\beta})(\mathbf{b} - \boldsymbol{\beta})^T] &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E[(\mathbf{y} - \mathbf{X} \boldsymbol{\beta})(\mathbf{y} - \mathbf{X} \boldsymbol{\beta})^T] \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}, \end{aligned} \quad (2.89)$$

da aufgrund der Annahmen für dieses Beispiel $E[(\mathbf{y} - \mathbf{X} \boldsymbol{\beta})(\mathbf{y} - \mathbf{X} \boldsymbol{\beta})^T]$ eine Diagonalmatrix mit Elementen $Var(Y_i) = \sigma^2$ (für alle i) ist.

Wegen $\mathfrak{S} = (1/\sigma^2) \mathbf{X}^T \mathbf{X}$ bzw. $\sigma^2 \mathfrak{S} = \mathbf{X}^T \mathbf{X}$ geht (2.89) über in

$$E[(\mathbf{b} - \boldsymbol{\beta})(\mathbf{b} - \boldsymbol{\beta})^T] = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} = \sigma^2 (\sigma^2 \mathfrak{S})^{-1} = \sigma^2 (1/\sigma^2) \mathfrak{S}^{-1} = \mathfrak{S}^{-1}, \quad (2.90)$$

d.h., die Varianz-Kovarianz-Matrix der Stichprobenparameter \mathbf{b} ist die inverse Informationsmatrix.

\mathbf{b} als eine Linearkombination der normalverteilten Y_i ($i = 1, \dots, n$) ist ebenfalls normalverteilt:

$$\mathbf{b} \sim N(\boldsymbol{\beta}; \mathfrak{S}^{-1}) \quad (2.91)$$

und somit

$$\mathbf{b} - \boldsymbol{\beta} \sim N(\mathbf{0}; \mathfrak{S}^{-1}) \quad (2.92)$$

bzw. äquivalent

$$(\mathbf{b} - \boldsymbol{\beta})^T \mathfrak{S}(\mathbf{b} - \boldsymbol{\beta}) \sim \chi_p^2. \quad (2.93)$$

Betrachtet man nun wieder die verallgemeinerten linearen Modelle in ihrer Definition gemäß Abschnitt 2.2 und setzt große Stichproben voraus, so gelten die Ergebnisse approximativ und (2.91) bis (2.93) zumindest asymptotisch.

Mittels (2.90) können Konfidenzintervalle für die Schätzer bestimmt werden:

$$b_j \pm c_{1-\alpha/2} s(b_j), \quad (2.94)$$

worin $c_{1-\alpha/2}$ das $1 - \alpha/2$ Quantil der Standardnormalverteilung und $s(b_j)$ der Standardfehler des Schätzers b_j als Wurzel aus dem j -ten Diagonalelement der Matrix \mathfrak{S}^{-1} sind.

Beurteilung der Güte der Anpassung

Die Güte der Anpassung wird beurteilt, indem das spezifizierte Modell mit einem maximalen Modell (full model) verglichen wird, wobei das maximale Modell ein verallgemeinertes lineares Modell mit der gleichen Verteilung aus der exponentiellen Familie und der gleichen link Funktion wie das interessierende Modell ist. Im maximalen Modell ist jedoch die Anzahl der Parameter gleich der Anzahl der Beobachtungen, d.h. gleich n . Dadurch soll mit dem maximalen Modell eine vollständige Beschreibung der Daten erreicht werden. Sind die Vektoren $\hat{\boldsymbol{\theta}}$ bzw. $\hat{\boldsymbol{\theta}}_{max}$ die ML-Schätzer, so können die Werte der Likelihood-Funktionen $L(\hat{\boldsymbol{\theta}}; \mathbf{y})$ und $L(\hat{\boldsymbol{\theta}}_{max}; \mathbf{y})$ berechnet und verglichen werden. Für den Vergleich bildet man das Verhältnis

$$\lambda = L(\hat{\boldsymbol{\theta}}_{max}; \mathbf{y}) / L(\hat{\boldsymbol{\theta}}; \mathbf{y}), \quad (2.95)$$

das als (verallgemeinerte) **likelihood ratio** bezeichnet wird. Äquivalent kann auch

$$\log \lambda = l(\hat{\boldsymbol{\theta}}_{max}; \mathbf{y}) - l(\hat{\boldsymbol{\theta}}; \mathbf{y}) \quad (2.96)$$

verwendet werden. Wenn das interessierende Modell die Daten „gut“ beschreibt, so wird λ bzw. $\log \lambda$ große Werte annehmen.

2 Verallgemeinerte lineare Modelle (generalized linear models, GLM)

Um einen Test durchführen zu können, benötigt man kritische Werte für diese Statistik, was wiederum die Kenntnis der Stichprobenverteilung voraussetzt.

Aus diesem Grund wird letztendlich als Teststatistik

$$D = 2 \log \lambda = 2[l(\hat{\boldsymbol{\theta}}_{max}; \mathbf{y}) - l(\hat{\boldsymbol{\theta}}; \mathbf{y})] \quad (2.97)$$

verwendet. D wird als **log-likelihood ratio statistic** oder **Deviance** bezeichnet. Unter der Voraussetzung großer Stichproben ist D approximativ chi-quadrat-verteilt. Die Anzahl der Freiheitsgrade ergibt sich als Differenz der Parameteranzahl im maximalen Modell und der Parameteranzahl im interessierenden Modell, d.h. n - p, so dass

$$D \approx \chi_{n-p}^2.$$

Dies soll wieder an einem einfachen Beispiel demonstriert werden.

Beispiel 2.4:

Die unabhängigen response Variablen Y_i ($i = 1, \dots, n$) seien normalverteilt mit Mittelwert μ_i (die verschieden sein können) und gemeinsamer (bekannter) Varianz σ^2 . Für die log Likelihood-Funktion ergibt sich dann

$$l(\boldsymbol{\theta}; \mathbf{y}) = l(\boldsymbol{\mu}; \mathbf{y}) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu_i)^2 - \frac{1}{2} n \log(2\pi\sigma^2). \quad (2.98)$$

Für das Modell von Interesse, dass keine erklärenden X-Variablen enthält, wird angenommen, dass $E(Y_i) = \mu$ für alle i ist, so dass $\boldsymbol{\theta}$ nur eine Element, d.h. μ , aufweist. Hierfür wurde bereits gezeigt, dass $\hat{\boldsymbol{\theta}} = \hat{\mu} = \bar{y}$ ist. Der Wert der log Likelihood-Funktion an der Stelle $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ ist somit

$$l(\hat{\boldsymbol{\theta}}; \mathbf{y}) = l(\hat{\mu}; \mathbf{y}) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \bar{y})^2 - \frac{1}{2} n \log(2\pi\sigma^2). \quad (2.99)$$

Betrachtet man dagegen das maximale Modell, bei dem angenommen wird, dass alle μ_i verschieden sind, d.h., $E(Y_i) = \mu_i$ für jedes i ist, dann ist $\boldsymbol{\theta} = \boldsymbol{\mu}$ mit den Elementen μ_1, \dots, μ_n . Als ML-Schätzer erhält man $\hat{\mu}_i = y_i$ bzw. $\hat{\boldsymbol{\mu}} = \mathbf{y}$. Der Wert der log Likelihood-Funktion an der Stelle $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{max}$ ist in diesem Fall

$$\begin{aligned}
 l(\hat{\boldsymbol{\theta}}_{max}; \mathbf{y}) &= l(\mathbf{y}; \mathbf{y}) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu_i)^2 - \frac{1}{2}n \log(2\pi\sigma^2) \\
 &= -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \bar{y})^2 - \frac{1}{2}n \log(2\pi\sigma^2) = -\frac{1}{2}n \log(2\pi\sigma^2).
 \end{aligned}
 \tag{2.100}$$

Somit resultiert für D:

$$D = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \bar{y})^2. \tag{2.101}$$

Mit der Stichprobenvarianz

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 \tag{2.102}$$

kann D auch wie folgt geschrieben werden:

$$D = (n-1)S^2/\sigma^2. \tag{2.103}$$

Nun ist aber bekannt, dass $(n-1)S^2/\sigma^2 \sim \chi_{n-1}^2$ und somit auch $D \sim \chi_{n-1}^2$, wenn das Modell mit $Y_i \sim N(\mu; \sigma^2)$ das korrekte Modell ist.

Das interessierende Modell wird nun in der Weise abgeändert, dass μ_i von p ($p < n$) Parametern abhängt, z.B. $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$. Bezeichnet man mit $\hat{\mu}_i = \mathbf{x}_i^T \mathbf{b}$ die Schätzung für μ_i aufgrund des ML-Schätzers \mathbf{b} , so nimmt die log Likelihood-Funktion an der Stelle $\boldsymbol{\mu} = \hat{\boldsymbol{\mu}}$ den Wert

$$l(\hat{\boldsymbol{\mu}}; \mathbf{y}) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \hat{\mu}_i)^2 - \frac{1}{2}n \log(2\pi\sigma^2) \tag{2.104}$$

und, mit demselben maximalen Modell wie oben, die Deviance den Wert

$$D = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \hat{\mu}_i)^2 \tag{2.105}$$

an. Wenn das spezifizierte Modell korrekt ist, dann ist $D \sim \chi_{n-p}^2$.

Man sollte sich jedoch nicht allein mit einer solchen globalen Beurteilung der Anpassung des Modells an die Daten zufrieden geben, sondern in eine detaillierte Analyse der Residuen einsteigen. $\hat{\mu}_i$, berechnet aufgrund der ML-Schätzung \mathbf{b} , wird als theoretischer

2 Verallgemeinerte lineare Modelle (generalized linear models, GLM)

oder angepasster Wert (fitted value) und die Abweichung $(y_i - \hat{\mu}_i)$ als Residuum bezeichnet. Des weiteren werden standardisierte Residuen berechnet, für die es allerdings verschiedene Definitionen gibt.

$$r_{P,i} = \frac{y_i - \hat{\mu}_i}{s_i} \quad (2.106)$$

ist das sogenannte Pearson Residuum, worin s_i die geschätzte Standardabweichung von μ_i ist.

$$r_{D,i} = \text{sign}(y_i - \hat{\mu}_i) \sqrt{d_i} \quad (2.107)$$

mit

$$d_i = 2[l_i(\hat{\theta}_{max}; y_i) - l_i(\hat{\theta}; y_i)] \quad (2.108)$$

wird als Deviance-Residuum bezeichnet, da es den Beitrag der i-ten Beobachtung zur Deviance beinhaltet.

Diese standardisierten Residuen können für eine tiefergehende Analyse der Anpassung des Modells verwendet werden, indem ihre Verteilung mit einer Normalverteilung verglichen wird. Dies kann mittels eines Wahrscheinlichkeitsplots geschehen, bei dem die geordneten Residuen gegen die unter einer Normalverteilung zu erwartenden Residuen abgetragen werden. Die Punkte sollten auf einer Geraden liegen. Systematische Abweichungen sind Anzeichen für Verletzungen des Modells bzw. für Ausreißer oder ungewöhnliche Beobachtungen. Gleichfalls sollten die standardisierten Residuen gegen die $\hat{\mu}_i$, gegen die Werte jeder erklärenden Variablen bzw. gegen die Werte potentieller X-Variablen geplottet werden. Sind dabei systematische Muster zu erkennen, so beschreibt das Modell die abhängige Variable nicht vollständig und es sollten weitere erklärende Variablen aufgenommen bzw. erklärende Variable ausgetauscht werden.

Hypothesenprüfung

Im letztgenannten Sinne können alternative Modelle spezifiziert werden, wobei vorausgesetzt wird, dass die Modelle die gleiche Verteilung aus der exponentiellen Familie und die gleiche link Funktion, allerdings mit einer unterschiedlichen Anzahl von Parametern,

aufweisen. Es ist dann die Hypothese

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 \text{ mit } \boldsymbol{\beta}_0^T = (\beta_1 \dots \beta_q)$$

gegen

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_1 \text{ mit } \boldsymbol{\beta}_1^T = (\beta_1 \dots \beta_q)$$

zu prüfen, wobei $q < p < n$ gilt. Für den Test wird die Differenz der Deviance beider Modelle verwendet:

$$\begin{aligned} \Delta D = D_0 - D_1 &= 2[l(\mathbf{b}_{max}; \mathbf{y}) - l(\mathbf{b}_0; \mathbf{y})] - 2[(\mathbf{b}_{max}; \mathbf{y}) - l(\mathbf{b}_1; \mathbf{y})] \\ &= 2[l(\mathbf{b}_1; \mathbf{y}) - l(\mathbf{b}_0; \mathbf{y})]. \end{aligned} \tag{2.109}$$

Wenn beide Modelle die Daten angemessen beschreiben und unabhängig voneinander sind, dann ist approximativ $D_0 \sim \chi_{n-q}^2$ und $D_1 \sim \chi_{n-p}^2$ und somit $\Delta D \sim \chi_{p-q}^2$.

Überschreitet der Wert von ΔD den kritischen Wert für ein vorgegebenes Signifikanzniveau α , so wird man das Modell unter H_1 , andernfalls das Modell unter H_0 wählen.

Da in den Devianzen und somit auch in ΔD die meist unbekannte Varianz σ^2 enthalten ist, wird in der Regel auf den F-Test

$$F = \frac{\frac{D_0 - D_1}{p - q}}{\frac{D_1}{n - p}} \tag{2.110}$$

zurückgegriffen. Diese Teststatistik folgt unter der Nullhypothese $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ und für hinreichend großen Stichprobenumfang approximativ einer F-Verteilung mit $f_1 = p - q$ und $f_2 = n - p$ Freiheitsgraden.

3 Modellierung binärer Daten

3.1 Das Logit-Modell (logistische Regressionen)¹²

Bei vielen praktischen Untersuchungen ist die response Variable eine nominalskalierte binäre Zufallsvariable, d.h., sie kann im Ergebnis des Zufallsexperiments nur zwei mögliche Realisationen annehmen.

Beispiel:

- Bei einer Person tritt eine bestimmte Krankheit auf oder nicht.
- Eine bestimmte Behandlung führt zum Erfolg oder nicht.
- Ein Insekt wird bei einer bestimmten Dosierung eines Insektenvernichtungsmittels getötet oder nicht.
- Ein Haushalt besitzt einen privaten PKW oder nicht.
- Ein wahlberechtigter Bürger geht zur Wahl oder nicht.
- Banken stufen einen Kunden als kreditwürdig oder nicht kreditwürdig ein.
- Ein Kunde kauft ein Produkt oder nicht.

¹²Zur logistischen Regression siehe u.a. Dobson, A.J. (1991), S. 104 ff.; McCullagh, P., Nelder, J.A. (1991), S. 98 ff.; Kleinbaum, D.G., (1994); Hosmer, D.W., Lemeshew, S. (1989); Collet, D. (1991); Cramer, J.S. (1991).

3 Modellierung binärer Daten

- Ein Erwerbstätiger ist arbeitslos bzw. nicht arbeitslos.

Wird generell die interessierende Realisation mit „Erfolg“ und die andere Realisation mit „Misserfolg“ bezeichnet, so erfolgt im allgemeinen eine Kodierung mit 1 und 0. Die Zufallsvariable, symbolisiert mit Z , ist somit

$$Z = \begin{cases} 1 & \text{wenn Erfolg} \\ 0 & \text{wenn Misserfolg} \end{cases}$$

mit den Wahrscheinlichkeiten $P(Z = 1) = \pi$ und $P(Z = 0) = 1 - \pi$. Die Wahrscheinlichkeitsverteilung für Z lässt sich in der folgenden Weise schreiben:

$$f(z; \pi) = P(Z = z) = \pi^z(1 - \pi)^{1-z}, \quad z = 0, 1, \quad (3.1)$$

bekannt als Bernoulli-Verteilung mit $E(Z) = \pi$ und $Var(Z) = \pi(1 - \pi)$. Allerdings ist in der Regel die Erfolgswahrscheinlichkeit π unbekannt und muss aufgrund einer Stichprobe geschätzt werden.

Wird das Zufallsexperiment nicht nur einmal, sondern n -mal unter gleichen Bedingungen und unabhängig voneinander durchgeführt, so erhält man n unabhängige Zufallsvariablen Z_i ($i = 1, \dots, n$) mit $P(Z_i = 1) = \pi_i$ und $P(Z_i = 0) = 1 - \pi_i$, d.h. mit den Wahrscheinlichkeitsverteilungen

$$f(z_i; \pi_i) = P(Z_i = z_i) = \pi_i^{z_i}(1 - \pi_i)^{1-z_i}, \quad z_i = 0, 1, \quad i = 1, \dots, n \quad (3.2)$$

sowie $E(Z_i) = \pi_i$ und $Var(Z_i) = \pi_i(1 - \pi_i)$.

Beispiel 3.1:

Die Zufallsvariable Z ist die Kaufentscheidung eines Kunden für eine bestimmte Ware, wobei $Z = 1$ die Entscheidung für den Kauf mit Wahrscheinlichkeit π und $Z = 0$ die Entscheidung gegen den Kauf mit Wahrscheinlichkeit $1 - \pi$ symbolisieren. Werden zum Beispiel $n = 100$ Kunden zufällig und unabhängig voneinander ausgewählt, so erhält man 100 Zufallsvariablen Z_i ($i = 1, \dots, 100$) mit den in (3.2) angegebenen Wahrscheinlichkeitsverteilungen. Die Kaufwahrscheinlichkeiten π_i sind jedoch unbekannt und müssen geschätzt werden.

3.1 Das Logit Modell (logistische Regression)

Wodurch ergeben sich die unterschiedlichen Kaufwahrscheinlichkeiten π_i ? Jeder Kunde wird durch sein „Umfeld“ in der Kaufentscheidung beeinflusst. Zu diesem „Umfeld“ gehören u.a. Geschlecht und Alter des Kunden, zu welcher Haushaltsgröße gehört der Kunde, wie hoch ist das Haushaltsnettoeinkommen, wie hoch ist der Preis der Ware und wie schätzt der Kunde ihn subjektiv ein.

Betrachtet man zunächst nur das Geschlecht als eine die Kaufentscheidung erklärende Variable X , die für männlich mit 1 und für weiblich mit 0 codiert wird. X ist somit eine nominalskalierte, dichotome Variable. Es ergeben sich zwei Gruppen von Kunden, die männlichen und die weiblichen Kunden. Bezeichnet man die Anzahl der Gruppen mit K , so ist $K = 2$. Jeder Kunde ist somit neben seiner Kaufentscheidung (response) z_i durch einen Wert der Variablen $X = \text{„Geschlecht“}$ charakterisiert. Eine fallweise Auflistung aller 100 Kunden ergibt z.B.:

Tabelle 3.1: Fallweise Auflistung der Beobachtungsdaten

Kunde Nr.	Kaufentscheidung (z_i)	Geschlecht (x_i)
1	0	1
2	0	1
3	1	0
\vdots	\vdots	\vdots
100(n)	1	0

Betrachtet man die Gesamtzahl der Kaufentscheidungen für eine gegebene Ausprägung von X , so ergeben sich $K = 2$ Zufallsvariablen Y_1 und Y_2 :

$$Y_1 = \sum_{i=1}^{n_1} Z_{1i}, \quad Y_2 = \sum_{i=1}^{n_2} Z_{2i}, \quad (3.3)$$

die die Anzahl der Käufe Y_1 bei den n_1 männlichen Kunden bzw. die Anzahl der Käufe Y_2 bei den n_2 weiblichen Kunden beinhalten. Es gilt $n = n_1 + n_2$. Es sei nun angenommen, dass die Kaufwahrscheinlichkeit bei den männlichen Kunden und die Kaufwahrscheinlichkeit bei den weiblichen Kunden jeweils gleich ist: $\pi_m = \pi_1$ und $\pi_w = \pi_2$. Setzt man allgemein Y_k ($k = 1, 2$), so folgt jede Zufallsvariable Y_k einer Binomialverteilung mit den

3 Modellierung binärer Daten

Parametern π_k und n_k : $Y_k \sim B(n_k; \pi_k)$:

$$P(Y_k = y_k | x_k) = f(y_k, \pi_k | x_k) = \binom{n_k}{y_k} \pi_k^{y_k} (1 - \pi_k)^{n_k - y_k}, \quad y_k = 0, 1, \dots, n_k. \quad (3.4)$$

Die Daten der Tabelle 3.1 lassen sich damit in einer 2×2 Kontingenztafel zusammenfassen.

Tabelle 3.2: 2×2 Kontingenztafel

	Gruppen		Gruppen	
	männlich ($x_1 = 1$)	weiblich ($x_2 = 0$)	männlich ($x_1 = 1$)	weiblich ($x_2 = 0$)
Erfolge (Käufe)	Y_1	Y_2	π_1	π_2
Misserfolge	$n_1 - Y_1$	$n_2 - Y_2$	$1 - \pi_1$	$1 - \pi_2$
Gesamt	n_1	n_2	1	1

Wie leicht ersichtlich, sind π_1 und π_2 bedingte Wahrscheinlichkeiten, bedingt durch die jeweilige Ausprägung von X.

Betrachtet man als nächstes nur die Haushaltsgröße als eine die Kaufentscheidung erklärende Variable X, die folgende Ausprägungen annehmen kann: 1 - Einpersonenhaushalt, 2 - Zweipersonenhaushalt, 3 - Dreipersonenhaushalt, 4 - sonst, d.h., X ist eine nominalskalierte, mehrkategoriale Variable, die einen Faktor mit mehr als zwei Faktorstufen repräsentiert. Es ergeben sich jetzt $K = 4$ Gruppen von Kunden, die jeweils durch einen bestimmten Wert der Variablen X (Haushaltsgröße) gekennzeichnet sind, d.h. $x_1 = 1$, $x_2 = 2$, $x_3 = 3$ oder $x_4 = 4$. Je Gruppe kann eine Zufallsvariable Y_k ($k = 1, \dots, 4$) definiert werden, die die Anzahl der Käufe für Kunden aus der k-ten Haushaltsgröße beinhalten, wobei n_k die Anzahl von Kunden aus der k-ten Haushaltsgröße ist und $n = \sum_k n_k$ gilt. Setzt man wiederum voraus, dass die Kaufwahrscheinlichkeit bei Kunden aus einer gegebenen Haushaltsgröße jeweils gleich ist (d.h. π_k), so folgt jede Zufallsvariable Y_k ($k = 1, \dots, 4$) einer Binomialverteilung (3.4). Die Ausgangsdaten können in einer 2×4 Kontingenztafel zusammengefasst werden.

Nimmt man jetzt an, dass nur die Variable Haushaltsnettoeinkommen eine die Kaufentscheidung erklärende Variable X ist. Diese Variable X ist eine metrisch skalierte Va-

riable, die sehr viele mögliche Ausprägungen annehmen kann. Es kann somit davon ausgegangen werden, dass sich die Angaben über das Haushaltsnettoeinkommen bei allen befragten Kunden unterscheiden. Nunmehr ergeben sich $K = n = 100$ Gruppen, wobei jede Gruppe durch einen spezifischen Einkommenswert charakterisiert ist und $n_k = 1$ für alle $k = 1, \dots, K$ gilt. Folglich resultieren $K = n$ Zufallsvariablen Y_k ($k = 1, \dots, n$), die (wegen $Y_k = Z_i$) einer Bernoulli-Verteilung folgen, die sich jedoch als Binomialverteilung gemäß (3.4) schreiben lässt:

$$P(Y_k = y_k | x_k) = f(y_k, \pi_k | x_k) = \binom{1}{y_k} \pi_k^{y_k} (1 - \pi_k)^{1-y_k}, \quad y_k = 0, 1. \quad (3.5)$$

Eine kleinere Anzahl von Gruppen könnte erreicht werden, wenn entweder bei der Befragung bereits Einkommenklassen vorgegeben werden und der befragte Kunde angibt, in welche Einkommenklasse er fällt, oder nachträglich eine Klassierung des Haushaltsnettoeinkommens vorgenommen wird. In einem solchen Fall würde entweder die Nummer der Einkommensklasse oder die Klassenmitte als Wert der Variablen X verwendet.

In der Regel wird jedoch bei der Schätzung der Erfolgswahrscheinlichkeit ihre Abhängigkeit von mehreren erklärenden X -Variablen unterstellt, was zu folgenden allgemeinen Überlegungen führt:

Zusätzlich zur response Variablen Y gibt es eine Reihe von erklärenden Variablen X_j , $j = 1, \dots, J$. Diese erklärenden Variablen können beliebiges Skalenniveau haben, d.h., sie können Faktoren (kategoriale Variablen, Gruppierungsvariablen) und/oder beobachtete metrisch skalierte Variablen sein. Die Werte der X -Variablen werden als fest vorgegeben angesehen (d.h., sie sind keine Zufallsvariablen). Aufgrund der unterschiedlichen Werte der X -Variablen ergeben sich K verschiedene Vektoren von Kovariablen-Werten $\mathbf{x}_k^T = (x_{k1} \dots x_{kj} \dots x_{kJ})$, $k = 1, \dots, K$. Jeder Vektor \mathbf{x}_k charakterisiert eine Gruppe, in der die statistischen Einheiten (Fälle) bezüglich der erklärenden Variablen X_j ($j = 1, \dots, J$) dieselben Beobachtungswerte aufweisen.

Vorausgesetzt, dass in jeder Gruppe die Erfolgswahrscheinlichkeit π_k konstant ist (d.h., für alle Fälle innerhalb der Gruppe gleich ist), folgt die response Variable $Y_k =$ "Anzahl der Erfolge in der k -ten Gruppe" ($k = 1, \dots, K$) einer Binomialverteilung (3.4): $Y_k \sim B(n_k; \pi_k)$.

3 Modellierung binärer Daten

Wie bereits im Abschnitt 2.1 gezeigt, gehört die Binomialverteilung

$$\begin{aligned}
 f(y_k, \pi_k | \mathbf{x}_k) &= \binom{n_k}{y_k} \pi_k^{y_k} (1 - \pi_k)^{n_k - y_k} \\
 &= \exp \left[y_k \log \pi_k + (n_k - y_k) \log(1 - \pi_k) + \log \binom{n_k}{y_k} \right] \\
 &= \exp \left[y_k \log \pi_k + n_k \log(1 - \pi_k) - y_k \log(1 - \pi_k) + \log \binom{n_k}{y_k} \right] \\
 &= \exp \left[y_k \log \left(\frac{\pi_k}{1 - \pi_k} \right) + n_k \log(1 - \pi_k) + \log \binom{n_k}{y_k} \right]
 \end{aligned} \tag{3.6}$$

zur exponentiellen Familie

$$f(y_k; \pi_k) = \exp[a(y_k)b(\pi_k) + c(\pi_k) + d(y_k)] \tag{2.1}$$

mit $\theta = \pi_k$ und

$$\begin{aligned}
 a(y_k) &= y_k \\
 b(\pi_k) &= \log \left(\frac{\pi_k}{1 - \pi_k} \right) \\
 c(\pi_k) &= n_k \log(1 - \pi_k) \\
 d(y_k) &= \log \binom{n_k}{y_k}.
 \end{aligned}$$

Die Beobachtungsdaten können in einer $2 \times K$ Kontingenztafel zusammengefasst werden.

Tabelle 3.3: Kontingenztafel mit K Kovariaten-Gruppen

	Gruppen				Gruppen			
	1	2	...	K	1	2	...	K
Erfolge	Y_1	Y_2	...	Y_K	π_1	π_2	...	π_K
Misserfolge	$n_1 - Y_1$	$n_2 - Y_2$...	$n_K - Y_K$	$1 - \pi_1$	$1 - \pi_2$...	$1 - \pi_K$
Gesamt	n_1	n_2	...	n_K	1	1	...	1

Die Erfolgswahrscheinlichkeiten π_k sind bedingte Wahrscheinlichkeiten, bedingt durch die jeweiligen Ausprägungen $\mathbf{x}_k^T = (x_{k1} \dots x_{kj} \dots x_{kJ})$ der erklärenden X-Variablen.

3.1 Das Logit Modell (logistische Regression)

Mit n_k für jede Gruppe folgt für den Gesamtstichprobenumfang: $n = n_1 + \dots + n_k + \dots + n_K$. Wird eine Konstante in das Modell aufgenommen (wie z.B. beim linearen Regressionsmodell), so sei $x_{k1} = 1$ für alle k .

In Matrixnotation erhält man:

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_k \\ \vdots \\ y_K \end{pmatrix}; \quad \mathbf{X} = \begin{pmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_k^T \\ \vdots \\ \mathbf{x}_K^T \end{pmatrix} = \begin{pmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1J} \\ \vdots & \dots & \vdots & \dots & \vdots \\ x_{k1} & \dots & x_{kj} & \dots & x_{kJ} \\ \vdots & \dots & \vdots & \dots & \vdots \\ x_{K1} & \dots & x_{Kj} & \dots & x_{KJ} \end{pmatrix}. \quad (3.7)$$

Wenn stetige Kovariable gegeben sind, so kann man im allgemeinen nicht davon ausgehen, dass unter gleichen Bedingungen kontrollierbare Versuchsdurchführungen möglich sind, so dass die Anzahl n_k der Messwiederholungen sehr klein bzw. gleich 1 sein wird. Dann ist $Y_k \sim B(1; \pi_k)$ und $K = n$, d.h., die Anzahl der Gruppen ist gleich dem Gesamtstichprobenumfang. Gegebenenfalls wird $\mathbf{x}_k^T = (x_{k1} \dots x_{kj})$ mehrmals in \mathbf{X} aufgelistet, was ein Notationsdetail, aber kein konzeptionelles Problem ist.

Für die Y_k zusammen ergibt sich die gemeinsame Likelihood-Funktion als

$$L(\pi_1, \dots, \pi_K; y_1, \dots, y_K) = \prod_{k=1}^K f(y_k; \pi_k) = \prod_{k=1}^K \binom{n_k}{y_k} \pi_k^{y_k} (1 - \pi_k)^{n_k - y_k} \quad (3.8)$$

und die log-Likelihood-Funktion als

$$l(\pi_1, \dots, \pi_K; y_1, \dots, y_K) = \sum_{k=1}^K \left[y_k \log \left(\frac{\pi_k}{1 - \pi_k} \right) + n_k \log(1 - \pi_k) + \log \binom{n_k}{y_k} \right]. \quad (3.9)$$

$\log(\text{Binomialkoeffizient})$ als letzter Term in (3.9), der π_k nicht enthält, kann auch vernachlässigt werden, da er für die Schätzung der π_k keine Rolle spielt:

$$l(\pi_1, \dots, \pi_K; y_1, \dots, y_K) = \sum_{k=1}^K \left[y_k \log \left(\frac{\pi_k}{1 - \pi_k} \right) + n_k \log(1 - \pi_k) \right]. \quad (3.10)$$

Das Ziel der statistischen Analyse ist die Schätzung der Abhängigkeit der Erfolgswahrscheinlichkeiten π_k von den erklärenden Variablen:

$$g(\pi_k) = \eta_k = \mathbf{x}_k^T \boldsymbol{\beta} = \sum_j x_{kj} \beta_j; \quad k = 1, \dots, K, \quad (3.11)$$

3 Modellierung binärer Daten

worin g eine link Funktion und $\boldsymbol{\beta}^T = (\beta_1 \dots \beta_j)$ ein Vektor unbekannter Parameter sind. Die Schätzung der K unbekannten Wahrscheinlichkeiten π_k wird auf die Schätzung der geringeren Anzahl J von unbekannten Parameter β_j verlagert. Nach deren Schätzung können auch die Wahrscheinlichkeiten π_k bestimmt werden. Vorher muss jedoch die link Funktion $g(\pi_k)$ spezifiziert werden.

Beispiel 3.2¹³:

Von $n = 100$ zufällig ausgewählten Personen wird erfasst, ob eine Herzkranzgefäßerkrankung (HKE) vorliegt ($= 1$) oder nicht ($= 0$). Zusätzlich wurde die Variable $X = \text{“Alter“}$ für jede Person notiert.

Tabelle 3.4: Herzkranzgefäßerkrankung (HKE) und Alter von 100 Personen

HKE	Alter	HKE	Alter	HKE	Alter	HKE	Alter	HKE	Alter
0	20	0	34	0	41	1	48	0	57
0	23	0	34	0	42	1	48	1	57
0	24	1	34	0	42	0	49	1	57
0	25	0	34	0	42	0	49	1	57
1	25	0	34	1	42	1	49	1	57
0	26	0	35	0	43	0	50	0	58
0	26	0	35	0	43	1	50	1	58
0	28	0	36	1	43	0	51	1	58
0	28	1	36	0	44	0	52	1	59
0	29	0	36	0	44	1	52	1	59
0	30	0	37	1	44	1	53	0	60
0	30	1	37	1	44	1	53	1	60
0	30	0	37	0	45	1	54	1	61
0	30	0	38	1	45	0	55	1	62
0	30	0	38	0	46	1	55	1	62
1	30	0	39	1	46	1	55	1	63
0	32	1	39	0	47	1	56	0	64
0	32	0	40	0	47	1	56	1	64
0	33	1	40	1	47	1	56	1	65
0	33	0	41	0	48	0	57	1	69

Wie ersichtlich treten verschiedene Altersangaben und somit $K = 43$ Kovariablen-Gruppen auf. Für jede Gruppe gibt die Zufallsvariable Y_k ($k = 1, \dots, 43$) die Anzahl des Auftre-

¹³Das Beispiel wurde entnommen aus: Hosmer, D.W., Lemeshow, S. (1989) S. 2 ff.

3.1 Das Logit Modell (logistische Regression)

tens von HKE an.

Tabelle 3.5: Realisationen der Zufallsvariablen Y_k für gegebenes x_k

Alter x_k	Y_k (für HKE=1)	$n_k - Y_k$ (für HKE=0)	Gesamt n_k	Alter x_k	Y_k (für HKE=1)	$n_k - Y_k$ (für HKE=0)	Gesamt n_k
20	0	1	1	47	1	2	3
23	0	1	1	48	2	1	3
24	0	1	1	49	1	2	3
25	1	1	2	50	1	1	2
26	0	2	2	51	0	1	1
28	0	2	2	52	1	1	2
29	0	1	1	53	2	0	2
30	1	5	6	54	1	0	1
32	0	2	2	55	2	1	3
33	0	2	2	56	3	0	3
34	1	4	5	57	4	2	6
35	0	2	2	58	2	1	3
36	1	2	3	59	2	0	2
37	1	2	3	60	1	1	2
38	0	2	2	61	1	0	1
39	1	1	2	62	2	0	2
40	1	1	2	63	1	0	1
41	0	2	2	64	1	1	2
42	1	3	4	65	1	0	1
43	1	2	3	69	1	0	1
44	2	2	4				
45	1	1	2	$\bar{\Sigma}$	$\bar{43}$	$\bar{57}$	$\bar{100}$
46	1	1	2				

Führt man eine Klassierung des Alters in die Klassen 20-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59 und 60-69 durch, so reduziert sich die Anzahl der Gruppen, die durch unterschiedliche x-Werte gekennzeichnet sind, auf $K = 8$ (siehe Tabelle 3.6).

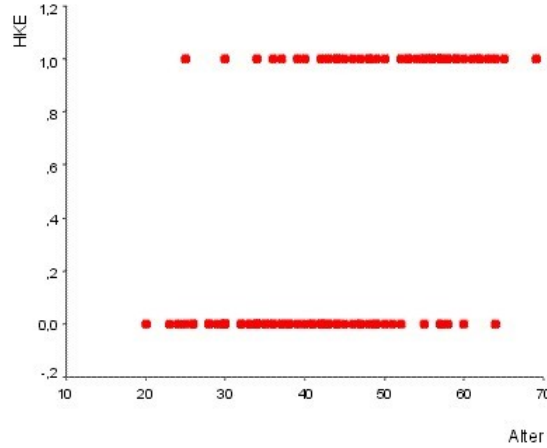
Die Frage lautet:

Inwieweit hängt die Wahrscheinlichkeit des Auftretens von HKE vom Alter ab?

Ad hoc würde zur Beantwortung dieser Frage das Modell der einfachen linearen Regressi-

on einfallen, um $\boldsymbol{\beta}^T = (\beta_0 \beta_1)$ zu schätzen, worin die link Funktion die Identitätsfunktion $g(\pi_k) = \pi_k = \mathbf{x}_k^T \boldsymbol{\beta}$ ist. Abbildung 3.1 zeigt einen Scatterplot des Alters gegen HKE.

Abbildung 3.1: Scatterplot Alter - HKE



Mit diesem Scatterplot kann keine Aussage über die Form und Stärke der Beziehung HKE - Alter getroffen werden, da die Punkte entweder auf der Linie $HKE = 0$ bzw. $HKE = 1$ liegen. Es ist jedoch eine Verschiebung der Punkte bei $HKE = 1$ zu höheren Alterswerten im Vergleich zu $HKE = 0$ zu beobachten.

Verwendet man die angegebenen Altersklassen und berechnet für jede Altersklasse die bedingte relative Häufigkeit für das Auftreten von HKE (fett markiert in der Tab. 3.6), so ergibt sich die Kontingenztafel Tabelle 3.6.

Aus dieser Tabelle ist ersichtlich, dass mit höherem Alter die bedingte relative Häufigkeit des Auftretens von HKE größer wird. Dies lässt sich auch grafisch in einem Scatterplot der bedingten relativen Häufigkeiten für das Auftreten von HKE über den Altersklassen verdeutlichen (siehe Abb. 3.2).

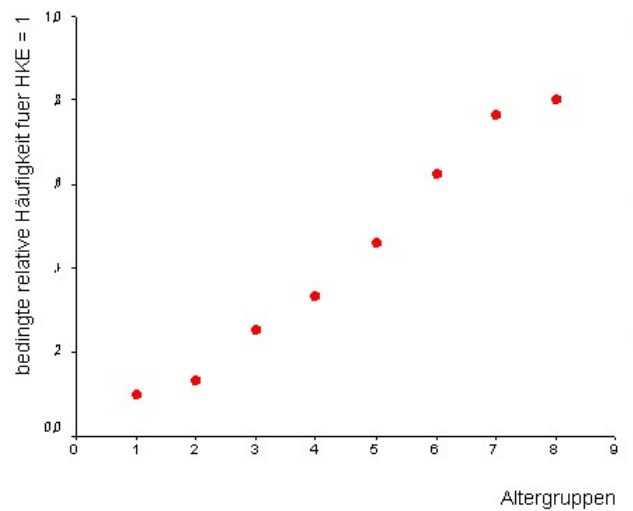
Eine gedachte Kurve durch die Punktwolke geht mit jüngerem Alter allmählich gegen Null und mit zunehmenden Alter allmählich gegen Eins und weist eine S-förmige Gestalt auf. Eine solche S-förmige Gestalt findet man bei einigen bekannten Verteilungsfunktionen (z.B. der Normalverteilung). Eine lineare Beziehung erscheint somit nicht angebracht.

3.1 Das Logit Modell (logistische Regression)

Tabelle 3.6: Altersgruppen und bedingte relative Häufigkeit des Auftretens von HKE

Altersgruppe		HKE		Total
		nein = 0	ja = 1	
20-29	Count	9	1	10
	% within Altersgruppe	90,0%	10,0%	100,0%
30-34	Count	13	2	15
	% within Altersgruppe	86,7%	13,3%	100,0%
35-39	Count	9	3	12
	% within Altersgruppe	75,0%	25,0%	100,0%
40-44	Count	10	5	15
	% within Altersgruppe	66,7%	33,3%	100,0%
45-49	Count	7	6	13
	% within Altersgruppe	53,8%	46,2%	100,0%
50-54	Count	3	5	8
	% within Altersgruppe	37,5%	62,5%	100,0%
55-59	Count	4	13	17
	% within Altersgruppe	23,5%	76,5%	100,0%
60-69	Count	2	8	10
	% within Altersgruppe	20,0%	80,0%	100,0%
Total	Count	57	43	100
	% within Altersgruppe	57,0%	43,0%	100,0%

Abbildung 3.2: Scatterplot Alter - bedingte relative Häufigkeit für HKE = 1



Ein weiteres Argument gegen eine lineare Beziehung ist mit der Tatsache gegeben, dass in der oberen erwähnten Identitätsfunktion $\pi_k = \mathbf{x}_k^T \boldsymbol{\beta}$ die rechte Seite in der Regel Werte im Bereich vom $-\infty$ bis $+\infty$ annehmen kann, während der Wertebereich der Wahrscheinlichkeit auf der linken Seite auf $(0;1)$ beschränkt ist.

Des Weiteren ist die Voraussetzung normalverteilter Fehlervariablen, wie sie für ein lineares Regressionsmodell gestellt wird, nicht haltbar. In einem Modell mit einer dichotomen response Variablen

$$y_k = E(Y_k | \mathbf{x}_k) + \varepsilon_k = \pi_k + \varepsilon_k \quad (3.12)$$

kann auch der Fehlerterm nur zwei mögliche Werte annehmen:

für $y_k = 1$ nimmt ε_k den Wert $\varepsilon_k = 1 - \pi_k$ mit einer Wahrscheinlichkeit von π_k an und für $y_k = 0$ den Wert $\varepsilon_k = -\pi_k$ mit einer Wahrscheinlichkeit $1 - \pi_k$:

y_k	$\varepsilon_k = y_k - \pi_k$	$P(\varepsilon_k)$
1	$1 - \pi_k$	π_k
0	$-\pi_k$	$1 - \pi_k$

Für den Erwartungswert von ε_k folgt:

$$E(\varepsilon_k) = (1 - \pi_k) \cdot \pi_k + (-\pi_k) \cdot (1 - \pi_k) = 0$$

und für die Varianz

$$Var(\varepsilon_k) = (1 - \pi_k)^2 \cdot \pi_k + (-\pi_k)^2 \cdot (1 - \pi_k) = \pi_k \cdot (1 - \pi_k)$$

Als Schlussfolgerung ergibt sich, dass eine andere Spezifikation der link Funktion (3.11) vorzunehmen ist, wofür es verschiedene Möglichkeiten gibt. Hier soll der Fall betrachtet werden, dass zur Absicherung von $0 \leq \pi_k \leq 1$ die Modellierung von π_k als Verteilungsfunktion erfolgt:

$$\pi_k = g^{-1}(\eta_k) = \int_{-\infty}^{\eta_k} f(\eta_k) d\eta_k, \quad \text{mit } f(\eta_k) \geq 0, \quad \int_{-\infty}^{\infty} f(\eta_k) d(\eta_k) = 1. \quad (3.13)$$

3.1 Das Logit Modell (logistische Regression)

Die Dichtefunktion $f(\eta_k)$ wird oftmals auch als Toleranz-Verteilung bezeichnet. Für binomialverteilte Daten wird häufig die logistische Verteilung verwendet, da sie viele medizinische, biologische und sozioökonomische Problemstellungen angemessen beschreibt. Die Dichtefunktion der logistischen Verteilung ist gegeben mit

$$f(\eta_k) = \frac{e^{\eta_k}}{(1 + e^{\eta_k})^2} \quad (3.14)$$

und die zugehörige Verteilungsfunktion mit

$$\pi_k = g^{-1}(\eta) = \frac{1}{1 + e^{-\eta_k}} = \frac{e^{\eta_k}}{1 + e^{\eta_k}} \quad (3.15)$$

Abbildung 3.3 und 3.4 zeigen die Dichtefunktion (3.14) bzw. die Verteilungsfunktion (3.15) für η_k im Bereich - 5 bis 5. Wie zu erkennen ist, weist die Gestalt der logistischen Verteilung große Ähnlichkeit mit der Normalverteilung auf. Abweichungen gibt es vor allem an den Enden der Kurven.

Abbildung 3.3: Dichtefunktion der logistischen Verteilung

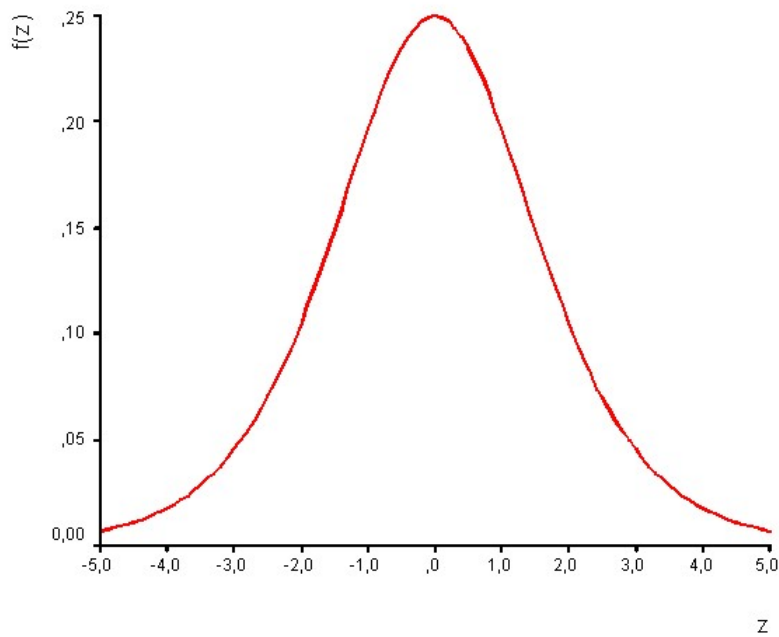
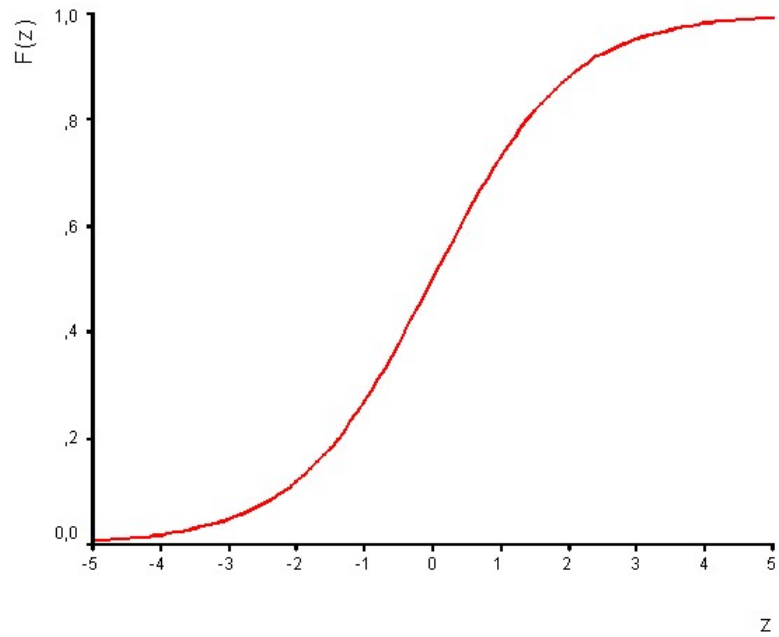


Abbildung 3.4: Verteilungsfunktion der logistischen Verteilung



(3.15) nach e_k^η aufgelöst ergibt:

$$\frac{\pi_k}{1 - \pi_k} = e^{\eta_k} = \exp(\eta_k). \quad (3.16)$$

$\pi_k/(1 - \pi_k)$ als Verhältnis der Wahrscheinlichkeit eines Erfolges zur Wahrscheinlichkeit eines Misserfolges wird als odds (Chancen) eines Erfolges bezeichnet. Aus (3.16) folgt unmittelbar als logarithmierte odds (oder kurz: log odds):

$$\log\left(\frac{\pi_k}{1 - \pi_k}\right) = \eta_k. \quad (3.17)$$

Dies ist die gesuchte Spezifikation der link Funktion (3.11) eines logit Modells:

$$g(\pi_k) = \eta_k = \log[\pi_k/(1 - \pi_k)] = \mathbf{x}_k^T \boldsymbol{\beta} = \sum_j x_{kj} \beta_j \quad (3.18)$$

und wird oft als logit Funktion oder kurz logit bezeichnet. (3.15) ist die inverse Funktion von (3.18).

Die logit Funktion $g(\pi_k)$ ist linear in den Parametern und nimmt je nach den Werten der erklärenden X-Variablen Werte im Bereich von $-\infty$ bis $+\infty$ an.

Der Parameter β_j gibt die Veränderung in den log odds an, wenn die Variable X_j um eine Einheit erhöht wird und alle anderen erklärenden Variablen konstant bleiben. Gleichbedeutend ist die Interpretation, dass eine Erhöhung von X_j um eine Einheit (bei Konstanz der Werte aller anderen X-Variablen) die odds eines Erfolgs multiplikativ mit dem Faktor $\exp(\beta_j)$ verändert (siehe Formel (3.16)).

Die Interpretation bezüglich der Wahrscheinlichkeit eines Erfolges π_k ist dagegen nicht unmittelbar gegeben, da der Effekt einer Veränderung um eine Einheit in X_j auf π_k von den Werten aller X-Variablen abhängt.

3.2 Die Schätzung des Modells und Hypothesenprüfung

In der log-Likelihood-Funktion

$$l(\pi_k; y_k) = l_k = y_k \log \left(\frac{\pi_k}{1 - \pi_k} \right) + n_k \log(1 - \pi_k) + \log \left(\frac{n_k}{y_k} \right)$$

wird $\log[\pi_k/(1 - \pi_k)]$ durch (3.18) und $\log(1 - \pi_k)$ unter Verwendung von (3.15) durch

$$\begin{aligned} \log(1 - \pi_k) &= \log \left(1 - \frac{e^{\eta_k}}{1 + e^{\eta_k}} \right) = \log \left(\frac{1 + e^{\eta_k} - e^{\eta_k}}{1 + e^{\eta_k}} \right) = \log \left(\frac{1}{1 + e^{\eta_k}} \right) \\ &= -\log(1 + e^{\eta_k}) = -\log \left(1 + \exp \left(\sum_{j=1}^J x_{kj} \beta_j \right) \right) \end{aligned}$$

ersetzt. Dies führt zu:

$$l(\beta; y_k) = l_k = \sum_{j=1}^J y_k x_{kj} \beta_j - n_k \log \left[1 + \exp \left(\sum_{j=1}^J x_{kj} \beta_j \right) \right] + \log \left(\frac{n_k}{y_k} \right). \quad (3.19)$$

Die gemeinsame log-Likelihood-Funktion resultiert damit in:

$$\begin{aligned} l(\beta; \mathbf{y}) &= \sum_{k=1}^K l_k \\ &= \sum_{k=1}^K \left\{ \sum_{j=1}^J y_k x_{kj} \beta_j - n_k \log \left[1 + \exp \left(\sum_{j=1}^J x_{kj} \beta_j \right) \right] + \log \left(\frac{n_k}{y_k} \right) \right\}, \end{aligned} \quad (3.20)$$

3 Modellierung binärer Daten

wobei $\boldsymbol{\beta}^T = (\beta_1 \dots \beta_J)$ und $\mathbf{y}^T = (y_1 \dots y_K)$ sind und nunmehr $l(\boldsymbol{\beta}; \mathbf{y})$ statt $l(\boldsymbol{\pi}; \mathbf{y})$ geschrieben wird, da die unbekannten $\boldsymbol{\pi}$ durch die unbekannten $\boldsymbol{\beta}$ ersetzt wurden.

Die Maximum-Likelihood-Schätzer der Parameter β_j und damit der Wahrscheinlichkeiten π_k erhält man durch Maximierung der log-Likelihood-Funktion (3.20). Ableitung von (3.20) nach β_j ($j = 1, \dots, J$) ergibt:

$$\begin{aligned} \frac{\partial l}{\partial \beta_j} &= \sum_{k=1}^K \frac{\partial l_k}{\partial \beta_j} = \sum_{k=1}^K \left\{ y_k x_{kj} - n_k x_{kj} \left[\frac{\exp \left(\sum_{j=1}^J x_{kj} \beta_j \right)}{1 + \exp \left(\sum_{j=1}^J x_{kj} \beta_j \right)} \right] \right\} \\ &= \sum_{k=1}^K x_{kj} (y_k - n_k \pi_k). \end{aligned} \quad (3.21)$$

Zu diesem Ergebnis gelangt man analog der im Abschnitt 2.3 beschriebenen Vorgehensweise (siehe Formel (2.48)). Danach ergibt sich die Ableitung von l_k nach β_j als

$$\frac{\partial l_k}{\partial \beta_j} = \frac{\partial l_k}{\partial \pi_k} \cdot \frac{\partial \pi_k}{\partial \eta_k} \cdot \frac{\partial \eta_k}{\partial \beta_j}. \quad (3.22)$$

Den 1. Faktor auf der rechten Seite von (3.22) erhält man als Ableitung von $l(\pi_k; y_k)$

$$\begin{aligned} l(\pi_k; y_k) &= y_k \log \left(\frac{\pi_k}{1 - \pi_k} \right) + n_k \log(1 - \pi_k) + \log \binom{n_k}{y_k} \\ &= y_k \log \pi_k - y_k \log(1 - \pi_k) + n_k \log(1 - \pi_k) + \log \binom{n_k}{y_k} \end{aligned}$$

nach π_k :

$$\frac{\partial l_k}{\partial \pi_k} = \frac{y_k}{\pi_k} - (-1) \frac{y_k}{1 - \pi_k} + (-1) \frac{n_k}{1 - \pi_k} = \frac{y_k - n_k \pi_k}{\pi_k (1 - \pi_k)}. \quad (3.23)$$

Den 2. Faktor auf der rechten Seite von (3.22) erhält man als Ableitung der Verteilungsfunktion der logistischen Verteilung (3.15) nach η_k unter Berücksichtigung von $\eta_k = \sum_j x_{kj} \beta_j$, was im Ergebnis zur Dichtefunktion der logistischen Verteilung (3.14) führt:

$$\frac{\partial \pi_k}{\partial \eta_k} = \frac{\exp \left(\sum_{j=1}^J x_{kj} \beta_j \right)}{\left[1 + \exp \left(\sum_{j=1}^J x_{kj} \beta_j \right) \right]^2}. \quad (3.14)$$

Der 3. Faktor auf der rechten Seite von (3.22) beinhaltet die Ableitung von $\eta_k = \sum_j x_{kj}\beta_j$ nach β_j :

$$\partial\eta_k/\partial\beta_j = x_{kj}. \quad (3.24)$$

Fügt man (3.23), (3.14) und (3.24) gemäß (3.22) zusammen und berücksichtigt

$$1 - \pi_k = 1 - \frac{\exp\left(\sum_{j=1}^J x_{kj}\beta_j\right)}{1 + \exp\left(\sum_{j=1}^J x_{kj}\beta_j\right)} = \frac{1}{1 + \exp\left(\sum_{j=1}^J x_{kj}\beta_j\right)} \quad (3.25)$$

sowie

$$\begin{aligned} \frac{\exp\left(\sum_{j=1}^J x_{kj}\beta_j\right)}{\left[1 + \exp\left(\sum_{j=1}^J x_{kj}\beta_j\right)\right]^2} &= \frac{\exp\left(\sum_{j=1}^J x_{kj}\beta_j\right)}{1 + \exp\left(\sum_{j=1}^J x_{kj}\beta_j\right)} \cdot \frac{1}{1 + \exp\left(\sum_{j=1}^J x_{kj}\beta_j\right)} \\ &= \pi_k \cdot (1 - \pi_k), \end{aligned} \quad (3.26)$$

so erhält man (3.21):

$$\begin{aligned} \frac{\partial l}{\partial\beta_j} &= \sum_{k=1}^K \left(\frac{y_k - n_k\pi_k}{\pi_k(1 - \pi_k)} \right) \cdot \frac{\exp\left(\sum_{j=1}^J x_{kj}\beta_j\right)}{\left[1 + \exp\left(\sum_{j=1}^J x_{kj}\beta_j\right)\right]^2} \cdot x_{kj} \\ &= \sum_{k=1}^K \left(\frac{y_k - n_k\pi_k}{\pi_k(1 - \pi_k)} \right) \cdot \pi_k(1 - \pi_k) \cdot x_{kj} = \sum_{k=1}^K x_{kj}(y_k - n_k\pi_k). \end{aligned} \quad (3.21)$$

(3.21) gleich Null gesetzt, führt zu nichtlinearen Gleichungen. Diese müssen iterativ bei Vorgabe von Startwerten für β gelöst werden, da alle β_j noch in (3.21) enthalten sind. Für die Elemente $\mathfrak{S}_{jj\star}$ der Informationsmatrix \mathfrak{S} erhält man gemäß (2.22) unter Berücksichtigung, dass der Erwartungswert der binomialverteilten Zufallsvariablen Y_k gleich $E(Y_k) = n_k\pi_k$ und die Varianz gleich $Var(Y_k) = E[(Y_k - E(Y_k))^2] = n_k\pi_k(1 - \pi_k)$

3 Modellierung binärer Daten

sind, als Beitrag jedes Y_k zu $\mathfrak{S}_{jj\star}$

$$\begin{aligned} E\left(\frac{\partial l_k}{\partial \beta_j} \frac{\partial l_k}{\partial \beta_{j\star}}\right) &= E(x_{kj}(Y_k - n_k \pi_k) \cdot x_{kj\star}(Y_k - n_k \pi_k)) \\ &= x_{kj} x_{kj\star} \cdot E[(Y_k - n_k \pi_k)^2] \\ &= n_k \pi_k (1 - \pi_k) x_{kj} x_{kj\star} \end{aligned}$$

und damit

$$\mathfrak{S}_{jj\star} = E(U_j U_{j\star}) = E\left(\frac{\partial l}{\partial \beta_j \partial \beta_{j\star}}\right) = \sum_{k=1}^K n_k \pi_k (1 - \pi_k) x_{kj} x_{kj\star}. \quad (3.27)$$

Die Gewichtsmatrix \mathbf{W} in

$$\mathfrak{S} = \mathbf{X}^T \mathbf{W} \mathbf{X} \quad (2.57)$$

ist somit eine $K \times K$ Diagonalmatrix mit den Elementen

$$w_{kk} = n_k \pi_k (1 - \pi_k).$$

Startwerte $\boldsymbol{\beta}_{(0)}$ gegeben, können $\hat{\pi}_{(0)}$ und $\hat{\eta}_{(0)}$ und daraus die angepassten Werte von \mathbf{z} in

$$\mathbf{X}^T \mathbf{W} \mathbf{X} \mathbf{b}_{(m)} = \mathbf{X}^T \mathbf{W} \mathbf{z} \quad (2.64)$$

berechnet werden, wobei sich die Elemente von \mathbf{z} zu

$$z_k = \hat{\eta}_k + \frac{y_k - n_k \hat{\pi}_k}{n_k} \frac{\partial \eta_k}{\partial \pi_k} \quad (3.28)$$

ergeben.

Die geschätzten Werte von Y_k ergeben sich zu:

$$\hat{y}_k = n_k \hat{\pi}_k \quad (3.29)$$

und die geschätzte Varianz-Kovarianz-Matrix für \mathbf{b} ist $[\mathfrak{S}(\mathbf{b})]^{-1}$.

Die log-Likelihood-Funktion (3.9) nimmt ihren maximalen Wert an der Stelle $\hat{\pi}_k$ an:

$$\begin{aligned} l(\hat{\boldsymbol{\pi}}; \mathbf{y}) &= \sum_{k=1}^K \left[y_k \log \left(\frac{\hat{\pi}_k}{1 - \hat{\pi}_k} \right) + n_k \log(1 - \hat{\pi}_k) + \log \binom{n_k}{y_k} \right] \\ &= \sum_{k=1}^K \left[y_k \log \hat{\pi}_k + (n_k - y_k) \log(1 - \hat{\pi}_k) + \log \binom{n_k}{y_k} \right] \end{aligned} \quad (3.30)$$

Beim maximalen (saturierten) Modell, in dem keine Konstante und keine erklärenden X-Variablen enthalten sind, sind die π_k die unmittelbar zu schätzenden Parameter, d.h., die Anzahl der zu schätzenden Parameter beträgt K und die geschätzten Y-Werte fallen mit den beobachteten Werten zusammen. Ausgehend von der log-Likelihood-Funktion

$$l(\boldsymbol{\pi}; \mathbf{y}) = \sum_{k=1}^K \left[y_k \log \pi_k - y_k \log(1 - \pi_k) + n_k \log(1 - \pi_k) + \log \binom{n_k}{y_k} \right] \quad (3.9)$$

ergeben sich die Schätzungen zu:

$$\begin{aligned} \frac{\partial l(\boldsymbol{\pi}; \mathbf{y})}{\partial \pi_k} &= \frac{y_k}{\pi_k} + \frac{y_k}{1 - \pi_k} - \frac{n_k}{1 - \pi_k} \stackrel{!}{=} 0 \\ \frac{y_k(1 - \hat{\pi}_k) + y_k \hat{\pi}_k - n_k \hat{\pi}_k}{\hat{\pi}_k(1 - \pi_k)} &= 0 \\ y_k - n_k \hat{\pi}_k &= 0 \\ \hat{\pi}_k &= \frac{y_k}{n_k} \end{aligned} \quad (3.31)$$

d.h., sie sind die beobachteten Anteile eines Erfolges in der k-ten Gruppe. Für dieses maximale Modell nimmt die log-Likelihood-Funktion ihren maximalen Wert bei

$$l(\hat{\boldsymbol{\pi}}_{max}; \mathbf{y}) = \sum_{k=1}^K \left[y_k \log \left(\frac{y_k}{n_k} \right) + (n_k - y_k) \log \left(1 - \frac{y_k}{n_k} \right) + \log \binom{n_k}{y_k} \right] \quad (3.32)$$

an.

Damit ist die Deviance gemäß (2.97)

$$\begin{aligned} D &= 2[l(\hat{\boldsymbol{\pi}}_{max}; \mathbf{y}) - l(\hat{\boldsymbol{\pi}}; \mathbf{y})] \\ &= 2 \sum_{k=1}^K \left[y_k \left(\log \left(\frac{y_k}{n_k} \right) - \log \hat{\pi}_k \right) + (n_k - y_k) \left(\log \left(1 - \frac{y_k}{n_k} \right) - \log(1 - \hat{\pi}_k) \right) \right] \\ &= 2 \sum_{k=1}^K \left[y_k \log \left(\frac{y_k}{n_k \hat{\pi}_k} \right) + (n_k - y_k) \log \left(\frac{n_k - y_k}{n_k - n_k \hat{\pi}_k} \right) \right] \\ &= 2 \sum_{k=1}^K \left[y_k \log \left(\frac{y_k}{\hat{y}_k} \right) + (n_k - y_k) \log \left(\frac{n_k - y_k}{n_k - \hat{y}_k} \right) \right], \end{aligned} \quad (3.33)$$

worin $n_k \hat{\pi}_k = \hat{y}_k$ ist. Letzterer Ausdruck zeigt deutlich, dass die Deviance die Beobachtungen y_k mit den geschätzten Werten \hat{y}_k aus dem interessierenden Modell vergleicht.

Die Deviance wird in der Literatur oftmals auch in folgender Form geschrieben:

$$D = -2 \log \left(\frac{L(\hat{\boldsymbol{\pi}}; \mathbf{y})}{L(\hat{\boldsymbol{\pi}}_{max}; \mathbf{y})} \right), \quad (3.34)$$

worin $L(\hat{\boldsymbol{\pi}}; \mathbf{y})$ der Wert der Likelihood-Funktion (3.8) des Modells mit J erklärenden X -Variablen an der Stelle $\hat{\boldsymbol{\pi}}$ und $L(\hat{\boldsymbol{\pi}}_{max}; \mathbf{y})$ der Wert der Likelihood-Funktion (3.8) des maximalen Modells an der Stelle $\hat{\boldsymbol{\pi}}_{max}$ sind. Es ist

$$\begin{aligned} D &= -2 \log \left(\frac{L(\hat{\boldsymbol{\pi}}; \mathbf{y})}{L(\hat{\boldsymbol{\pi}}_{max}; \mathbf{y})} \right) \\ &= -2[l(\hat{\boldsymbol{\pi}}; \mathbf{y}) - l(\hat{\boldsymbol{\pi}}_{max}; \mathbf{y})] \\ &= 2[l(\hat{\boldsymbol{\pi}}_{max}; \mathbf{y}) - l(\hat{\boldsymbol{\pi}}; \mathbf{y})]. \end{aligned}$$

Setzt man voraus, dass die Anzahl der Gruppen mit K festgelegt ist und $K < n$ gilt, so tendieren die n_k größer zu werden, wenn n größer wird.

D folgt unter dieser Voraussetzung approximativ einer Chi-Quadrat-Verteilung mit $f = K - J$ Freiheitsgraden, wenn das Modell korrekt ist. J ist dabei die Anzahl der zu schätzenden Parameter bzw. die Anzahl der X -Variablen. Ist eine Konstante β_0 in (3.18) eingeschlossen, so impliziert dies eine Variable X_1 mit $\mathbf{x}_1 = \mathbf{1}$. Die Approximation an die Chi-Quadrat-Verteilung ist jedoch schlecht, wenn einige n_k klein sind.

Für den Fall ungruppiertter Daten, d.h. $K = n$ und $n_k = 1$, folgt D noch nicht einmal approximativ einer Chi-Quadrat-Verteilung (da im maximalen Modell die Anzahl der zu schätzenden Parameter gegen unendlich geht, wenn n gegen unendlich geht). In diesem Fall sagt die Deviance nichts über die Güte des Modells aus. Für das interessierende Modell nimmt die log-Likelihood-Funktion ihren maximalen Wert an der Stelle $\hat{\boldsymbol{\pi}}$ an:

$$l(\hat{\boldsymbol{\pi}}; \mathbf{y}) = \sum_{k=1}^K \left[y_k \log \left(\frac{\hat{\pi}_k}{1 - \hat{\pi}_k} \right) + \log(1 - \hat{\pi}_k) \right]. \quad (3.35)$$

Für das maximale Modell ergibt sich aus (3.31) wegen $n_k = 1$:

$$\hat{\pi}_k = y_k \quad \text{und} \quad l(\hat{\boldsymbol{\pi}}_{max}; \mathbf{y}) = 0,$$

da $y_k \log(y_k) = 0$ und $(1 - y_k) \log(1 - y_k) = 0$ für die beiden möglichen Werte y_k (1 oder 0) ist. Somit ist die Deviance

$$D = -2 \sum_{k=1}^K \left[y_k \log \left(\frac{\hat{\pi}_k}{1 - \hat{\pi}_k} \right) + \log(1 - \hat{\pi}_k) \right]. \quad (3.36)$$

Im Gegensatz zur Deviance (3.33), bei der die Beobachtungen y_k mit den geschätzten Werten \hat{y}_k aus dem interessierenden Modell verglichen werden, hängt die Deviance (3.36)

nur von den geschätzten Wahrscheinlichkeiten $\hat{\pi}_k$ ab und sagt nichts über die Anpassung der beobachteten zu den geschätzten Werten aus.

Grundsätzlich sollte deshalb die Deviance nicht als Maß der Güte des Modells (goodness-of-fit) verwendet werden. Vielmehr sollten zwei Modelle mit unterschiedlicher Parameterzahl verglichen werden. Unter der Nullhypothese H_0 wird ein Modell M_0 mit J erklärenden X -Variablen und damit J zu schätzenden Parametern spezifiziert, dass gegen ein Modell M_1 mit $J + t$ erklärenden X -Variablen und somit $J + t$ zu schätzenden Parametern unter der Alternativhypothese H_1 getestet wird. Bezeichnet man mit D_0 die zum Modell M_0 gehörende Deviance und mit D_1 die zum Modell M_1 gehörende Deviance, so ergibt sich eine Reduktion in der Deviance von

$$\begin{aligned}\Delta D &= D_0 - D_1 = 2[l(\hat{\boldsymbol{\pi}}_{max}; \mathbf{y}) - l_0(\hat{\boldsymbol{\pi}}; \mathbf{y})] - 2[l(\hat{\boldsymbol{\pi}}_{max}; \mathbf{y}) - l_1(\hat{\boldsymbol{\pi}}; \mathbf{y})] \\ &= 2[l_1(\hat{\boldsymbol{\pi}}; \mathbf{y}) - l_0(\hat{\boldsymbol{\pi}}; \mathbf{y})].\end{aligned}\tag{3.37}$$

Für die Freiheitsgrade ergibt sich:

$$f = (K - J) - [K - (J + t)] = t.\tag{3.38}$$

ΔD ist approximativ chi-quadrat-verteilt mit $f = t$ Freiheitsgraden. Dies gilt auch für den Fall $K = n$ mit $n_k = 1$, da $l(\hat{\boldsymbol{\pi}}_{max}; \mathbf{y})$ verschwindet.

Ein derartiges Vorgehen, zwei verschiedene Modelle gegeneinander zu prüfen, wird im allgemeinen auch bei der Verwendung statistischer Software realisiert, wobei das Modell unter der Nullhypothese nur die Konstante und das Modell unter der Alternativhypothese ein Modell mit J erklärenden X -Variablen sind.

Ein anderer Test ist der Score-Test, der auf der Verteilung der Scores basiert (siehe Abschnitt 2.4). Auf die Darstellung soll hier wegen der umfangreichen Formeln verzichtet werden. Der Vorteil der Verwendung des Score-Tests besteht darin, dass er nicht die explizite Berechnung der Parameterschätzungen erfordert, was zu einer erheblichen Rechenzeitverkürzung vor allem bei großen Stichprobenumfängen und logistischen Regressionsmodellen mit vielen X -Variablen führt. Dieser Test wird deshalb oftmals für die Sektion von erklärenden X -Variablen verwendet.

Um diejenigen erklärenden Variablen unter den gegebenen X-Variablen herauszufinden, die einen wesentlichen Einfluss ausüben, kann man sich verschiedener Selektionsverfahren bedienen, z.B. Forward-Selection und Backward-Selection, wobei als Kriterium für die Aufnahme bzw. den Ausschluss einer Variablen u.a. die Wald-Statistik oder die Veränderung im maximalen Wert der log-Likelihood-Funktion verwendet werden können. Die Verwendung verschiedener Selektionsmethoden auf die gleichen X-Variablen kann durchaus zu einer unterschiedlichen Auswahl von X-Variablen führen. Keine dieser Selektionsmethoden kann als „beste“ angesehen werden. Entscheidend für die letztendliche Wahl des Modells ist neben den statistischen Kriterien seine Interpretierbarkeit.

Zwei Statistiken, die eine Aussage über den Anteil der erklärten Variation in der logistischen Regression treffen, sind Cox&Snell R_{CS}^2 und Nagelkerke R_N^2 .

Cox&Snell R_{CS}^2 ist definiert als:

$$R_{CS}^2 = 1 - \left[\frac{L_0}{L_1} \right]^{\frac{2}{n}} \quad (3.39)$$

und Nagelkerke R_N^2 als

$$R_N^2 = \frac{R_{CS}^2}{1 - L_0^{2/n}}. \quad (3.40)$$

Darin sind:

- L_0 der Wert der Likelihood-Funktion für das Logit Model mit nur einer Konstanten,
- L_1 der Wert der Likelihood-Funktion für das interessierende Logit-Modell mit einer bestimmten Anzahl erklärenden X-Variablen.

Der Nachteil von Cox&Snell R_{CS}^2 besteht darin, dass der maximale Wert von 1 nicht erreicht werden kann. Nagelkerke schlug deshalb seine Modifikation von Cox&Snell R_{CS}^2 vor.

Ein Test zur Prüfung der einzelnen Parameter β_j ($j = 1, \dots, J$) auf Signifikanz ist der Wald-Test. Die Teststatistik für die Prüfung der Nullhypothese $H_0 : \beta_j = 0$ ist:

$$W = [b_j / s(b_j)]^2, \quad (3.41)$$

worin $s(b_j)$ der Standardfehler des geschätzten Regressionskoeffizienten b_j ist, den man als Quadratwurzel aus dem j-ten Diagonalelement der inversen Informationsmatrix erhält. W ist unter H_0 approximativ chi-quadrat-verteilt mit $f = 1$ Freiheitsgrad.

Ebenso wie bei der linearen Regressionsschätzung sollte eine Modellprüfung vorgenommen werden, wofür u.a. folgende Diagnostikkriterien zur Verfügung stehen:

- die Residuen

$$\hat{e}_k = y_k - \hat{\pi}_k \quad (3.42)$$

- die standardisierten Residuen

$$r_k = \frac{y_k - \hat{\pi}_k}{\sqrt{\hat{\pi}_k(1 - \hat{\pi}_k)}} \quad (3.43)$$

- die logit Residuen

$$\tilde{e}_k = \frac{y_k - \hat{\pi}_k}{\hat{\pi}_k(1 - \hat{\pi}_k)} \quad (3.44)$$

- die Deviance Residuen

$$d_k = \begin{cases} \sqrt{2y_k \log\left(\frac{y_k}{\hat{\pi}_k}\right)} & \text{für } y_k = 1 \\ \sqrt{2(1 - y_k) \log\left(\frac{1 - y_k}{1 - \hat{\pi}_k}\right)} & \text{für } y_k = 0 \end{cases} \quad (3.45)$$

- die DFBETAS

Sie reflektieren die Veränderung in den einzelnen geschätzten Parametern, wenn der Fall k aus der Analyse ausgeschlossen wird:

$$dfbeta_j = (b_j - b_{j(k)}), \quad (3.46)$$

worin $b_{j(k)}$ der j-te geschätzte Parameter bei Ausschluss des k-ten Falles ist. Ein grosser absoluter Wert von $dfbeta_j$ zeigt an, dass der k-te Fall den Koeffizienten beeinflusst.

- der Leverage

Da die Inverse der Informationsmatrix die Varianz-Kovarianz-Matrix von \mathbf{b} ist, kann (analog zur Projektions- oder hat-Matrix in der linearen Regression) ein Maß angegeben werden, dass den Einfluss (leverage) der erklärenden Variablen auf die Schätzung angibt:

$$\mathbf{H} = \mathbf{W}^{1/2}(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2} \quad (3.47)$$

Die Diagonalelemente h_{kk} ($k = 1, \dots, K$) liegen im Intervall $[0,1]$ und der Mittelwert der Diagonalelemente ist J/n . Sie geben den Einfluss der Beobachtungen der erklärenden Variablen des k -ten Falles auf die geschätzten Parameter an und sind deshalb zum Auffinden beeinflussender Beobachtungen nützlich. Die hat-Werte hängen jedoch von $\hat{\pi}_k$ und von \mathbf{x}_k ab. Im Falle einer Beobachtung für jedes \mathbf{x}_k ist $h_{kk} \leq 1$, sonst gilt $h_{kk} \leq 1/n_k$ (n_k ist die Häufigkeit des Auftretens von \mathbf{x}_k).

- Cook's Distanz

Cook's Distanz misst den Einfluss des Ausschlusses des k -ten Falles auf die Schätzung des Modells. Dieses Maß ist gegeben mit:

$$C_k = r_k^2 \cdot h_{kk} / (1 - h_{kk})^2, \quad (3.48)$$

d.h., es hängt von dem standardisierten Residual r_k und dem Leverage h_{kk} des jeweiligen Falles ab.

Vorhersage (prediction):

Das geschätzte logistische Modell wird verwendet, um die Wahrscheinlichkeit für jede Gruppe von Fällen mit identischen \mathbf{x}_k gemäß (3.15) zu schätzen. Vereinbart man einen Trennpunkt (cut point) z.B. in der Weise, dass $y = 0$ für $\hat{\pi}_k \leq 0,5$ und $y = 1$ für $\hat{\pi}_k > 0,5$ gesetzt wird, dann lässt sich eine Klassifikationstabelle der Vorhersage in der folgenden Form erstellen.

\hat{m}_{11} , \hat{m}_{10} , \hat{m}_{01} und \hat{m}_{00} sind darin die geschätzten absoluten Häufigkeiten der Zellen der Vorhersagetabelle. Ihre Summe muss den geschätzten Gesamtstichprobenumfang n reproduzieren.

Tabelle 3.7: Klassifikationstabelle der Vorhersage

Beobachtet	Vorhersage		Anteil korrekter Vorhersage
	y = 1	y = 0	
y = 1	\hat{m}_{11}	\hat{m}_{10}	$\hat{m}_{11}/(\hat{m}_{11} + \hat{m}_{10})$
y = 0	\hat{m}_{01}	\hat{m}_{00}	$\hat{m}_{00}/(\hat{m}_{01} + \hat{m}_{00})$
Overall			$(\hat{m}_{11} + \hat{m}_{00})/n$

3.3 Zur Interpretation der Ergebnisse des logistischen Modells

Allgemein kann der Effekt einer erklärenden X-Variablen auf die Wahrscheinlichkeit eines Erfolges über (3.15)

$$\hat{\pi} = \frac{e^{b_0 + b_1 x_1 + \dots + b_J x_J}}{1 + e^{b_0 + b_1 x_1 + \dots + b_J x_J}} \quad (3.15)$$

wie folgt beschrieben werden:

Ist der Parameter b_j positiv, dann erhöht sich die link Funktion $g(\hat{\pi}_k) = \log[\hat{\pi}_k/(1 - \hat{\pi}_k)]$ bei Erhöhung der Variablen X_j um eine Einheit und Konstanz aller anderen X-Variablen um b_j und die odds $\hat{\pi}_k/(1 - \hat{\pi}_k)$ erhöhen sich multiplikativ um $\exp(b_j)$. Je höher jedoch $\exp(\bullet)$ ist, desto höher ist die Wahrscheinlichkeit eines Erfolges.

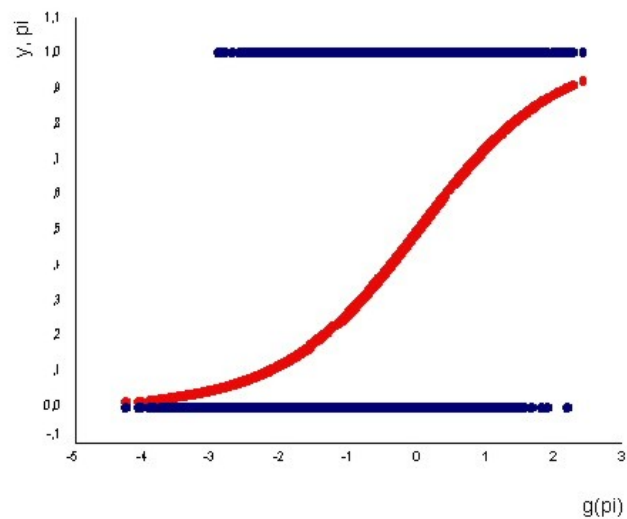
Ist der Parameter b_j negativ, dann sinkt die link Funktion $g(\hat{\pi}_k) = \log[\hat{\pi}_k/(1 - \hat{\pi}_k)]$ bei Erhöhung der Variablen X_j um eine Einheit und Konstanz aller anderen X-Variablen um b_j und die odds $\hat{\pi}_k/(1 - \hat{\pi}_k)$ sinken multiplikativ um $\exp(b_j)$. Je niedriger jedoch $\exp(\bullet)$ ist, desto kleiner ist die Wahrscheinlichkeit eines Erfolges.

Diese tendenziellen Veränderungen verdeutlichen nochmals Tab. 3.8 und Abb. 3.5.

In der Abb. 3.5 sind auf der Abszisse $g(\hat{\pi}_k) = b_0 + b_1 x_1 + \dots + b_J x_J$ und auf der Ordinate die Werte für Y (entweder 1 oder 0) bzw. die geschätzte Wahrscheinlichkeit $\hat{\pi}_k$ abgetragen.

Tabelle 3.8: log odds, odds und Wahrscheinlichkeit

$g(\hat{\pi}_k) = \log[\hat{\pi}_k/(1 - \hat{\pi}_k)]$ $= b_0 + b_1x_1 + \dots + b_Jx_J$	$\hat{\pi}_k/(1 - \hat{\pi}_k) =$ $\exp(b_0 + b_1x_1 + \dots + b_Jx_J)$	$\hat{\pi}_k = \frac{\exp(b_0 + b_1x_1 + \dots + b_Jx_J)}{[1 + \exp(b_0 + b_1x_1 + \dots + b_Jx_J)]}$
\vdots	\vdots	\vdots
0,8	2,22	0,69
0,5	1,65	0,62
0,2	1,22	0,55
\vdots	\vdots	\vdots
-0,2	0,82	0,45
-0,5	0,61	0,38
-0,8	0,45	0,31
\vdots	\vdots	\vdots

Abbildung 3.5: Plot von $g(\hat{\pi}_k)$ gegen y bzw. $\hat{\pi}_k$ 

Logistische Regression ohne erklärende X-Variablen

Bevor irgendeine erklärende X-Variable aufgenommen wird, ist das Modell der logistischen Regression

$$g(\pi_k) = \eta_k = \log[\pi_k/(1 - \pi_k)] = \beta_0 \quad (3.49)$$

und die zu testenden Hypothesen

$$H_0 : \beta_0 = 0 \text{ und } H_1 : \beta_0 \neq 0.$$

Ausgehend von (3.21) und unter Beachtung von $x_{k0} = 1$, $n_k = 1$ für alle k und $K =$

3.3 Zur Interpretation der Ergebnisse des logistischen Modells

n (da wegen fehlender erklärender X-Variablen keine Gruppen gebildet werden können) sowie unter Berücksichtigung von (3.15) erhält man den ML-Schätzer für β_0 wie folgt:

$$\frac{\partial l}{\partial \beta_0} = \sum_{k=1}^n (y_k - \pi_k) = \sum_{k=1}^n \left\{ y_k - \left(\frac{\exp \beta_0}{1 + \exp \beta_0} \right) \right\} \stackrel{!}{=} 0. \quad (3.50)$$

Da $\exp(\beta_0)/[1 + \exp(\beta_0)]$ ein konstanter Term für alle k ist, folgt:

$$\begin{aligned} \sum_{k=1}^n y_k - \frac{n \cdot \exp b_0}{1 + \exp b_0} &= 0 \\ \sum_{k=1}^n y_k + \exp b_0 \cdot \sum_{k=1}^n y_k - n \cdot \exp b_0 &= 0 \\ \sum_{k=1}^n y_k - \exp b_0 (n - \sum_{k=1}^n y_k) &= 0 \\ \exp b_0 &= \frac{\sum_{k=1}^n y_k}{n - \sum_{k=1}^n y_k} \\ b_0 &= \log \left(\frac{\sum_{k=1}^n y_k}{n - \sum_{k=1}^n y_k} \right). \end{aligned} \quad (3.51)$$

Die geschätzte Wahrscheinlichkeit $\hat{\pi}_k$ ist somit konstant für alle n und entspricht der Stichprobenhäufigkeit für Y = 1:

$$\hat{\pi}_k = \frac{e^{b_0}}{1 + e^{b_0}} = \frac{\frac{\sum_{k=1}^n y_k}{n - \sum_{k=1}^n y_k}}{\frac{\sum_{k=1}^n y_k}{n - \sum_{k=1}^n y_k} + 1} = \frac{\sum_{k=1}^n y_k}{n} \quad (3.52)$$

Für den Wert der log-Likelihood-Funktion an der Stelle b_0 folgt gemäß (3.20):

$$\begin{aligned}
 l(b_0; \mathbf{y}) &= b_0 \sum_{k=1}^n y_k - n \cdot \log(1 + \exp b_0) \\
 &= b_0 \sum_{k=1}^n y_k - n \cdot \log \left(1 + \frac{\sum_{k=1}^n y_k}{n - \sum_{k=1}^n y_k} \right) \\
 &= b_0 \sum_{k=1}^n y_k - n \cdot \log \left(\frac{n}{n - \sum_{k=1}^n y_k} \right)
 \end{aligned} \tag{3.53}$$

Die Nullhypothese $H_0 : \beta_0 = 0$ ist wahr, wenn $\pi_k = 0,5$, denn es ist: $\log[0,5/0,5] = \log(1) = 0 = \beta_0$.

Dieser Test ist äquivalent zu einem Test auf gleiche Anteile von $Y = 1$ und $Y = 0$.

Bei Verwendung statistischer Software zur Schätzung eines logistischen Modells ist das Modell nur mit der Konstanten im allgemeinen das Ausgangsmodell, im weiteren symbolisiert mit M_0 .

Beispiel 3.2 (Fortsetzung 1):

In Fortsetzung des Beispiels 3.2 wird die Schätzung der Parameter unter Verwendung der Prozedur Regression Binary Logistic von SPSS unter Windows Release 10.0.7¹⁴ durchgeführt. Dafür ist die Auflistung aller Einzelfälle, d.h. die Daten der Tabelle 3.4, erforderlich.

Geprüft werden soll auf einem Signifikanzniveau von $\alpha = 0,05$.

Block 0: Beginning Block

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	-,282	,202	1,947	1	,163	,754

¹⁴Siehe SPSS for Windows, Advanced Statistics, SPSS Regression Models 9.0

Der geschätzte Parameter des Modells (3.49) kann der Tabelle “Variables in the Equation“ des Blocks 0 entnommen werden: $b_0 = -0,282$, so dass resultiert:

$$g(\hat{\pi}_k) = \hat{\eta}_k = -0,282.$$

Der Parameter b_0 in diesem Modell ist zum 5%-Niveau nicht signifikant, wie der Wald-Test anzeigt.

Für das Beispiel sind $n = 100$; $\sum y_k = 43$. Einsetzen in (3.51) und (3.52) ergibt:

$$b_0 = \log(43/57) = \log(0,7544) = -0,28185;$$

$$\hat{\pi} = 43/100 = 0,43.$$

Für den Wert der log-Likelihood-Funktion an der Stelle b_0 folgt gemäß (3.53):

$$\begin{aligned} l(\hat{\pi}; \mathbf{y}) &= -0,28185 \cdot 43 - 100 \cdot \log(100/57) \\ &= -12,11955 - 56,21189 = -68,3314. \end{aligned}$$

Dieser Wert der log-Likelihood-Funktion für das Ausgangsmodell wird für die Berechnung von ΔD für ein Modell mit J erklärenden X -Variablen verwendet.

Einfache logistische Regression mit einer dichotomen X -Variablen

Es wird hier der Fall einer einfachen logistischen Regression

$$g(\pi_k) = \eta_k = \beta_0 + \beta_1 x_k \quad (k = 1, \dots, K)$$

betrachtet, wobei die erklärende Variable X eine dichotome Variable ist, die die Werte Null und Eins annimmt.

Die Werte für π_k und $1 - \pi_k$ in Abhängigkeit von x_k gemäß (3.15) und (3.25) sind in der folgenden Tabelle zusammengestellt, wobei zur Vereinfachung auf den Index k verzichtet wird und $\pi(1)$ bzw. $\pi(0)$ die Wahrscheinlichkeit an der Stelle $x = 1$ bzw. $x = 0$ symbolisieren.

Tabelle 3.9: Wahrscheinlichkeiten des logistischen Modells mit dichotomem X

Response Variable	erklärende Variable X	
	x = 1	x = 0
y = 1	$\pi(1) = \frac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)}$	$\pi(0) = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)}$
y = 0	$1 - \pi(1) = \frac{1}{1 + \exp(\beta_0 + \beta_1)}$	$1 - \pi(0) = \frac{1}{1 + \exp(\beta_0)}$
Gesamt	1	1

Die odds gemäß (3.16) sind für $x = 1$ bzw. $x = 0$:

$$\pi(1)/[1 - \pi(1)] \text{ bzw. } \pi(0)/[1 - \pi(0)]$$

und die log odds (logits) gemäß (3.17):

$$g(\pi(1)) = \log\{\pi(1)/[1 - \pi(1)]\} \text{ bzw. } g(\pi(0)) = \log\{\pi(0)/[1 - \pi(0)]\}.$$

Nach Einsetzen der Formeln erhält man für die odds:

$$\begin{aligned} \frac{\pi(1)}{1 - \pi(1)} &= \exp(\beta_0 + \beta_1) \\ \frac{\pi(0)}{1 - \pi(0)} &= \exp(\beta_0) \end{aligned} \tag{3.54}$$

und für die link Funktion (log odds):

$$g(\pi(1)) = \beta_0 + \beta_1$$

$$g(\pi(0)) = \beta_0$$

Als odds ratio (es soll hier mit or symbolisiert werden) wird das Verhältnis der beiden odds bezeichnet:

$$or = \frac{\frac{\pi(1)}{1 - \pi(1)}}{\frac{\pi(0)}{1 - \pi(0)}} = \frac{\pi(1)[1 - \pi(0)]}{\pi(0)[1 - \pi(1)]}. \tag{3.55}$$

Nach Einsetzen von (3.54) erhält man:

$$or = \frac{\exp(\beta_0 + \beta_1)}{\exp(\beta_0)} = \exp(\beta_1). \tag{3.56}$$

3.3 Zur Interpretation der Ergebnisse des logistischen Modells

Dieses odds ratio ist ein Maß der Assoziation, das direkt aus dem logistischen Modell geschätzt werden kann. Es gibt an, wie viel wahrscheinlicher es ist, dass das Ereignis eintritt ($y = 1$) unter den Fällen mit $x = 1$ im Gegensatz zu den Fällen mit $x = 0$.

Daraus folgt als $\log(\text{or})$ bzw. Differenz in den \log odds

$$\log(\text{or}) = g(\pi(1)) - g(\pi(0)) = \beta_1. \quad (3.57)$$

Entsprechende Formeln erhält man mit den geschätzten Parameterwerten.

Beispiel 3.2 (Fortsetzung 2):

Die zu prüfende Hypothese lautet: Hat die Variable Alter einen signifikanten Einfluss auf die response Variable HKE? Dabei soll das Alter zunächst als dichotomisierte Variable verwendet werden, wobei die folgende Kodierung erfolgt: kleiner 55 Jahre = 0 und größer gleich 55 Jahre = 1.

Die Hypothesen lauten:

$$H_0 : \beta_1 = 0 \text{ und } H_1 : \beta_1 \neq 0$$

mit

$$g(\pi_k) = \eta_k = \beta_0 \quad (3.49)$$

als Modell M_0 unter der Nullhypothese und

$$g(\pi_k) = \eta_k = \beta_0 + \beta_1 x_k \quad (3.58)$$

als Modell M_1 unter der Alternativhypothese gemäß (3.18).

Mit der dichotomisierten Variablen Alter erhält man die folgende 2×2 -Kontingenztafel zur Variablen HKE.

Herzkranzgefäßerkrankung * Alter dichotom Crosstabulation

Count

		Alter dichotom		Total
		-54	55-	
Herzkranzge-	nein = 0	51	6	57
fäßerkrankung	ja = 1	22	21	43
Total		73	27	100

Als Schätzung ergibt sich:

Block 1: Method = Enter

Omnibus Test of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	18,704	1	,000
	Block	18,704	1	,000
	Model	18,704	1	,000

Model Summary

Step 1	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	117,959	,171	,229

Classification Table^a

Observed			Predicted		
			Herzkranzgefäßerkrankung		Percentage
			nein	ja	Correct
Step 1	Herzkranzge-	nein	51	6	89,5
	fäßerkrankung	ja	22	21	48,8
	Overall Percentage				72,0

^a. The cut value is 0.5

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a Alter dichotom	2,094	,529	15,690	1	,000	8,114
Constant	-,841	,255	10,865	1	,001	,431

^a. Variable(s) entered on step 1: Alter dichotom

In der Tabelle “Omnibus Tests of Model Coefficients“ des Blocks 1 steht unter Model Chi-square der Wert der Differenz der Deviancen des Modells M_0 und des Modells M_1 , d.h. ΔD gemäß (3.37): $\Delta D = 18,704$. Die Anzahl der Freiheitsgrade für ΔD gemäß (3.38) ist $f = 1$, was im Output unter df angegeben ist. Der Wert von ΔD wird mit dem Quantil der Chi-Quadrat-Verteilung für das vorgegebene Signifikanzni-

3.3 Zur Interpretation der Ergebnisse des logistischen Modells

veau $\alpha = 0,05$ verglichen. Aus der Tabelle der Chi-Quadrat-Verteilung findet man für $f = 1$ Freiheitsgrade $\chi_{0,05;1}^2 = 3,841$. H_0 wird somit auf dem 5%-Niveau abgelehnt, da $\Delta D = 18,704 > \chi_{0,05;1}^2 = 3,841$ ist.

Zu diesem Ergebnis gelangt man auch durch den Vergleich des vorgegebenen Signifikanzniveaus $\alpha = 0,05$ mit der Ausgabe Sig. in der Tabelle “Omnibus Tests of Model Coefficients”: Da $Sig. = ,000 < \alpha = 0,05$ ist, wird H_0 abgelehnt. Das dichotomisierte Alter hat einen signifikanten Einfluss auf das Auftreten von HKE.

Aus der Tabelle “Model Summary“ des Blocks 1 folgt für den maximalen Wert der log-Likelihood-Funktion des Modells M_1 :

$$l_1(\hat{\pi}; \mathbf{y}) = -58,9795.$$

Aus der Fortsetzung 1 des Beispiels 3.2 kann der maximale Wert der log-Likelihood-Funktion des Modells M_0 entnommen werden:

$$l(b_0; \mathbf{y}) = -68,3314.$$

Somit folgt:

$$\Delta D = 2[l_1(\hat{\pi}; \mathbf{y}) - l_0(\hat{\pi}; \mathbf{y})] = 2[(-58,9795) - (-68,3314)] = 18,704.$$

Das Nagelkerke $R_N^2 = 0,229$ gibt an, dass jedoch nur rund 23 % der Variation durch dieses logistische Modell erklärt werden.

Unter Verwendung des geschätzten Modells und einem cut value = 0,5 werden 89,5% der Fälle ohne Herzkrankgefäßerkrankung und 48,8% der Fälle mit Herzkrankgefäßerkrankung korrekt vorhergesagt, was 72% korrekte Vorhersagen insgesamt ergibt (siehe Classification Table).

Die geschätzten Parameter des Modells M_1 können der Tabelle “Variables in the Equation“ des Blocks 1 entnommen werden:

$$b_0 = -0,841 \text{ und } b_1 = 2,094.$$

Der Wald-Test zeigt für beide Parameter Signifikanz zum vorgegebenen 5%-Niveau an. Diese geschätzten Parameter-Werte in (3.58) eingesetzt, ergibt:

$$g(\hat{\pi}_k) = \hat{\eta}_k = -0,841 + 2,094x_k.$$

Für die beiden möglichen Werte von X resultiert als Wert der link Funktion:

$$g(\hat{\pi}(1)) = \log\{\pi(1)/[1 - \pi(1)]\} = -0,841 + 2,094 = 1,253$$

$$g(\hat{\pi}(0)) = \log\{\pi(0)/[1 - \pi(0)]\} = -0,841.$$

Daraus ergeben sich die geschätzten odds zu:

$$\hat{\pi}(1)/[1 - \hat{\pi}(1)] = \exp(1,253) = 3,5$$

$$\hat{\pi}(0)/[1 - \hat{\pi}(0)] = \exp(-0,841) = 0,4313$$

Bei den “älteren“ Personen (55- Jahre) sind in die Stichprobe die Chancen für das Auftreten von HKE 3,5 mal größer als für das Nichtauftreten von HKE.

Bei den “jüngeren“ Personen (-54 Jahre) betragen in der Stichprobe die Chancen für das Auftreten von HKE nicht einmal die Hälfte im Gegensatz zum Nichtauftreten von HKE.

Für das geschätzte log odds ratio (3.57) folgt:

$$\log(or) = g(\hat{\pi}(1)) - g(\hat{\pi}(0)) = b_1 = 2,094$$

und für das geschätzte odds ratio (3.56), was in der Tabelle unter Exp(B) zu finden ist:

$$or = \exp(2,094) = 8,114.$$

HKE ($Y = 1$) tritt in der Stichprobe rund achtmal häufiger bei Personen 55 Jahre und älter ($X = 1$) auf im Vergleich zu den “jüngeren“ Personen ($X = 0$).

Für die geschätzten bedingten Wahrscheinlichkeiten $\hat{\pi}_k$ des Eintretens von HKE nach (3.15) folgt:

$$\begin{aligned}\hat{\pi}_k &= \exp(b_0 + b_1 x_k) / [1 + \exp(b_0 + b_1 x_k)] \\ &= \exp(-0,841 + 2,094 x_k) / [1 + \exp(-0,841 + 2,094 x_k)],\end{aligned}$$

und zwar gegeben die älteren Personen ($X = 1$):

$$\begin{aligned}\hat{\pi}(1) &= \exp(b_0 + b_1) / [1 + \exp(b_0 + b_1)] \\ &= \exp(-0,841 + 2,094) / [1 + \exp(-0,841 + 2,094)] \\ &= 3,5/4,5 = 0,7778\end{aligned}$$

und gegeben die jüngeren Personen ($X = 0$):

$$\begin{aligned}\hat{\pi}(0) &= \exp(b_0)/[1 + \exp(b_0)] = \exp(-0,841)/[1 + \exp(-0,841)] \\ &= 0,4313/1,4313 = 0,3013.\end{aligned}$$

In diesem Fall einer einfachen logistischen Regression mit einer dichotomen X-Variablen kann das Ergebnis auch direkt aus der Kontingenztafel (Herzkranzgefäßerkrankung * Alter dichotom Crosstabulation) ermittelt werden.

Es sind:

$$\begin{aligned}\hat{\pi}(1)/[1 - \hat{\pi}(1)] &= 0,21/0,06 = 3,5 \\ \hat{\pi}(0)/[1 - \hat{\pi}(0)] &= 0,22/0,51 = 0,4314 \\ or &= (21/6)/(22/51) = 8,11 \quad \text{und} \\ \log(8,11) &= 2,094.\end{aligned}$$

Dies zeigt deutlich, dass die Möglichkeit der Präsentation der Daten in einer Kontingenztafel die Basis für die Interpretation des geschätzten Parameters b_1 der logistischen Regression als $\log(or)$ gibt.

Da das odds ratio or bei kleinen Stichprobenumfängen eine schiefe Verteilung aufweist, werden Konfidenzintervalle für or in der Weise berechnet, indem zuerst ein Konfidenzintervall für den Koeffizienten b_1 gebildet wird und dann das Ergebnis exponentiert wird:

$$\exp[b_1 \pm z_{1-\alpha/2}s(b_1)].$$

Einfache logistische Regression mit einer mehrkategorialen X-Variablen

Nun sei angenommen, dass die erklärende Variable X eine mehrkategoriale (nominal oder ordinal skalierte) Variable mit h Ausprägungen ist (z.B. Bundesland, Wirtschaftszweig, Rasse, Beruf).

Eine solche Variable ist nicht unmittelbar in der logistischen Regression verwendbar. Die

3 Modellierung binärer Daten

Werte dieser Variablen müssen neu kodiert werden, indem eine Reihe von neuen Variablen (als Kontrast-Variablen bezeichnet) geschaffen wird. Die Anzahl der erforderlichen Kontrast-Variablen entspricht der Anzahl der verschiedenen Ausprägungen der mehrkategorialen Variablen minus Eins: $h - 1$, da eine Kategorie als Bezugskategorie dient¹⁵. Es gibt verschiedene Möglichkeiten der Erstellung der Kontrast-Variablen, von denen zwei im nachfolgenden Beispiel demonstriert werden sollen.

Das Modell der logistischen Regression mit einer mehrkategorialen X-Variablen lautet:

$$g(\pi_k) = \eta_k = \beta_0 + \beta_1 X_{k1} + \beta_2 X_{k2} + \dots + \beta_{h-1} X_{k;h-1} \quad (k = 1, \dots, K) \quad (3.59)$$

Beispiel 3.2 (Fortsetzung 3):

Das Alter wird wie folgt kodiert:

bis 40 Jahre = 1

41 - 54 Jahre = 2

55 Jahre und älter = 3,

wodurch eine mehrkategoriale Variable mit 3 möglichen Ausprägungen entsteht. Die 1. Alterklasse wird als Bezugskategorie gewählt.

Die Hypothesen lauten:

H_0 : Die mehrkategoriale Variable Alter hat keinen Einfluss auf die Wahrscheinlichkeit des Auftretens von HKE, d.h., die Parameter bei allen drei Alterskategorien sind gleich Null.

H_1 : Die mehrkategoriale Variable Alter beeinflusst die Wahrscheinlichkeit des Auftretens von HKE, d.h., mindestens ein Parameter bei den drei Alterskategorien ist verschieden von Null.

Das impliziert die folgenden Modelle:

$$g(\pi_k) = \eta_k = \beta_0 \quad (3.49)$$

¹⁵Unter SPSS kann über die entsprechenden Dialogfelder als Bezugskategorie die letzte oder die erste gewählt werden. Über die Syntax kann aber auch jede andere Kategorie als Bezugskategorie ausgewählt werden.

3.3 Zur Interpretation der Ergebnisse des logistischen Modells

als Modell M_0 unter der Nullhypothese und

$$g(\pi_k) = \eta_k = \beta_0 + \beta_1 x_{k1} + \beta_2 x_{k2} \quad (3.60)$$

als Modell M_1 unter der Alternativhypothese. Zu beachten ist, dass die Kontrat-Variable X_1 mit dem zugehörigen Parameter β_1 für die 2. Alterskategorie und die Kontrast-Variable X_2 mit dem zugehörigen Parameter β_2 für die 3. Alterskategorie steht (da die 1. Alterskategorie die Bezugskategorie ist). Die inhaltliche Bedeutung der Parameter hängt von der Art der Erstellung der Kontrast-Variablen ab.

Indikator-Kodierung zur Erstellung der Kontrast-Variablen:

Die Kontrast-Variablen zeigen das Vorhandensein bzw. Nicht-Vorhandensein der Kategorie an. Dieser Kontrast entspricht der herkömmlichen Kodierung in Scheinvariablen (0;1-Variablen). Die Werte für die Bezugskategorie werden in allen h-1 Kontrast-Variablen gleich Null gesetzt. Jede andere Kategorie erhält in einer Kontrast-Variablen eine Eins und in allen anderen Kontrast-Variablen eine Null. Dies ist die am häufigsten verwendete Art der Kodierung, da sie eine leichte Interpretation erlaubt.

Für das Beispiel ergibt die Indikator-Kodierung mit der ersten Kategorie als Bezugskategorie die folgende Design-Matrix:

Categorical Variables Codings

		Frequency	Parameter Coding	
			(1)	(2)
Alter mehrkategorial	-40	39	,000	,000
	41-54	34	1,000	,000
	55-	27	,000	1,000

Für die Schätzung erhält man:

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Step 1	24,689	2	,000
Block	24,689	2	,000
Model	24,689	2	,000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	111,974	,219	,294

Classification Table^a

Observed			Predicted		
			Herzkranzgefäßerkrankung		Percentage
			nein	ja	Correct
Step 1	Herzkranzge-	nein	51	6	89,5
	fäßerkrankung	ja	22	21	48,8
	Overall Percentage				72,0

a. The cut value is 0.5

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a A_K			19,799	2	,000	
A_K(1)	1,283	,542	5,614	1	,018	3,609
A_K(2)	2,773	,623	19,792	1	,000	15,999
Constant	-1,520	,417	13,266	1	,000	,219

a. Variables(s) entered on step 1: A_K (Alter mehrkategorial).

In der Tabelle “Omnibus Tests of Model Coefficients“ des Blocks 1 steht unter Model Chi-square der Wert der Differenz der Deviancen des Modells M_0 und des Modells M_1 , d.h. ΔD gemäß (3.37): $\Delta D = 24,689$. Die Anzahl der Freiheitsgrade für ΔD gemäß (3.38) ist $f = 2$, da gegenüber dem Modell M_0 zwei weitere Parameter zu schätzen sind, was im Output unter df angegeben ist. Der Wert von ΔD wird mit dem Quantil der Chi-Quadrat-Verteilung für das vorgegebene Signifikanzniveau $\alpha = 0,05$ verglichen. Aus der Tabelle der Chi-Quadrat-Verteilung findet man für $f = 2$ Freiheitsgrade $\chi_{0,05;2}^2 = 5,99$. H_0 wird somit auf dem 5%-Niveau abgelehnt, da $\Delta D = 24,689 > \chi_{0,05;2}^2 = 5,99$ ist.

Zu diesem Ergebnis gelangt man auch durch den Vergleich des vorgegebenen Signifikanzniveaus $\alpha = 0,05$ mit der Ausgabe Sig. in der Tabelle “Omnibus Tests of Model Coefficients“: Da $Sig. = ,000 < \alpha = 0,05$ ist, wird H_0 abgelehnt. Das mehrkategoriale Alter hat einen signifikanten Einfluss auf das Auftreten von HKE.

Aus der Tabelle “Model Summary“ des Blocks 1 folgt für den maximalen Wert der log-Likelihood-Funktion des Modells M_1 :

$$l_1(\hat{\pi}; \mathbf{y}) = -55,987.$$

Der maximale Wert der log-Likelihood-Funktion des Modells M_0 betrug (siehe Fortsetzung 1 des Beispiels 3.2):

$$l(b_0; \mathbf{y}) = -68,3314.$$

Somit folgt:

$$\Delta D = 2[l_1(\hat{\pi}; \mathbf{y}) - l_0(\hat{\pi}; \mathbf{y})] = 2[(-55,987) - (68,3314)] = 24,689.$$

Das Nagelkerke $R_N^2 = 0,294$ gibt an, dass jedoch nur rund 30% der Variation durch dieses logistische Modell erklärt werden.

Unter Verwendung dieses geschätzten Modells und einem cut value = 0,5 werden 89,5% der Fälle ohne Herzkrankgefäßerkrankung und 48,8% der Fälle mit Herzkrankgefäßerkrankung korrekt vorhergesagt, was 72% korrekte Vorhersagen insgesamt ergibt (siehe Classification Table).

Auch im Fall einer einfachen logistischen Regression mit einer mehrkategorialen X-Variablen und Verwendung der Indikator-Kodierung kann das Ergebnis direkt aus der Kontingenztabelle ermittelt werden. Die Kontingenztabelle dieser mehrkategorialen Variablen Alter zur Variablen HKE ist wie folgt:

Herzkrankgefäßerkrankung * Alter mehrkategorial Crosstabulation

Count

		Alter mehrkategorial			Total
		-40 (1)	41-54 (2)	55- (3)	
Herzkrankgefäß	nein = 0	32	19	6	57
erkrankung	ja = 1	7	15	21	43
Total		39	34	27	100

Wählt man die erste Kategorie (bis 40 Jahre) als Bezugskategorie, so ergeben sich aus dieser Kontingenztabelle folgende odds (Chancen des Auftretens von HKE bei gegebener Kategorie des Alters):

$$\hat{\pi}(1)/[1 - \hat{\pi}(1)] = 7/32 = 0,2188$$

$$\hat{\pi}(2)/[1 - \hat{\pi}(2)] = 15/19 = 0,7895$$

$$\hat{\pi}(3)/[1 - \hat{\pi}(3)] = 21/6 = 3,5$$

Bei den "älteren" Personen (55- Jahre) sind in der Stichprobe die Chancen für das Auftreten von HKE 3,5 mal größer als für das Nichtauftreten von HKE.

Bei den "mittleren" Personen (41-54 Jahre) betragen in der Stichprobe die Chancen für

das Auftreten von HKE etwa drei Viertel im Gegensatz zum Nichtauftreten von HKE. Bei den “jüngeren“ Personen (-40 Jahre) betragen in der Stichprobe die Chancen für das Auftreten von HKE etwa ein Viertel im Gegensatz zum Nichtauftreten von HKE.

Für die odds ratio und $\log(\text{or})$ folgt:

$$\begin{aligned} \text{Kategorie 1 (zu sich selbst): } & \text{or}_1 = 1 & \log(1) = 0 \\ \text{Kategorie 2 zur Kategorie 1: } & \text{or}_2 = (15/19)/(7/32) = 3,61 & \log(3,61) = 1,2834 \\ \text{Kategorie 3 zur Kategorie 1: } & \text{or}_3 = (21/6)/(7/32) = 16 & \log(16) = 2,7725 \end{aligned}$$

In der Stichprobe tritt HKE rund 3,6 mal häufiger bei Personen mit einem Alter von 41-54 Jahre ($X = 2$) als unter den Personen mit einem Alter bis 40 Jahre ($X = 0$, Bezugskategorie) auf.

In der Stichprobe tritt HKE rund 16 mal häufiger bei Personen mit einem Alter von 55 und mehr Jahren ($X = 3$) als unter Personen mit einem Alter bis 40 Jahre ($X = 0$, Bezugskategorie) auf.

Unter Beachtung der Ergebnisse aus der Kontingenztafel wird nun deutlich, dass die Interpretation der Parameter der 2. Alterskategorie = $A_K(1)$ und der 3. Alterskategorie = $A_K(2)$ immer im Vergleich zur Bezugskategorie erfolgen muss:

$$\begin{aligned} \log[\text{or}(41-54; -40)] &= b_1 = 1,283 & \text{or}(41-54; -40) &= \exp(1,2834) = 3,609 \\ \log[\text{or}(55-; -40)] &= b_2 = 2,773 & \text{or}(55-; -40) &= \exp(2,7725) = 15,999 \end{aligned}$$

Dies lässt sich unter Verwendung der angegebenen Design-Matrix wie folgt nachweisen, z.B. für

$$\begin{aligned} \log[\text{or}(41-54; -40)] &= g(\hat{\pi}_{41-54}) - g(\hat{\pi}_{-40}) \\ &= [b_0 + b_1(X_1 = 1) + b_2(X_2 = 0)] - [b_0 + b_1(X_1 = 0) + b_2(X_2 = 0)] = b_1. \end{aligned}$$

Kodierung Abweichung (deviation) zur Erstellung der Kontrast-Variablen:

Der Wert der Bezugskategorie wird für alle Kontrast-Variablen -1 gesetzt. Jede andere Kategorie erhält in einer Kontrast-Variablen eine Eins und in allen anderen Kontrast-Variablen eine Null.

Für das Beispiel ergibt die Abweichungskodierung mit der ersten Kategorie als Bezugskategorie die folgende Design-Matrix:

3.3 Zur Interpretation der Ergebnisse des logistischen Modells

Categorical Variables Codings

	Frequency	Parameter coding	
		(1)	(2)
Alter mehrkategorial	-40	39	-1,000
	41-54	34	1,000
	55-	27	,000

Für die Schätzung erhält man:

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Step 1	24,689	2	,000
Block	24,689	2	,000
Model	24,689	2	,000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	111,974	,219	,294

Classification Table^a

Observed			Predicted		
			Herzkranzgefäßerkrankung		Percentage
			nein	ja	Correct
Step 1	Herzkranzge-	nein	51	6	89,5
	fäßerkrankung	ja	22	21	48,8
Overall Percentage					72,0

a. The cut value is 0.5

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a						
A_K			19,799	2	,000	
A_K(1)	-,069	,310	,049	1	,825	,934
A_K(2)	1,421	,358	15,785	1	,000	4,139
Constant	-,168	,238	,499	1	,480	,846

a. Variables(s) entered on step 1: A_K (Alter mehrkategorial).

Wie leicht ersichtlich, ändern sich die Gütekriterien für die logistische Regression insgesamt durch die Veränderung der Erstellung der Kontrast-Variablen nicht, jedoch die geschätzten Parameter und möglicherweise auch deren Signifikanz.

Durch Auswahl der Abweichungskodierung wird der Effekt einer jeden Kategorie zum **mittleren Effekt** aller Kategorien verglichen, d.h., die geschätzten Parameter geben

3 Modellierung binärer Daten

die Differenz zum mittleren Effekt aller Kategorien an. Der geschätzte Parameter für die Bezugskategorie wird nicht ausgegeben; er ergibt sich als die negative Summe der Parameterwerte aller anderen Kategorien.

Um dies zu zeigen, werden aufgrund der obigen Design-Matrix Categorical Variables Codings die logits für die Alterskategorien und anschließend daraus das arithmetische Mittel berechnet:

$$\begin{aligned}
 g(\hat{\pi}(1)) &= b_0 + b_1(-1) + b_2(-1) = b_0 - b_1 - b_2 \\
 g(\hat{\pi}(2)) &= b_0 + b_1(1) + b_2(0) = b_0 + b_1 \\
 g(\hat{\pi}(3)) &= b_0 + b_1(0) + b_2(1) = b_0 + b_2 \\
 \bar{g}(\hat{\pi}) &= \frac{g(\hat{\pi}(1)) + g(\hat{\pi}(2)) + g(\hat{\pi}(3))}{3} \\
 &= \frac{(b_0 - b_1 - b_2) + (b_0 + b_1) + (b_0 + b_2)}{3} = b_0
 \end{aligned} \tag{3.61}$$

Als Abweichungen erhält man:

$$\begin{aligned}
 g(\hat{\pi}(1)) - \bar{g}(\hat{\pi}) &= b_0 - b_1 - b_2 - b_0 = -b_1 - b_2 \\
 g(\hat{\pi}(2)) - \bar{g}(\hat{\pi}) &= b_0 + b_1 - b_0 = b_1 \\
 g(\hat{\pi}(3)) - \bar{g}(\hat{\pi}) &= b_0 + b_2 - b_0 = b_2
 \end{aligned}$$

Für das Beispiel folgt unter Verwendung der obigen Kontingenztafel:

$$\begin{aligned}
 g(\hat{\pi}(1)) &= \log[(7/39)/(32/39)] = \log(7/32) = -1,51983 \\
 g(\hat{\pi}(2)) &= \log(15/19) = -0,23639 \\
 g(\hat{\pi}(3)) &= \log(21/6) = 1,25276 \\
 \bar{g}(\hat{\pi}) &= [\log(7/32) + \log(15/19) + \log(21/6)]/3 = -0,50345/3 = -0,16782.
 \end{aligned}$$

Als Abweichungen erhält man:

$$\begin{aligned}
 g(\hat{\pi}(1)) - \bar{g}(\hat{\pi}) &= -1,51983 - (-0,16782) = -1,352 \\
 &= -(b_1 + b_2) = -[(-0,0686) + 1,4205] = -1,352 \\
 g(\hat{\pi}(2)) - \bar{g}(\hat{\pi}) &= -0,23639 - (-0,16782) = -0,069 = b_1 \\
 g(\hat{\pi}(3)) - \bar{g}(\hat{\pi}) &= 1,25276 - (-0,16782) = 1,421 = b_2.
 \end{aligned}$$

Um die odds ratios (angegeben unter Exp(B) im obigen Output) zu verstehen, müssen die errechneten Werte exponentiert werden. Für $\bar{g}(\hat{\pi})$ ergibt sich:

$$\begin{aligned}\exp[\bar{g}(\hat{\pi})] &= \exp\left\{\frac{g(\hat{\pi}(1)) + g(\hat{\pi}(2)) + g(\hat{\pi}(3))}{3}\right\} \\ &= \sqrt[3]{g(\hat{\pi}(1)) \cdot g(\hat{\pi}(2)) \cdot g(\hat{\pi}(3))} \\ &= \sqrt[3]{\frac{7}{32} \cdot \frac{15}{19} \cdot \frac{21}{6}} = 0,84551\end{aligned}\tag{3.62}$$

d.h., es resultiert ein geometrisches Mittel.

Das odds ratio ist somit das Verhältnis der odds einer Kategorie zum geometrischen Mittel aller Kategorien, was die Interpretation erschwert, da es nicht das Verhältnis der odds zweier verschiedener Kategorien ist,

für die 1. Alterskategorie:

$$\exp(-1,352) = \exp[g(\hat{\pi}(1)) - \bar{g}(\hat{\pi})] = \frac{\exp[g(\hat{\pi}(1))]}{\exp[\bar{g}(\hat{\pi})]} = \frac{\frac{7}{32}}{\sqrt[3]{\frac{7}{32} \cdot \frac{15}{19} \cdot \frac{21}{6}}} = 0,259$$

für die 2. Alterskategorie:

$$\exp(-0,069) = \exp[g(\hat{\pi}(2)) - \bar{g}(\hat{\pi})] = \frac{\exp[g(\hat{\pi}(2))]}{\exp[\bar{g}(\hat{\pi})]} = \frac{\frac{15}{19}}{\sqrt[3]{\frac{7}{32} \cdot \frac{15}{19} \cdot \frac{21}{6}}} = 0,934$$

für die 3. Alterskategorie:

$$\exp(1,421) = \exp[g(\hat{\pi}(3)) - \bar{g}(\hat{\pi})] = \frac{\exp[g(\hat{\pi}(3))]}{\exp[\bar{g}(\hat{\pi})]} = \frac{\frac{21}{6}}{\sqrt[3]{\frac{7}{32} \cdot \frac{15}{19} \cdot \frac{21}{6}}} = 4,139$$

Der Inhalt von log(or) zweier verschiedener Kategorien ist komplizierter als bei der Indikator-Kodierung, z.B.

$$\begin{aligned}\log[or(41-54; -40)] &= g(\hat{\pi}_{41-54}) - g(\hat{\pi}_{-40}) = g(\hat{\pi}(2)) - g(\hat{\pi}(1)) \\ &= -0,23639 - (-1,51983) = 1,283 \\ &= [b_0 + b_1(X_1 = 1) + b_2(X_2 = 0)] - [b_0 + b_1(X_1 = -1) + b_2(X_2 = -1)] \\ &= 2b_1 + b_2 = 2 \cdot (-0,0686) + 1,4205 = 1,283.\end{aligned}$$

Dieses Ergebnis ist identisch mit dem $\log(or_2)$, berechnet aus der Kontingenztafel, was zu erwarten war.

Die Verwendung dieser Kodierung hängt somit davon ab, ob es eine sachgerechte Interpretation für das mittlere odds gibt.

Weitere Arten der Erstellung von Kontrast-Variablen sind möglich, auf die hier aber nicht eingegangen werden soll.

Einfache logistische Regression mit einer metrischen X-Variablen

In der link Funktion

$$g(\pi_k) = \eta_k = \beta_0 + \beta_1 x_k \quad (k = 1, \dots, K)$$

ist X nunmehr eine metrische Variable. Die Interpretation von β_1 besagt, dass sich die logits um β_1 verändern, wenn sich die Variable X um eine Einheit erhöht. Somit ist

$$\log[or(x+1; x)] = g[\pi_k(x+1)] - g[\pi_k(x)] = \beta_1$$

für jeden Wert von X. Für das odds ratio ergibt sich

$$or(x+1; x) = \exp(\beta_1).$$

Dieser Maßstab der Veränderung in X von einer Einheit ist oftmals nicht sinnvoll, da er zu klein oder zu groß ist, so dass eine Veränderung von a Einheiten angebracht erscheint. Es ist dann:

$$\begin{aligned} \log[or(x+a; x)] &= g[\pi_k(x+a)] - g[\pi_k(x)] = [\beta_0 + \beta_1(x+a)] - [\beta_0 + \beta_1 x] = a\beta_1 \\ or(x+a; x) &= \exp(a\beta_1). \end{aligned}$$

Beispiel 3.2 (Fortsetzung 4):

Die erklärende Variable X (Alter) geht jetzt als metrische Variable in das Modell ein. Die zu prüfende Hypothese lautet: Hat die (metrische) Variable Alter einen signifikanten Einfluss auf die reponse Variable Y = HKE?

Die statistische Formulierung dieser Hypothese ergibt:

3.3 Zur Interpretation der Ergebnisse des logistischen Modells

$H_0 : \beta_1 = 0$ und $h_1 : \beta_1 \neq 0$.

Gepriüft werden soll auf einem Signifikanzniveau von $\alpha = 0,05$.

Das Modell M_0 unter der Nullhypothese lautet:

$$g(\pi_k) = \eta_k = \beta_0 \quad (3.49)$$

Für das Modell M_1 unter der Alternativhypothese resultiert gemäß (3.18):

$$g(\pi_k) = \eta_k = \beta_0 + \beta_1 x_k \quad (k = 1, \dots, 100)$$

Die Schätzung der Parameter wird unter Verwendung von SPSS durchgeführt, wofür die Auflistung aller Einzelfälle, d.h. die Daten der Tabelle 3.4, erforderlich sind.

Für die Schätzung erhält man:

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Step 1	29,310	1	,000
Block	29,310	1	,000
Model	29,310	1	,000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	107,353	,254	,341

Classification Table^a

Observed			Predicted		
			Herzkranzgefäßerkrankung		Percentage
			nein	ja	Correct
Step 1	Herzkranzge-	nein	45	12	78,9
	fäßerkrankung	ja	14	29	67,4
Overall Percentage					74,0

a. The cut value is 0.5

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a AL	0,111	0,02	21,254	1	,000	1,117
Constant	-5,309	1,134	21,935	1	,000	0,005

a. Variables(s) entered on step 1: AL

In der Tabelle "Omnibus Tests of Model Coefficients" des Blocks 1 steht unter Model Chi-square der Wert der Differenz der Deviancen des Modells M_0 und des Modells

M_1 , d.h. ΔD gemäß (3.37): $\Delta D = 29, 31$. Die Anzahl der Freiheitsgrade für ΔD gemäß (3.38) ist $f = 1$, was im Output unter df angegeben ist. Der Wert von ΔD wird mit dem Quantil der Chi-Quadrat-Verteilung zum vorgegebenen Signifikanzniveau $\alpha = 0,05$ verglichen. Aus der Tabelle der Chi-Quadrat-Verteilung findet man für $f = 1$ Freiheitsgrade $\chi_{0,05;1}^2 = 3,841$. H_0 wird somit auf dem vorgegebenen Signifikanzniveau von 5% abgelehnt, da $\Delta D = 29, 31 > \chi_{0,05;1}^2 = 3,841$ ist.

Zu diesem Ergebnis gelangt man auch durch den Vergleich des vorgegebenen Signifikanzniveaus $\alpha = 0,05$ mit der Ausgabe $Sig.$ in der Tabelle “Omnibus Tests of Model Coefficients”: Da $Sig. = ,000 < \alpha = 0,05$ ist, wird H_0 abgelehnt. Das Alter hat einen signifikanten Einfluss auf das Auftreten von HKE.

Aus der Tabelle “Model Summary“ des Blocks 1 folgt für den maximalen Wert der log-Likelihood-Funktion des Modells M_1 : $l_1(\hat{\pi}; \mathbf{y}) = -53,6765$.

Aus Fortsetzung 1 des Beispiels 3.2 kann der maximale Wert der log-Likelihood-Funktion des Modells M_0 entnommen werden:

$$l(b_0; \mathbf{y}) = -68,3314.$$

Somit folgt:

$$\Delta D = 2[l_1(\hat{\pi}; \mathbf{y}) - l_0(\hat{\pi}; \mathbf{y})] = 2[(-53,6765) - (-68,3314)] = 29,31.$$

Das Nagelkerke $R_N^2 = 0,341$ gibt an, dass jedoch nur rund 34% der Variation durch dieses logistische Modell erklärt werden. Dies ist deutlich mehr als bei Verwendung der dichotomisierten bzw. mehrkategorialen Altersvariablen und ist auf die volle Ausnutzung der Informationen über das Alter in der Stichprobe zurückzuführen.

Unter Verwendung des geschätzten Modells und einem cut value = 0,5 werden 78,9% der Fälle ohne Herzkrankgefäßerkrankung und 67,4% der Fälle mit Herzkrankgefäßerkrankung korrekt vorhergesagt, was 74% korrekte Vorhersagen insgesamt ergibt (siehe Classification Table). Vor allem für den Fall von Herzkrankgefäßerkrankungen hat sich der Prozentsatz korrekt vorhergesagter Fälle erheblich erhöht.

Die geschätzten Parameter des Modells M_1 können der Tabelle “Variables in the Equation“ des Blocks 1 entnommen werden:

$$b_0 = -5,309 \text{ und } b_1 = 0,111.$$

3.3 Zur Interpretation der Ergebnisse des logistischen Modells

Der Wald-Test zeigt für beide Parameter Signifikanz zum vorgegebenen 5%-Niveau an. Diese geschätzten Parameterwerte in die link Funktion eingesetzt, ergibt:

$$g(\hat{\pi}_k) = \hat{\eta}_k = -5,309 + 0,111x_k, \quad k = 1, \dots, 100.$$

Für die geschätzten Wahrscheinlichkeiten $\hat{\pi}_k$ nach (3.15) folgt:

$$\hat{\pi}_k = \exp(-5,309 + 0,111x_k) / [1 + \exp(-5,309 + 0,111x_k)]$$

Setzt man für jede erfasste Person $x_k = \text{Alter}_k$ ein, erhält man die geschätzte Wahrscheinlichkeit $\hat{\pi}_k$ für das Auftreten von HKE.

Tabelle 3.10: Geschätzte Wahrscheinlichkeiten $\hat{\pi}_k$ des logistischen Modells mit metrischer Variable Alter

Alter	$\hat{\pi}_k$	Alter	$\hat{\pi}_k$	Alter	$\hat{\pi}_k$	Alter	$\hat{\pi}_k$
20	,043479	35	,193533	46	,448414	57	,733617
23	,059621	36	,211436	47	,475979	58	,754725
24	,066153	37	,230521	48	,503690	59	,774674
25	,073344	38	,250781	49	,531379	60	,793445
26	,081248	39	,272192	50	,558877	61	,811033
28	,099422	40	,294712	51	,586017	62	,827449
29	,109804	41	,318280	52	,612645	63	,842716
30	,121125	42	,342817	53	,638617	64	,856866
32	,146793	43	,368224	54	,663803	65	,869939
33	,161237	44	,394384	55	,688091	69	,912465
34	,176807	45	,421163	56	,711387		

Multiple logistische Regression

Wenn das Modell mehrere erklärende X-Variablen enthält, gleichgültig ob dichotome, mehrkategoriale, metrische oder ein Mix dieser Variablen, dann beinhalten die geschätzten Parameter b_j die (statistisch) bereinigten log(or) und entsprechend $\exp(b_j)$ die bereinigten odd ratios. “Bereinigt“ bedeutet dabei, dass nur bei der Variablen X_j eine Veränderung um eine Einheit auftritt und alle anderen Variablen konstant gehalten werden. Mit anderen Worten: Die bereinigten odds ratios erhält man durch den Vergleich

von Fällen, die sich nur in der Variablen X_j unterscheiden, aber gleiche Werte bei allen anderen Variablen aufweisen. Diese Interpretation ist jedoch nur dann sinnvoll, wenn die vorausgesetzte Linearität der Variablen in $g(\pi_k)$ und ein konstanter Anstieg zutreffen.

Beispiel 3.3¹⁶:

Zur Erkennung des Harnblasenkarzinoms wird die T-Zelltypisierung und der LAI-Test eingesetzt. Es soll die Erkennung der Erkrankung von der Anwendung dieser Behandlungsmethoden untersucht werden. Dazu wurde eine Gruppe bereits an Harnblasenkarzinom Erkrankter und eine Gruppe Gesunder untersucht. Die Variable Y (als Ergebnis der Diagnose) nimmt die Werte 0 für gesund und 1 für krank an. Die erklärende Variable X_1 = "tzell" beinhaltet die Zelltypisierung und ist eine metrisch skalierte (intervallskalierte) Variable. Die erklärende Variable X_2 = "lai" enthält die Ergebnisse des LAI-Tests, die nur die Realisationen positiv (1) und negativ (0) annehmen kann und somit eine nominalskalierte dichotome Variable ist. X_0 ist die Variable für die Konstante, die den Wert Eins für alle Fälle annimmt. Der Gesamtstichprobenumfang ist $n = 45$.

Betrachtet man nur X_1 = "tzell" als erklärende Variable, ergeben sich für $y = 0$ bzw. $y = 1$ bzw. für alle Y-Werte die nachstehenden Statistiken:

T-Zelltypisierung

Diagnose	Mean	n	Std. Deviation	Range	Minimum	Maximum
gesund	71,267	21	4,836	17,4	61,1	78,5
krank	63,854	24	5,613	25,0	48,5	73,5
Total	67,313	45	6,410	30,0	48,5	78,5

Anhand dieser Statistiken ist ersichtlich, dass im Mittel die Werte der T-Zelltypisierung für die Gesunden einen höheren Wert bei kleinerer Streuung ergeben. Daraus leitet sich der Gedanke ab, die Wahrscheinlichkeit für das Vorliegen eines Harnblasenkarzinoms in Abhängigkeit von tzell zu schätzen.

Wenn als erklärende Variable nur X_2 = "lai" verwendet wird, so ist die zugehörige Kontingenztafel wie folgt:

¹⁶Dieses Beispiel wurde Bühl, A., Zöfel, P. (1994) S.264 ff. entnommen; die Daten sind in der Datei hkarz.sav der dort beigefügten Diskette enthalten; die Kodierung wurde jedoch verändert.

LAI-Test * Diagnose Crosstabulation

Count

		Diagnose		Total
		gesund (0)	krank (1)	
Lai-Test	negativ = 0	16	3	19
	positiv = 1	5	21	26
Total		21	24	45

Auch hier zeigt sich ein Zusammenhang zwischen beiden Variablen.

Es wird somit die Wahrscheinlichkeit für das Vorliegen eines Harnblasenkarzinoms in Abhängigkeit von tzell und lai geschätzt. Geprüft werden soll auf einem Signifikanzniveau von $\alpha = 0,05$.

Block 0: Beginning Block**Variables in the equation**

	B	S.E.	Wald	df	Sig.	Exp(b)
Step 0 Constant	,134	,299	,200	1	,655	1,143

Variables not in the Equation

	Score	df	Sig.
Step 0 Variables TZELL	15,317	1	,000
LAI	18,624	1	,000
Overall Statistics	21,591	2	,000

Block 1: Method = Enter**Omnibus Tests of Model Coefficients**

	Chi-square	df	Sig.
Step 1 Step	25,668	2	,000
Block	25,668	2	,000
Model	25,668	2	,000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	36,515	,435	,580

Classification Table^a

Observed	Predicted		
	Harnblasenkarzinom		Percentage Correct
	nein	ja	
Step 1 Diagnose gesund	16	5	76,2
krank	4	20	83,3
Overall Percentage			74,0

^a. The cut value is 0.5

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a TZELL	-,201	,094	4,574	1	,032	,818
LAI	2,205	,877	6,324	1	,012	9,074
Constant	12,440	6,527	3,633	1	,057	252728,3

^a. Variables(s) entered on step 1: TZELL, LAI.

In der Tabelle “Omnibus Tests of Model Coefficients“ des Blocks 1 steht unter Model Chi-square der Wert der Differenz der Deviancen des Modells M_0 und des Modells M_1 , d.h. ΔD gemäß (3.37): $\Delta D = 25,668$. Die Anzahl der Freiheitsgrade für ΔD gemäß (3.38) ist $f = 2$, da gegenüber dem Modell M_0 (das nur die Konstante beinhaltet) zwei weitere Parameter zu schätzen sind. Der Wert von ΔD wird mit dem Quantil der Chi-Quadrat-Verteilung für das vorgegebene Signifikanzniveau $\alpha = 0,05$ verglichen. Aus der Tabelle der Chi-Quadrat-Verteilung findet man für $f = 2$ Freiheitsgrade $\chi_{0,05;2}^2 = 5,99$. H_0 wird somit auf dem 5%-Niveau abgelehnt, da $\Delta D = 25,668 > \chi_{0,05;2}^2 = 5,99$ ist. Zu diesem Ergebnis gelangt man auch durch den Vergleich des vorgegebenen Signifikanzniveaus $\alpha = 0,05$ mit der Ausgabe Sig. in der Tabelle “Omnibus Tests of Model Coefficients“: Da $Sig. = ,000 < \alpha = 0,05$ ist, wird H_0 abgelehnt. Die beiden Variablen tzell und lai haben zusammen einen signifikanten Einfluss auf die Response Variable Y. Durch die Aufnahme von $X_1 = \text{“tzell“}$ und $X_2 = \text{“lai“}$ verringert sich -2 Log Likelihood um Model Chi-square = 25,668 auf -2 Log Likelihood = 36,515 für das Modell unter H_1 (siehe Tabelle “Model Summary“ des Blocks 1). Daraus folgt für das Modell unter H_0 : -2 Log Likelihood = 62,183.

Der maximale Wert der log-Likelihood-Funktion des Modells M_1 ist:

$$l_1(\hat{\pi}; \mathbf{y}) = -18,2575,$$

der maximale Wert der log-Likelihood-Funktion des Modells M_0 :

$$l(b_0; \mathbf{y}) = -31,0915.$$

Somit folgt:

$$\Delta D = 2[l_1(\hat{\pi}; \mathbf{y}) - l_0(\hat{\pi}; \mathbf{y})] = 2[(-18,2575) - (-31,0915)] = 25,668.$$

Das Nagelkerke $R_N^2 = 0,58$ gibt an, dass 58% der Variation durch dieses logistische Modell erklärt werden.

Unter Verwendung dieses geschätzten Modells und einem cut value = 0,5 werden 76,2% der gesunden Fälle und 83,3% der kranken Fälle korrekt vorhergesagt, was 80% korrekte Vorhersagen insgesamt ergibt (siehe Classification Table).

Die link Funktion für die k-te Gruppe gemäß (3.18) nach Einsetzen der geschätzten Parameter lautet:

$$g(\hat{\pi}_k) = \log[\hat{\pi}_k / (1 - \hat{\pi}_k)] = 12,440 - 0,201x_{k1} + 2,205x_{k2},$$

worin die beiden Parameter tzell und lai nach dem Wald-Test zum 5% Niveau signifikant sind (siehe Tabelle Variables in the Equation).

Für den Wert der link Funktion und die geschätzte Wahrscheinlichkeit für $Y = 1 =$ "krank" erhält man:

- wenn z.B. für tzell der Wert 72 und für lai der Wert 0 (negativ) angenommen wird
 $g[\hat{\pi}(72; 0)] = 12,440 - 0,201 \cdot 72 = 12,440 - 14,472 = -2,032;$
 $\hat{\pi}(72; 0) = \exp(-2,032) / [1 + \exp(-2,032)] = 0,1311 / 1,1311 = 0,1159.$

Der betrachtete Fall ist somit mit einer Wahrscheinlichkeit von 11,6% krank, wenn die T-Zelltypisierung einen Wert 72 und der LAI-Test zu einem negativen Ergebnis führt. Die geschätzte Wahrscheinlichkeit, dass er gesund ist, beträgt $(1 - \hat{\pi}_k) = 0,884$.

- wenn z.B. für tzell der Wert 77 und für lai der Wert 0 (negativ) angenommen wird
 $g[\hat{\pi}(77; 0)] = 12,440 - 0,201 \cdot 77 = 12,440 - 15,477 = -3,037;$
 $\hat{\pi}(77; 0) = \exp(-3,037) / [1 + \exp(3,037)] = 0,048 / 1,048 = 0,046.$

Der betrachtete Fall ist somit mit einer Wahrscheinlichkeit von 4,6% krank, wenn die T-Zelltypisierung einen Wert 77 und der LAI-Test zu einem negativen Ergebnis führt. Die geschätzte Wahrscheinlichkeit, dass er gesund ist, beträgt $(1 - \hat{\pi}_k) = 0,954$.

- wenn z.B. für tzell der Wert 72 und für lai der Wert 1 (positiv) angenommen wird
 $g[\hat{\pi}(72; 1)] = 12,440 - 0,201 \cdot 72 + 2,205 = 12,440 - 14,472 + 2,205 = 0,173$;
 $\hat{\pi}(72; 1) = \exp(0,173) / [1 + \exp(0,173)] = 1,189 / 2,189 = 0,543$.

Der betrachtete Fall ist somit mit einer Wahrscheinlichkeit von 54,3% krank, wenn die T-Zelltypisierung einen Wert 72 und der LAI-Test zu einem positiven Ergebnis führt. Die geschätzte Wahrscheinlichkeit, dass er gesund ist, beträgt $(1 - \hat{\pi}_k) = 0,457$.

Nimmt man bei der T-Zelltypisierung eine Veränderung von $a = 5$ Einheiten an, so ergibt sich:

$$\log[or(77; 0|72; 0)] = g[\hat{\pi}(77; 0)] - g[\hat{\pi}(72; 0)] = -3,037 - (-2,032) = -1,005,$$

was identisch ist mit $5 \cdot (-0,201)$.

Des weiteren ist:

$$or(77; 0|72; 0) = \exp(-1,005) = 0,366.$$

In der Stichprobe beträgt die Häufigkeit des Diagnoseergebnisses “krank“ bei Anwendung der T-Zelltypisierung und des LAI-Tests rund 1/3 bei Personen mit einem um 5 Einheiten höheren Wert der T-Zelltypisierung, wenn der LAI-Test ein negatives Ergebnis (0) liefert.

Hält man die T-Zelltypisierung bei einem Wert 72 konstant und geht beim LAI-Test von einem negativen zu einem positiven Wert über, so ergibt sich:

$$\log[or(72; 1|72; 0)] = g[\hat{\pi}(72; 1)] - g[\hat{\pi}(72; 0)] = 0,173 - (-2,032) = 2,205,$$

was identisch ist mit b_2 für den LAI-Test.

Des weiteren ist:

$$or(72; 1|72; 0) = \exp(2,205) = 9,07.$$

In der Stichprobe tritt die Diagnose “krank“ bei Anwendung der T-Zelltypisierung und des LAI-Tests rund 9 mal häufiger beim LAI-Test mit einem positiven Ergebnis (1) im Vergleich zum LAI-Test mit einem negativen Ergebnis (0) auf, wenn die T-Zelltypisierung einen konstanten Wert von 72 aufweist.

Auf eine Reihe weiterer Probleme soll hier nicht eingegangen werden. Dazu gehören u.a.

- Interaktionen und ihre Einbeziehung in das Modell

Interaktion kann anhand des folgenden einfachen Beispiels erklärt werden: In (3.18) sei X_1 eine dichotome X-Variable (z.B. Geschlecht) und X_2 eine metrische Variable (z.B. Alter) enthalten. Das Vorhandensein einer Interaktion zwischen diesen beiden erklärenden Variablen zeigt sich darin, dass der Parameter β_2 bei der Variablen X_2 sich für die beiden Geschlechter unterscheidet.

- die Auswahl der erklärenden Variablen¹⁷

Hierbei sollte man zunächst für jede X-Variable eine univariate Analyse durchführen, um anhand der Gütekriterien eine erste Einschätzung ihrer Bedeutung für die Schätzung von Y zu erhalten. Für den multivariaten Fall kann dann eine schrittweise Methode gewählt werden, wobei entweder die sogenannte Vorwärts-Selektion, die Rückwärts-Selektion oder die Stepwise-Selektion Anwendung findet.

- die Modelldiagnose¹⁸

Ist ein (im statistischen Sinne) “bestes“ Modell gefunden, sollten nicht nur die Gütekriterien zur Einschätzung des Gesamtmodells herangezogen werden, sondern auch eine eingehende Analyse der Residuen erfolgen.

- Nichtlineare logits

Hier wurde nur die logistische Regression unter der Voraussetzung der Linearität der logits (3.18) behandelt. Diese Voraussetzung muss geprüft werden, was u.a.

¹⁷Hierzu sei u.a. auf Hosmer, D.W., Lemeshow, S. (1989), Kapitel 4 verwiesen, wo dieses Problem anhand eines Beispiels ausführlich diskutiert wird.

¹⁸Siehe u.a. Collet, D. (1991), Kapitel 5, in dem ausführlich die Modelldiagnose behandelt wird.

mittels eines Scatterplots der response Variablen gegen jede erklärende Variable erfolgen kann. Gegebenenfalls muss dafür die X-Variable in Klassen eingeteilt werden.

Beispiel 3.4: Soziale Umfrage¹⁹

Mit diesem Beispiel soll die Schätzung eines Logit Modells demonstriert werden, wenn die Daten gruppiert vorliegen. Bei Verwendung von SPSS unter Windows Release 10.0.7²⁰ ist zur Schätzung die Prozedur Regression Probit bei Wahl von Model Logit zu verwenden.

Die Daten beziehen sich auf zwei soziale Umfragen in den Jahren 1974 und 1975 durch das National Opinion Research Center, University of Chicago, Illinois, und beinhalten das Problem der Rolle der Frau in der Gesellschaft. Keine der befragten Personen wurde bei beiden Umfragen erfasst. Jede ausgewählte Person wurde gefragt, ob sie/er mit der Behauptung „Frauen sollten sich um den Haushalt kümmern und die Führung der Gesellschaft den Männern überlassen“ übereinstimmen oder nicht. Die vorliegenden Variablen sind:

ed	education (Bildungsdauer, in Jahren)
sex	Geschlecht mit 0 - male (männlich) und 1 - female (weiblich),
na _k	Anzahl der Zustimmungen (agree) je Bildungsdauer
nd _k	Anzahl der Ablehnungen (disagree) je Bildungsdauer
n _k	Anzahl der Befragten je Bildungsdauer; $n_k = na_k + nd_k$
pa _k	Anteil der Zustimmungen je Bildungsdauer; $pa_k = na_k/n_k$
pd _k	Anteil der Ablehnungen je Bildungsdauer; $pd_k = nd_k/n_k$

Die interessierende Frage ist hier:

In wieweit wird die Häufigkeit für Zustimmung bzw. Ablehnung von der Bildungsdauer beeinflusst und gibt es dabei Unterschiede zwischen Männern und Frauen?

Es wird ein Signifikanzniveau $\alpha = 0,05$ vereinbart.

Die Ausgangsdaten für das Beispiel sind in Tabelle 3.11 angegeben.

Aus der Tabelle ergibt sich:

Gesamtstichprobenumfang $n = 2871$; $n(\text{männlich}) = 1305$; $n(\text{weiblich}) = 1566$

¹⁹Dieses Beispiel wurde Collet, D. (1991), S. 9 ff. entnommen.

²⁰Siehe SPSS for Windows, Advanced Statistics, SPSS Regression Models 9.0

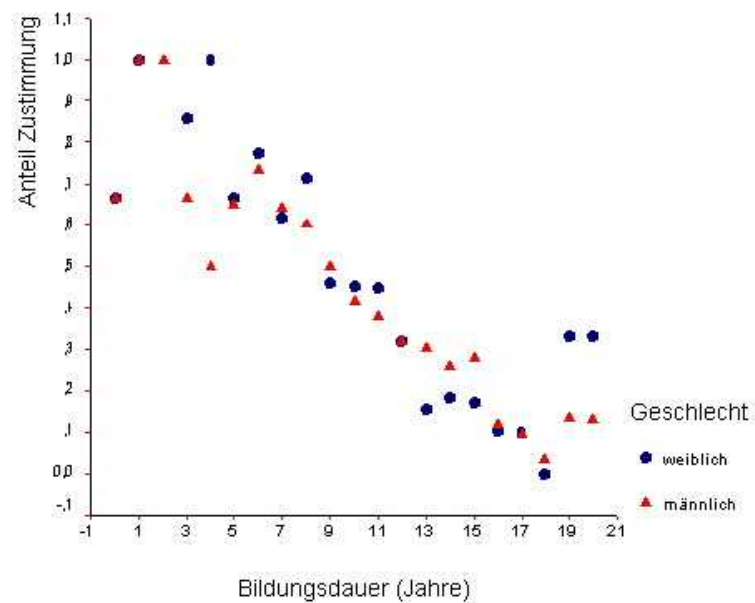
3.3 Zur Interpretation der Ergebnisse des logistischen Modells

$n(\text{agree}) = 1020$ $n(\text{disagree}) = 1851$
 $n(\text{männlich} \cap \text{agree}) = 465$ $n(\text{männlich} \cap \text{disagree}) = 840$
 $n(\text{weiblich} \cap \text{agree}) = 555$ $n(\text{weiblich} \cap \text{disagree}) = 1011$

Tabelle 3.11: Ausgangsdaten für das Beispiel 3.4

ed	sex	n_k	na_k	nd_k	pa_k	pd_k
0	0	6	4	2	,6667	,3333
1	0	2	2	0	1,0000	,0000
2	0	4	4	0	1,0000	,0000
3	0	9	6	3	,6667	,3333
4	0	10	5	5	,5000	,5000
5	0	20	13	7	,6500	,3500
6	0	34	25	9	,7353	,2647
7	0	42	27	15	,6429	,3571
8	0	124	75	49	,6048	,3952
9	0	58	29	29	,5000	,5000
10	0	77	32	45	,4156	,5844
11	0	95	36	59	,3789	,6211
12	0	360	115	245	,3194	,6806
13	0	101	31	70	,3069	,6931
14	0	107	28	79	,2617	,7383
15	0	32	9	23	,2813	,7188
16	0	125	15	110	,1200	,8800
17	0	32	3	29	,0938	,9063
18	0	29	1	28	,0345	,9655
19	0	15	2	13	,1333	,8667
20	0	23	3	20	,1304	,8696
0	1	6	4	2	,6667	,3333
1	1	1	1	0	1,0000	,0000
3	1	7	6	1	,8571	,1429
4	1	10	10	0	1,0000	,0000
5	1	21	14	7	,6667	,3333
6	1	22	17	5	,7727	,2273
7	1	42	26	16	,6190	,3810
8	1	127	91	36	,7165	,2835
9	1	65	30	35	,4615	,5385
10	1	122	55	67	,4508	,5492
11	1	112	50	62	,4464	,5536
12	1	593	190	403	,3204	,6796
13	1	109	17	92	,1560	,8440
14	1	99	18	81	,1818	,8182
15	1	41	7	34	,1707	,8293
16	1	128	13	115	,1016	,8984
17	1	31	3	28	,0968	,9032
18	1	21	0	21	,0000	1,0000
19	1	3	1	2	,3333	,6667
20	1	6	2	4	,3333	,6667

Abbildung 3.6: Scatterplot Anteil Zustimmung (pa) gegen Bildungsdauer (ed)



Aus den Daten der Tabelle 3.11 und aus Abb. 3.6 ist ersichtlich, dass mit längerer Bildungsdauer eine deutliche Tendenz zur Ablehnung der Behauptung auftritt. Um die Beziehung zwischen Dauer der Bildung und dem Antwortverhalten zu schätzen, wird das Logit Modell

$$g(\pi_k) = \log[\pi_k / (1 - \pi_k)] = \beta_0 + \beta_1 x_k$$

verwendet.

Schätzung A: Schätzung ohne Verwendung der Variablen sex

Parameter Estimates (Logit model: $\text{LOG}(p/(1-p)) = \text{Intercept} + \text{BX}$):

	Regression Coeff.	Standard Error	Coeff./S.E.
ED	-,27065	,01541	-17,56136
Intercept	2,50334	,17843	14,02981
Pearson Goodness-of-Fit	Chi Square = 72,625	DF = 39	P = ,001

3.3 Zur Interpretation der Ergebnisse des logistischen Modells

Für die geschätzte link Funktion resultiert aus dieser Schätzung:

$$g(\hat{\pi}_k) = \log[\hat{\pi}_k / (1 - \hat{\pi}_k)] = 2,50334 - 0,27065x_k,$$

worin die x_k die Beobachtungswerte der Variablen Bildungsdauer (ed) sind.

Mit Coeff./S.E. = -17,56136 lässt sich der Wert des Wald Test

$$W = [b_j / s(b_j)]^2 \quad (3.41)$$

mit $W = 308,40$ berechnen, dessen Überschreitungswahrscheinlichkeit $\text{Sig} = ,0000$ ist.

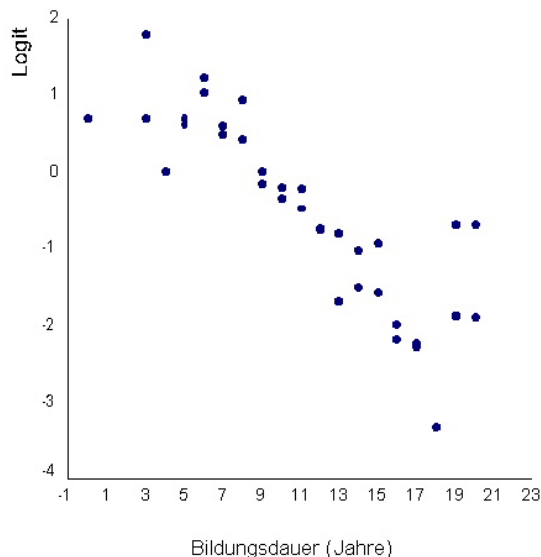
Der Parameter b_1 für die Bildungsdauer ist auf dem 5%-Niveau signifikant.

Jedoch ist die Modellanpassung nicht gut, wie Pearsons Goodness-of-Fit Chi Square

$$\chi^2 = \sum_{k=1}^K \frac{(\text{Residuum}_k)^2}{n_k \pi_k (1 - \pi_k)} \quad (3.63)$$

zeigt. Die Nullhypothese, dass das angegebene Logit Modell das angemessene Modell ist, wird auf dem 5% Niveau abgelehnt. Die Ursache dafür kann darin liegen, dass die Daten heterogen sind, was in einer unterschiedlichen Streuung der beobachteten logit Werte um das Modell resultiert. Dies ist offensichtlich für dieses Beispiel vor allem bei geringer bzw. hoher Bildungsdauer gegeben.

Abbildung 3.7: Logit-transformierte Responses



3 Modellierung binärer Daten

Die Aufnahme der Variablen sex als weitere Covariate führt zu keinen substantiellen Verbesserungen, da der geschätzte Parameter der Variablen sex nicht signifikant zum 5% Niveau ist.

Um Unterschiede im Geschlecht zu erkennen, werden nunmehr Schätzungen des Logit Modells getrennt nach dem Geschlecht durchgeführt.

Schätzung B: Schätzung für Männer und Frauen getrennt

Männlich (sex = 0):

Parameter Estimates (Logit model: $\text{LOG}(p/(1-p)) = \text{Intercept} + \text{BX}$):

	Regression Coeff.	Standard Error	Coeff./S.E.
ED	-,23403	,02019	-11,59226
	Intercept	Standard Error	Intercept/S.E.
	2,09820	,23550	8,90959
Pearson Goodness-of-Fit Chi Square = 19,407 DF = 19 P = ,431			

Weiblich (sex = 1):

Parameter Estimates (Logit model: $\text{LOG}(p/(1-p)) = \text{Intercept} + \text{BX}$):

	Regression Coeff.	Standard Error	Coeff./S.E.
ED	-,31541	,02365	-13,33799
	Intercept	Standard Error	Intercept/S.E.
	3,00295	,27238	11,02484
Pearson Goodness-of-Fit Chi Square = 53,336 DF = 18 P = ,000			

Während bei den Männern die Modellanpassung gut ist, ist sie bei den Frauen nicht akzeptabel. Die nicht ausreichende Modellanpassung der Schätzung A ist somit vor allem auf die Heterogenität der Daten bei den Frauen zurückzuführen.

Abbildung 3.8: Logit-transformierte Responses (Männer)

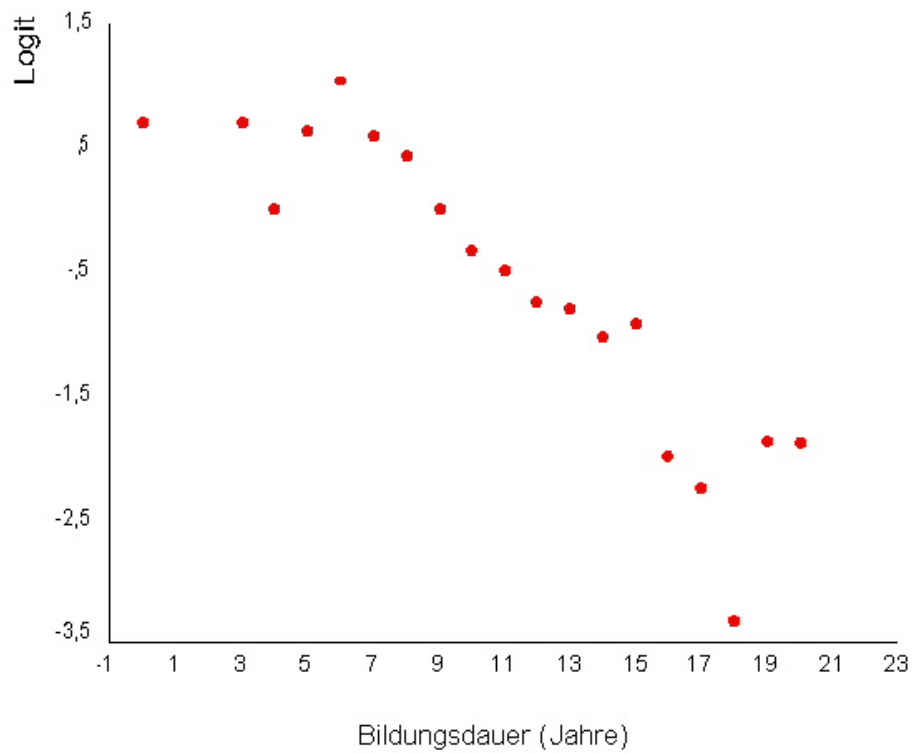
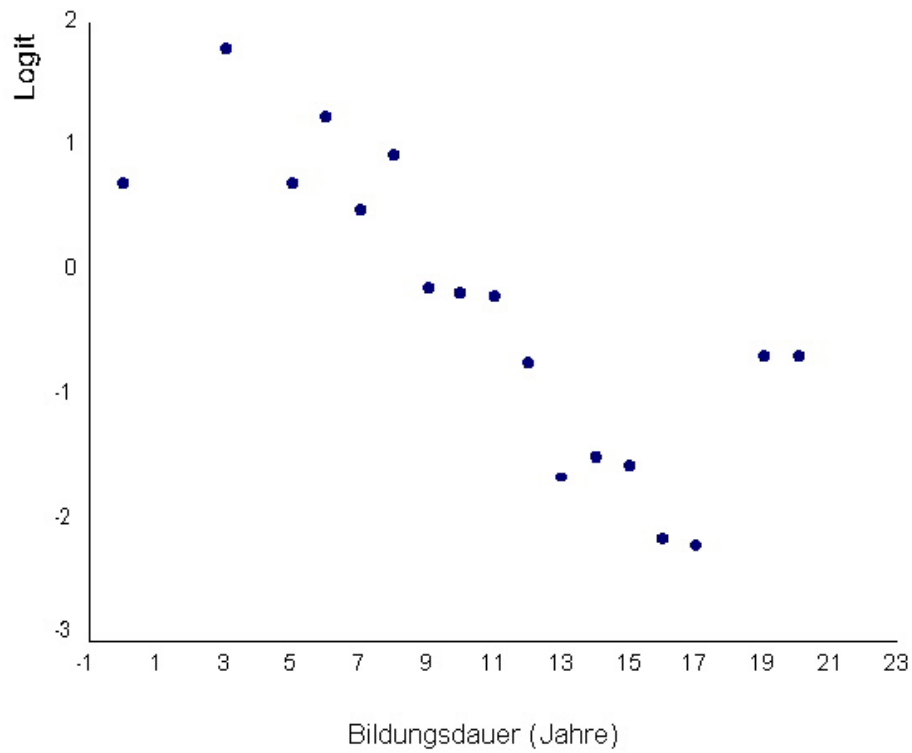


Abbildung 3.9: Logit-transformierte Responses (Frauen)



3.4 Das Probit-Modell

Als Toleranz-Verteilung kann auch die Dichtefunktion der Standardnormalverteilung $N(0;1)$ verwendet werden, so dass in (3.13) die Verteilungsfunktion der $N(0;1)$, bezeichnet mit Φ , steht:

$$\pi_k = g^{-1}(\eta_k) = \Phi(\eta_k) = \int_{-\infty}^{\eta_k} f(z)dz, \quad (3.64)$$

d.h., η_k ist derjenige Wert, für den die Verteilungsfunktion der Standardnormalverteilung $\Phi(\eta_k)$ den Wert π_k annimmt.

Daraus folgt unmittelbar

$$\eta_k = \Phi^{-1}(\pi_k) \quad (3.65)$$

mit Φ^{-1} als Inverse der Verteilungsfunktion der $N(0;1)$. Es ist diese inverse Transformation $\Phi^{-1}(\pi_k)$, die als probit bezeichnet wird²¹:

$$g(\pi_k) = \eta_k = \text{probit}(\pi_k) = \Phi^{-1}(\pi_k). \quad (3.66)$$

Die Probit-Funktion ist symmetrisch in π , und für jeden Wert von π im Bereich $(0;1)$ liegt der korrespondierende Wert von $\text{probit}(\pi)$ zwischen $-\infty$ und ∞ .

Obwohl es keinen einfachen analytischen Ausdruck für die Inverse der Verteilungsfunktion $\Phi(\eta)$ gibt, kann man das Argument η für gegebene Werte Φ unter Verwendung der Tabelle der Verteilungsfunktion von $N(0;1)$ finden.

Ein einfaches Beispiel möge die Probit-Transformation verdeutlichen. Betrachtet man eine zufällige Auswahl von Kunden und deren Kaufentscheidung für ein bestimmtes Produkt bei gegebenem Preis dieses Produktes. Ein niedriger Preis wird viele Kunden zum Kauf veranlassen, während ein hoher Preis das Gegenteil bewirkt. Der Preis ist eine die Kaufentscheidung (response Variable mit ja = 1 oder nein = 0) beeinflussende Variable. Ob ein Kunde kauft oder nicht, wird dabei von seiner eigenen "Toleranz" gegenüber dem Preis bestimmt; Kunden mit einem niedrigen individuellen Toleranzniveau kaufen eher

²¹Die ursprüngliche Definition von probit war: $\text{probit}(\pi) = 5 + \Phi^{-1}(\pi)$, um negative Werte zu vermeiden.

als jene mit einem hohen Toleranzniveau. In der Grundgesamtheit, aus der die Kunden zufällig gezogen wurden, existiert eine Verteilung des Toleranzniveaus. S sei die mit der Verteilung des Toleranzniveaus verbundene Zufallsvariable, s das Toleranzniveau eines speziellen Kunden und $f(s)$ die Dichtefunktion von S . Dann ist die Wahrscheinlichkeit eines Kaufs bei einem Preis x_k gegeben mit

$$\pi_k = P(S \leq x_k) = \int_{-\infty}^{x_k} f(s) ds.$$

Wenn nunmehr angenommen wird, dass die Toleranz normalverteilt ist, folgt S einer Normalverteilung mit Mittelwert μ und Varianz σ^2 :

$$f(s) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{s - \mu}{\sigma} \right)^2 \right\}, \quad -\infty < s < \infty$$

und

$$\pi_k = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_k} \exp \left\{ -\frac{1}{2} \left(\frac{s - \mu}{\sigma} \right)^2 \right\} ds = \Phi \left(\frac{x_k - \mu}{\sigma} \right).$$

Mit $\beta_0 = -\mu/\sigma$ und $\beta_1 = 1/\sigma$ kann dies auch als $\pi_k = \Phi(\beta_0 + \beta_1 x_k)$ geschrieben werden, und es ist

$$\eta_k = \text{probit}(\pi_k) = \Phi^{-1}(\pi_k) = \beta_0 + \beta_1 x_k.$$

Dies ist das Probit-Modell der Beziehung zwischen Kaufwahrscheinlichkeit π_k und Preis des Produkts x_k und resultiert als Ergebnis der Annahme einer Normalverteilung für die Toleranz.

Allgemein ist die gesuchte link Funktion des Probit-Modells:

$$g(\pi_k) = \eta_k = \text{probit}(\pi_k) = \mathbf{x}_k^T \boldsymbol{\beta}. \quad (3.67)$$

Für eine einfache Zufallsstichprobe von n Beobachtungen über eine dichotome (punktbinomiale) Zufallsvariable Y_k mit $y_k = 1$ oder 0 ($k = 1, \dots, n$) ist die Likelihood-Funktion

$$L(\pi_1, \dots, \pi_n; y_1, \dots, y_n) = \prod_{k=1}^n f(y_k; \pi_k) = \prod_{k=1}^n \pi_k^{y_k} (1 - \pi_k)^{1-y_k} \quad (3.68)$$

und die log-Likelihood-Funktion

$$l(\pi_1, \dots, \pi_n; y_1, \dots, y_n) = \sum_{k=1}^n [y_k \log(\pi_k) + (1 - y_k) \log(1 - \pi_k)]. \quad (3.69)$$

Einsetzen von π_k gemäß (3.64) und η_k gemäß (3.67) ergibt:

$$\begin{aligned} l(\boldsymbol{\pi}; \mathbf{y}) &= \sum_{k=1}^n \{y_k \log[\Phi(\eta_k)] + (1 - y_k) \log[1 - \Phi(\eta_k)]\} \\ &= \sum_{k=1}^n \{y_k \log[\phi(x_k^T \boldsymbol{\beta})] + (1 - y_k) \log[1 - \Phi(x_k^T \boldsymbol{\beta})]\}. \end{aligned} \quad (3.70)$$

Die partiellen Ableitungen der log-Likelihood-Funktion (3.70) nach β_j ($j = 1, \dots, J$) führen zu nichtlinearen Gleichungen. Allgemein wird auch hier der iterative Newton-Raphson-Algorithmus²² zur Lösung angewandt. Die Schätzung der unbekannten Parameter $\boldsymbol{\beta}$ des Probit-Modells und die Einschätzung der Modellgüte verläuft in analoger Weise wie beim Logit Modell.

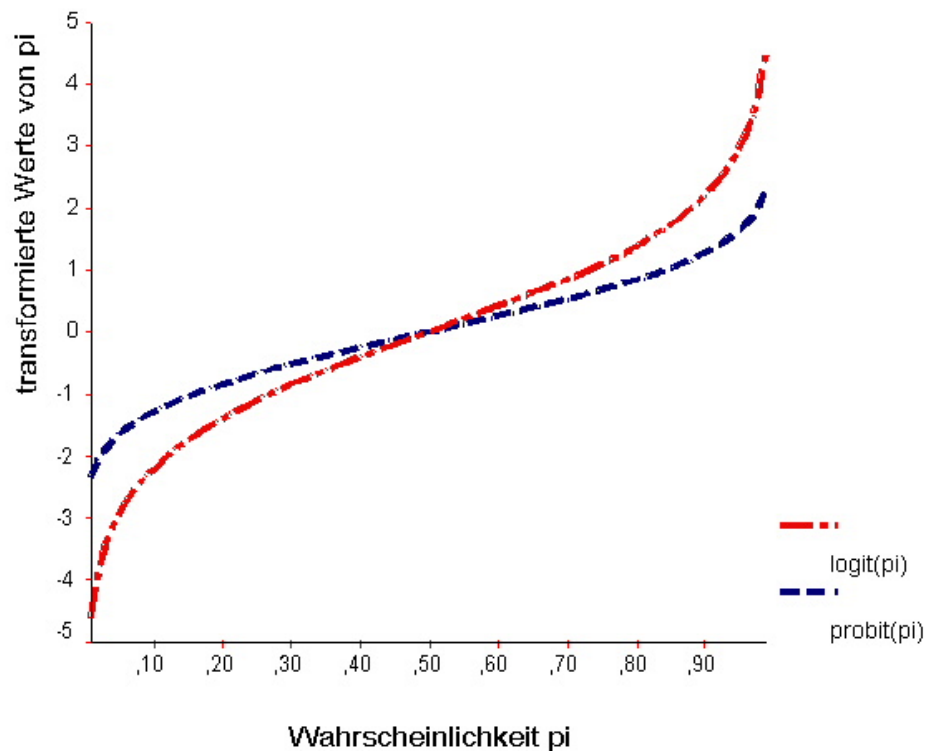
Die Schätzergebnisse des Probit-Modells für $\hat{\pi}_k$ unterscheiden sich im allgemeinen wenig von denen des Logit Modells. Während das Logit Modell im allgemeinen angemessener für Beobachtungsdaten ist, wird das Probit-Modell hauptsächlich auf die Ergebnisse von Experimenten angewandt.

Die Probit-Transformation hat die gleiche allgemeine Gestalt wie die Logit-Transformation. Dies zeigen Tabelle 3.12, in der für ausgewählte Werte von π sowohl $\logit(\pi)$ als auch $\text{probit}(\pi)$ angegeben sind, und die Abbildung 3.10.

²²Siehe u.a. Judge, G.G., Hill, R.C., Griffiths, W.E., Lütkepohl, H., Lee, T.-Chao (1988), Introduction to the Theory and Practice of Econometrics, Wiley, New York et al., S. 792

Tabelle 3.12: $\text{logit}(\pi)$ und $\text{probit}(\pi)$ für $\pi = 0,01(0,01)1,00$

π	$\text{logit}(\pi)$	$\text{probit}(\pi)$	π	$\text{logit}(\pi)$	$\text{probit}(\pi)$
,01	-4,5951	-2,3263	,51	,0400	,0251
,02	-3,8918	-2,0537	,52	,0800	,0502
,03	-3,4761	-1,8808	,53	,1201	,0753
,04	-3,1781	-1,7507	,54	,1603	,1004
,05	-2,9444	-1,6449	,55	,2007	,1257
,06	-2,7515	-1,5548	,56	,2412	,1510
,07	-2,5867	-1,4758	,57	,2819	,1764
,08	-2,4423	-1,4051	,58	,3228	,2019
,09	-2,3136	-1,3408	,59	,3640	,2275
,10	-2,1972	-1,2816	,60	,4055	,2533
,11	-2,0907	-1,2265	,61	,4473	,2793
,12	-1,9924	-1,1750	,62	,4895	,3055
,13	-1,9010	-1,1264	,63	,5322	,3319
,14	-1,8153	-1,0803	,64	,5754	,3585
,15	-1,7346	-1,0364	,65	,6190	,3853
,16	-1,6582	-,9945	,66	,6633	,4125
,17	-1,5856	-,9542	,67	,7082	,4399
,18	-1,5163	-,9154	,68	,7538	,4677
,19	-1,4500	-,8779	,69	,8001	,4959
,20	-1,3863	-,8416	,70	,8473	,5244
,21	-1,3249	-,8064	,71	,8954	,5534
,22	-1,2657	-,7722	,72	,9445	,5828
,23	-1,2083	-,7388	,73	,9946	,6128
,24	-1,1527	-,7063	,74	1,0460	,6433
,25	-1,0986	-,6745	,75	1,0986	,6745
,26	-1,0460	-,6433	,76	1,1527	,7063
,27	-,9946	-,6128	,77	1,2083	,7388
,28	-,9445	-,5828	,78	1,2657	,7722
,29	-,8954	-,5534	,79	1,3249	,8064
,30	-,8473	-,5244	,80	1,3863	,8416
,31	-,8001	-,4959	,81	1,4500	,8779
,32	-,7538	-,4677	,82	1,5163	,9154
,33	-,7082	-,4399	,83	1,5856	,9542
,34	-,6633	-,4125	,84	1,6582	,9945
,35	-,6190	-,3853	,85	1,7346	1,0364
,36	-,5754	-,3585	,86	1,8153	1,0803
,37	-,5322	-,3319	,87	1,9010	1,1264
,38	-,4895	-,3055	,88	1,9924	1,1750
,39	-,4473	-,2793	,89	2,0907	1,2265
,40	-,4055	-,2533	,90	2,1972	1,2816
,41	-,3640	-,2275	,91	2,3136	1,3408
,42	-,3228	-,2019	,92	2,4423	1,4051
,43	-,2819	-,1764	,93	2,5867	1,4758
,44	-,2412	-,1510	,94	2,7515	1,5548
,45	-,2007	-,1257	,95	2,9444	1,6449
,46	-,1603	-,1004	,96	3,1781	1,7507
,47	-,1201	-,0573	,97	3,4761	1,8808
,48	-,0800	-,0502	,98	3,8918	2,0537
,49	-,0400	-,0251	,99	4,5951	2,3263
,50	,0000	,0000	1,00	∞	∞

Abbildung 3.10: Logit- und probit-Transformation als Funktion von π Beispiel 3.2 (Fortsetzung):

Für die Schätzung eines Probit-Modells unter Verwendung von SPSS müssen die Daten gruppiert vorliegen. Es wird deshalb von der Altersgruppierung ausgegangen, die in der (Kontingenz-)Tabelle 3.6 enthalten ist. Die Schätzung wird mit der Prozedur Regression Probit von SPSS unter Windows Release 10.0.7 realisiert.

Als Output erhält man:

*****P R O B I T A N A L Y S I S *****

Parameter Estimates (PROBIT model: (PROBIT(p)) = Intercept + BX):

	Regression Coeff.	Standard Error	Coeff./S.E.
AGGR	,33422	,06736	4,96161
Intercept	-1,72015	,34556	-4,97785

Pearson Goodness-of-Fit Chi-Square = ,270 DF = 6 P = 1,000

Since Goodness-of-Fit Chi square is NOT significant, no heterogeneity factor is used in the calculation of confidence limits.

Observerd and Expected Frequencies

AGGR	Number of Subjects	Observed Responses	Expected Responses	Residual	Prob
1,00	10,0	1,0	,829	,171	,08288
2,00	15,0	2,0	2,197	-,197	,14646
3,00	12,0	3,0	2,838	,162	,23653
4,00	15,0	5,0	5,261	-,261	,35075
5,00	13,0	6,0	6,246	-,246	,48043
6,00	8,0	5,0	4,898	,102	,61223
7,00	17,0	13,0	12,447	,553	,73216
8,00	10,0	8,0	8,298	-,298	,82985

Für die geschätzte link Funktion des Probit-Modells resultiert:

$$g(\hat{\pi}_k) = \hat{\eta}_k = \text{probit}(\hat{\pi}_k) = -1,72015 + 0,33422 \cdot \text{Altersgruppe}.$$

Setzt man nacheinander für Altersgruppe die Werte 1 bis 8 ein, dann ergeben sich die in Tabelle 3.13 enthaltenen probits. Die Modellanpassung ist gut, wie Pearson Goodness-of-Fit Chi-Square anzeigt, d.h., die Nullhypothese, dass das Probit-Modell das angemessene Modell ist, wird nicht abgelehnt. Die geschätzten Parameter sind zum 5 % Niveau signifikant.

Zum Vergleich wird für die gleiche Altersgruppierung ein Logit Modell geschätzt, das zu nachstehenden Resultaten führt:

Parameter Estimates (LOGIT model: $(\text{LOG}(p/(1-p))) = \text{Intercept} + \text{BX}$):

	Regression Coeff.	Standard Error	Coeff./S.E.
AGGR	,55812	,12056	4,62934
	Intercept	Standard Error	Intercept/S.E.
	-2,87741	,62339	-4,61573

Pearson Goodness-of-Fit Chi-Square = ,218 DF = 6 P = 1,000

Since Goodness-of-Fit Chi square is NOT significant, no heterogeneity factor is used in the calculation of confidence limits.

Observerd and Expected Frequencies

AGGR	Number of Subjects	Observed Responses	Expected Responses	Residual	Prob
1,00	10,0	1,0	,895	,105	,08954
2,00	15,0	2,0	2,200	-,200	,14664
3,00	12,0	3,0	2,771	,229	,23093
4,00	15,0	5,0	5,162	-,162	,34413
5,00	13,0	6,0	6,218	-,218	,47831
6,00	8,0	5,0	4,926	,074	,61569
7,00	17,0	13,0	12,526	,474	,73680
8,00	10,0	8,0	8,303	-,303	,83027

3 Modellierung binärer Daten

Für die geschätzte link Funktion des Logit Modells resultiert:

$$g(\hat{\pi}_k) = \hat{\eta}_k = \log[\hat{\pi}_k/(1 - \hat{\pi}_k)] = -2,87741 + 0,55812 \cdot \text{Altersgruppe}.$$

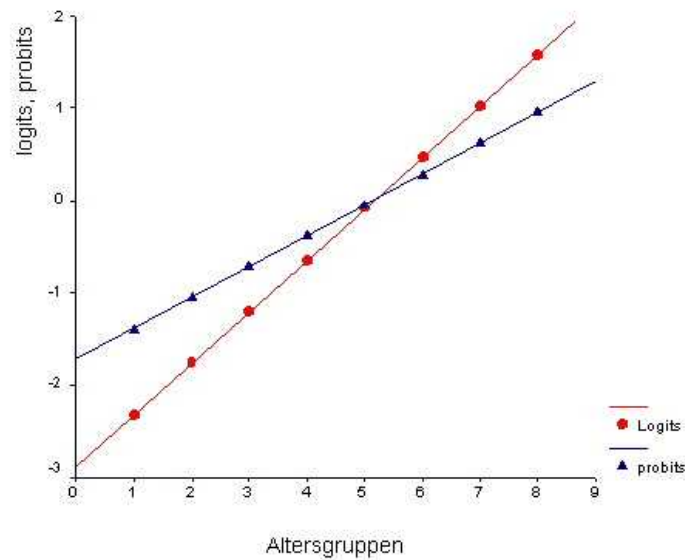
Setzt man nacheinander für Altersgruppe die Werte 1 bis 8 ein, dann ergeben sich die in Tabelle 3.13 enthaltenen logits. Die Modellanpassung ist ebenfalls gut und die geschätzten Parameter sind zum 5% Niveau signifikant.

Tabelle 3.13: probits und logits für die Altersgruppen

Alter	Altersgruppe	probits	$\hat{\pi}_k$	logits	$\hat{\pi}_k$
1	20 - 29	-1,38593	,08288	-2,31929	,08954
2	30 - 34	-1,05171	,14646	-1,76117	,14664
3	35 - 39	-,71749	,23653	-1,20305	,23093
4	40 - 44	-,38327	,35075	-,64493	,34413
5	45 - 49	-,04905	,48043	-,08681	,47831
6	50 - 54	,28517	,61223	,47131	,61569
7	55 - 59	,61939	,73216	1,02943	,73680
8	60 - 69	,95361	,82985	1,58755	,83027

Die geschätzten Wahrscheinlichkeiten aus dem Probit- und dem Logit Modell unterscheiden sich nur geringfügig.

Abbildung 3.11: Logit- und probit-Transformation für das Beispiel 3.2



4 Das multinomiale Logit Modell

Das Logit Modell kann auf den Fall erweitert werden, dass die response Variable mehr als zwei Kategorien aufweist. Ein solches Modell wird im allgemeinen als multinomiales Logit Modell²³ bezeichnet.

Beispielsweise haben befragte Personen bei vielen praktischen sozio-ökonomischen Untersuchungen, die auf Umfragen basieren, die Wahl zwischen verschiedenen Antwortmöglichkeiten. Die response Variable ist dann eine kategoriale Variable nominalen oder ordinalen Skalenniveaus mit mehreren Ausprägungen.

Im weiteren sei davon ausgegangen, dass die response Variable Y G Kategorien aufweist ($g = 1, \dots, G$).

Zusätzlich zur response Variablen Y gibt es eine Reihe von erklärenden Variablen X_j , $j = 1, \dots, J$, die beliebiges Skalenniveau aufweisen können. Die Werte der X -Variablen werden als fest vorgegeben angesehen (d.h., sie sind keine Zufallsvariablen). Aufgrund der unterschiedlichen Werte der X -Variablen in der Stichprobe ergeben sich K verschiedene Vektoren von X -Werten $\mathbf{x}_k^T = (x_{k0}, x_{k1}, \dots, x_{kj}, \dots, x_{kJ})$, $k = 1, \dots, K$, wobei vereinbarungsgemäß $x_{k0} = 1$ für die Konstante des Modells steht. Jeder Vektor \mathbf{x}_k charakterisiert eine Gruppe, in der die statistischen Einheiten (Fälle) bezüglich der erklärenden Variablen X_j ($j = 1, \dots, J$) dieselben Beobachtungswerte aufweisen.

²³Vgl. u.a. Hosmer, D.W., Lemeshow, S. (1989), S. 216 ff., Agresti, A. (1996), S. 205 ff., Tutz, G. (2000), S. 159 ff.

Annahmen:

- Für jede Kombination der Werte der X-Variablen \mathbf{x}_k folgen die Häufigkeiten der response Variablen einer Multinomialverteilung mit festen Gesamthäufigkeiten n_k .
- Die Verteilungen der Häufigkeiten der response Variablen für die verschiedenen Kombinationen \mathbf{x}_k sind unabhängig voneinander.
- Es liegt eine einfache Zufallsstichprobe vom Umfang n vor, was impliziert, dass die Beobachtungen unabhängig voneinander sind.

Ebenso wie bei binären Logit Modellen steht im Mittelpunkt des Interesses die Schätzung von Effekten der erklärenden X-Variablen auf die Wahrscheinlichkeit des Auftretens der Realisationen der response Variablen Y .

Bezeichnet $\pi_g(\mathbf{x}_k) = P(Y = g|\mathbf{x}_k)$ die bedingte Wahrscheinlichkeit der g -ten Kategorie der response Variablen Y ($g = 1, \dots, G$) für die k -te Kombination der Werte der erklärenden X-Variablen, so ergibt sich für die response Funktion:

$$\pi_g(\mathbf{x}_k) = P(Y = g|\mathbf{x}_k) = \frac{e^{\boldsymbol{\beta}_g^T \mathbf{x}_k}}{\sum_{g=1}^G e^{\boldsymbol{\beta}_g^T \mathbf{x}_k}}, \quad g = 1, \dots, G, \quad (4.1)$$

worin $\boldsymbol{\beta}_g = (\beta_{g0}, \beta_{g1}, \dots, \beta_{gj}, \dots, \beta_{gJ})^T$ ein Vektor unbekannter Parameter und $\eta_{gk} = \boldsymbol{\beta}_g^T \mathbf{x}_k$ eine lineare Vorhersagegleichung sind.

Zur Erfüllung der Nebenbedingung $\sum_g \pi_g(\mathbf{x}_k) = 1$ für jede Kombination \mathbf{x}_k der Werte der erklärenden X-Variablen und zur Identifizierung von (4.1) wird für eine als Basis ausgewählte Kategorie der response Variablen Y der Vektor $\boldsymbol{\beta}_g = \mathbf{0}$ gesetzt. Im weiteren wird die letzte Kategorie G als Referenzkategorie gewählt, für die dann folgt:

$$\eta_{Gk} = \boldsymbol{\beta}_G^T \mathbf{x}_k = 0 \text{ und } \exp(\boldsymbol{\beta}_G^T \mathbf{x}_k) = 1.$$

Damit sind:

$$\pi_g(\mathbf{x}_k) = P(Y = G|\mathbf{x}_k) = \frac{e^{\boldsymbol{\beta}_g^T \mathbf{x}_k}}{1 + \sum_{g=1}^{G-1} e^{\boldsymbol{\beta}_g^T \mathbf{x}_k}}, \quad g = 1, \dots, G-1 \quad (4.2)$$

$$\pi_G(\mathbf{x}_k) = P(Y = g|\mathbf{x}_k) = \frac{1}{1 + \sum_{g=1}^{G-1} e^{\boldsymbol{\beta}_g^T \mathbf{x}_k}}, \quad g = G.$$

Die $\pi_g(\mathbf{x}_k)$, $g = 1, \dots, G$, erfüllen eine multinomiale logistische Verteilung.

Ein Vergleich der g -ten Kategorie von Y zur Referenzkategorie G führt zu:

$$\frac{\pi_g(\mathbf{x}_k)}{\pi_G(\mathbf{x}_k)} = \frac{\frac{e^{\boldsymbol{\beta}_g^T \mathbf{x}_k}}{1 + \sum_{g=1}^{G-1} e^{\boldsymbol{\beta}_g^T \mathbf{x}_k}}}{\frac{1}{1 + \sum_{g=1}^{G-1} e^{\boldsymbol{\beta}_g^T \mathbf{x}_k}}} = e^{\boldsymbol{\beta}_g^T \mathbf{x}_k}, \quad g = 1, \dots, G - 1 \quad (4.3)$$

Für die link Funktion der $G - 1$ separaten (binären) logistischen Modelle folgt:

$$g[\pi_g(\mathbf{x}_k)] = \eta_{gk} = \log \left[\frac{\pi_g(\mathbf{x}_k)}{\pi_G(\mathbf{x}_k)} \right] = \boldsymbol{\beta}_g^T \mathbf{x}_k \quad (4.4)$$

Im multinomialen Logit Modell ist somit die abhängige Variable als das logarithmierte Verhältnis der Wahrscheinlichkeit $\pi_g(\mathbf{x}_k)$ des Eintretens der g -ten Kategorie der response Variablen zur Wahrscheinlichkeit $\pi_G(\mathbf{x}_k)$ des Eintretens der Referenzkategorie G in Abhängigkeit von den Werten \mathbf{x}_k der erklärenden Variablen definiert (log odds oder logits)²⁴. Man erhält für jede nicht als Basis ausgewählte Kategorie der response Variablen einen Parametervektor, d.h. insgesamt $G - 1$.

Jede bedingte Wahrscheinlichkeit $\pi_g(\mathbf{x}_k)$, $g = 1, \dots, G$, ist eine Funktion von $(G - 1) \cdot (J + 1)$ Parametern $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_{G-1})$.

Ein Vergleich von jeweils zwei Kategorien r und s ($r, s \neq G$; $r \neq s$) der response Variablen

²⁴Der ordinale Charakter der response Variablen wird hier vernachlässigt. Siehe dazu u.a. Agresti, A. (1996).

ergibt sich als Differenz der logits der Kategorie r und der Kategorie s:

$$\frac{\pi_r(\mathbf{x}_k)}{\pi_s(\mathbf{x}_k)} = \frac{\frac{e^{\boldsymbol{\beta}_r^T \mathbf{x}_k}}{1 + \sum_{g=1}^{G-1} e^{\boldsymbol{\beta}_g^T \mathbf{x}_k}}}{\frac{e^{\boldsymbol{\beta}_s^T \mathbf{x}_k}}{1 + \sum_{g=1}^{G-1} e^{\boldsymbol{\beta}_g^T \mathbf{x}_k}}} = \frac{e^{\boldsymbol{\beta}_r^T \mathbf{x}_k}}{e^{\boldsymbol{\beta}_s^T \mathbf{x}_k}} = e^{(\boldsymbol{\beta}_r - \boldsymbol{\beta}_s)^T \mathbf{x}_k} \quad (4.5)$$

$$\log \left[\frac{\pi_r(\mathbf{x}_k)}{\pi_s(\mathbf{x}_k)} \right] = \boldsymbol{\beta}_r^T \mathbf{x}_k - \boldsymbol{\beta}_s^T \mathbf{x}_k = (\boldsymbol{\beta}_r - \boldsymbol{\beta}_s)^T \mathbf{x}_k.$$

Ausschließlich zum Zweck der Herleitung der log-Likelihood-Funktion wird für jede Kategorie von Y eine dichotome Variable Y_g eingeführt, mit $Y_g = 1$, wenn die Beobachtung die Kategorie g aufweist, und sonst 0. Diese binären Variablen kennzeichnen die Zugehörigkeit einer Beobachtung zu einer Kategorie von Y. Wenn z.B. $Y = 1$, dann sind $Y_1 = 1$ und $Y_2 = 0, Y_3 = 0, \dots, Y_G = 0$. Wie ersichtlich gilt für jeden möglichen Wert von Y: $\sum_g Y_g = 1$, und es ist: $\pi_g(\mathbf{x}_k) = P(Y = g|\mathbf{x}_k) = P(Y_g = 1|\mathbf{x}_k)$.

Die bedingte Likelihood-Funktion ist:

$$L(\boldsymbol{\beta}) = \prod_{k=1}^K \pi_1(\mathbf{x}_k)^{y_{1k}} \cdot \dots \cdot \pi_g(\mathbf{x}_k)^{y_{gk}} \cdot \dots \cdot \pi_G(\mathbf{x}_k)^{y_{Gk}} = \prod_{k=1}^K \prod_{g=1}^G \pi_g(\mathbf{x}_k)^{y_{gk}}, \quad (4.6)$$

womit für die log-Likelihood Funktion resultiert:

$$\begin{aligned} l(\boldsymbol{\beta}) &= \sum_{k=1}^K \sum_{g=1}^G y_{gk} \log \pi_g(\mathbf{x}_k) \\ &= \sum_{k=1}^K \left[\sum_{g=1}^{G-1} y_{gk} \log \pi_g(\mathbf{x}_k) + y_{Gk} \log \pi_G(\mathbf{x}_k) \right] \\ &= \sum_{k=1}^K \left[\sum_{g=1}^{G-1} y_{gk} \log \left(\frac{e^{\boldsymbol{\beta}_g^T \mathbf{x}_k}}{1 + \sum_{g=1}^{G-1} e^{\boldsymbol{\beta}_g^T \mathbf{x}_k}} \right) + y_{Gk} \log \left(\frac{1}{1 + \sum_{g=1}^{G-1} e^{\boldsymbol{\beta}_g^T \mathbf{x}_k}} \right) \right] \\ &= \sum_{k=1}^K \left[\sum_{g=1}^{G-1} y_{gk} \boldsymbol{\beta}_g^T \mathbf{x}_k - \sum_{g=1}^{G-1} y_{gk} \log \left(1 + \sum_{g=1}^{G-1} e^{\boldsymbol{\beta}_g^T \mathbf{x}_k} \right) - y_{Gk} \log \left(1 + \sum_{g=1}^{G-1} e^{\boldsymbol{\beta}_g^T \mathbf{x}_k} \right) \right] \end{aligned}$$

$$\begin{aligned}
&= \sum_{k=1}^K \left[\sum_{g=1}^{G-1} y_{gk} \boldsymbol{\beta}_g^T \mathbf{x}_k - \sum_{g=1}^G y_{gk} \log \left(1 + \sum_{g=1}^{G-1} e^{\boldsymbol{\beta}_g^T \mathbf{x}_k} \right) \right] \\
l(\boldsymbol{\beta}) &= \sum_{k=1}^K \left[\sum_{g=1}^{G-1} y_{gk} \boldsymbol{\beta}_g^T \mathbf{x}_k - \log \left(1 + \sum_{g=1}^{G-1} e^{\boldsymbol{\beta}_g^T \mathbf{x}_k} \right) \right] \tag{4.7}
\end{aligned}$$

(4.7) folgt unter Berücksichtigung, dass $\sum_g y_{gk} = 1$ für jedes k ist.

Die ersten partiellen Ableitungen der log-likelihood Funktion nach $\boldsymbol{\beta}_g$ ergibt:

$$\begin{aligned}
\frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_g} &= \sum_{k=1}^K y_{gk} \mathbf{x}_k - \frac{e^{\boldsymbol{\beta}_g^T \mathbf{x}_k}}{1 + \sum_{g=1}^{G-1} e^{\boldsymbol{\beta}_g^T \mathbf{x}_k}} \mathbf{x}_k \\
&= \sum_{k=1}^K [y_{gk} - \pi_g(\mathbf{x}_k)] \mathbf{x}_k \tag{4.8}
\end{aligned}$$

bzw. nach β_{gj}

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \beta_{gj}} = \sum_{k=1}^K [y_{gk} - \pi_g(\mathbf{x}_k)] \mathbf{x}_{kj} \tag{4.9}$$

Setzt man (4.8) bzw. (4.9) gleich Null und löst die Gleichungen für β_{gj} , so erhält man den ML-Schätzer. Die Schätzung des multinomialen Logit Modells erfordert ebenfalls einen iterativen Prozess (Newton-Raphson-Prozedur). Die Eigenschaften des Schätzers sind analog wie im binären Logit Modell.

Die zweiten partiellen Ableitungen werden für die $(G-1)(J+1) \times (G-1)(J+1)$ Informationsmatrix $\mathfrak{I}(\boldsymbol{\beta})$ benötigt:

$$\begin{aligned}
\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_{gj} \partial \beta_{gj^*}} &= - \sum_{k=1}^K x_{kj} x_{kj^*} \pi_g(\mathbf{x}_k) [1 - \pi_g(\mathbf{x}_k)] \\
\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_{gj} \partial \beta_{g^*j^*}} &= \sum_{k=1}^K x_{kj} x_{kj^*} \pi_g(\mathbf{x}_k) \pi_{g^*}(\mathbf{x}_k) \tag{4.10}
\end{aligned}$$

für $g, g^* = 1, \dots, G-1$ und $j, j^* = 0, 1, \dots, J$. Die Elemente der Informationsmatrix sind das Negative der Erwartungswerte von (4.10). Die asymptotische Varianz-Kovarianz-Matrix der ML-Schätzer ist die Inverse der Informationsmatrix: $\Sigma(\boldsymbol{\beta}) = \mathfrak{I}(\boldsymbol{\beta})^{-1}$.

Ebenso wie im binären Logit Modell werden für kategoriale X-Variablen Kontrast-Variablen geschaffen. Die Anzahl dieser Kontrast-Variablen entspricht der Anzahl der verschiedenen Ausprägungen der mehrkategorialen Variablen minus Eins, da eine Kategorie als Bezugskategorie dient. Weist die kategoriale X-Variable z.B. H Kategorien auf, sind H - 1 Kontrast-Variablen zu definieren.

Zur Interpretation der Parameter:

Die Konstante β_{g0} beinhaltet den Logarithmus der odds der g-ten Kategorie zur G-ten Kategorie (Referenzkategorie) der response Variablen Y, wenn alle kategorialen X-Variablen die Referenzkategorie und metrische X-Variable den Wert Null annehmen (symbolisiert mit dem Vektor \mathbf{x}_{k-R}):

$$\frac{\pi_g(\mathbf{x}_{k-R})}{\pi_G(\mathbf{x}_{k-R})} = \frac{\frac{e^{\beta_g^T \mathbf{x}_{k-R}}}{1 + \sum_{g=1}^{G-1} e^{\beta_g^T \mathbf{x}_{k-R}}}}{\frac{1}{1 + \sum_{g=1}^{G-1} e^{\beta_g^T \mathbf{x}_{k-R}}}} = e^{\beta_g^T \mathbf{x}_{k-R}} = e^{\beta_{g0}} \quad (4.11)$$

$$g[\pi_g(\mathbf{x}_{k-R})] = \log \left[\frac{\pi_g(\mathbf{x}_{k-R})}{\pi_G(\mathbf{x}_{k-R})} \right] = \beta_{g0}, \quad g = 1, \dots, G - 1.$$

Ein positive (negative) Konstante zeigt an, dass die Zellwahrscheinlichkeit von Y bei der g-ten Kategorie größer (kleiner) ist als bei der G-ten Kategorie.

Wenn die erklärende Variable X_j im Fall einer kategorialen Variablen die Kategorie h statt der Referenzkategorie H annimmt bzw. im Fall einer stetigen Variablen um eine Einheit erhöht wird und alle anderen X-Variablen konstant gehalten werden (symbolisiert mit dem Vektor \mathbf{x}_{k-h}), dann beinhaltet der Parameter β_{gj} den Effekt der Variablen X_j und entspricht dem Logarithmus des odds ratio:

$$\begin{aligned} g[\pi_g(\mathbf{x}_{k-h})] - g[\pi_g(\mathbf{x}_{k-R})] &= \log \left[\frac{\pi_g(\mathbf{x}_{k-h})}{\pi_G(\mathbf{x}_{k-h})} \right] - \log \left[\frac{\pi_g(\mathbf{x}_{k-R})}{\pi_G(\mathbf{x}_{k-R})} \right] \\ &= \log \left[\frac{\frac{\pi_g(\mathbf{x}_{k-h})}{\pi_G(\mathbf{x}_{k-h})}}{\frac{\pi_g(\mathbf{x}_{k-R})}{\pi_G(\mathbf{x}_{k-R})}} \right] = \beta_{gj} \end{aligned} \quad (4.12)$$

Ein positiver (negativer) Parameter besagt, dass beim Übergang von der Referenzkategorie zur h-ten Kategorie einer kategorialen X-Variablen bzw. bei Erhöhung einer stetigen X-Variablen um eine Einheit das Verhältnis der Zellwahrscheinlichkeit der g-ten Kategorie zur G-ten Kategorie von Y zunimmt (abnimmt).

Beispiel 4.1:

Die Problemstellung dieses Beispiels betrifft die Modellierung der Perzeption der gegenwärtigen und künftigen Wirtschaftslage der Bundesrepublik Deutschland in Abhängigkeit von anderen, die befragten Personen charakterisierenden Variablen. Dazu werden Daten des kumulierten ALLBUS ausgewertet.²⁵ Speziell werden die Befragten der Jahre 1991 und 1996 herangezogen. Ab 1991 wurden auch die neuen Bundesländer in die ALLBUS-Stichproben einbezogen. Mit der Analyse über verschiedene Jahre sollen die Veränderungen der Perzeption reflektiert werden.

Im ALLBUS der genannten Jahre wurden bezüglich der Perzeption der Wirtschaftslage u.a. folgende Fragen gestellt:

1. Wie beurteilen Sie ganz allgemein die heutige wirtschaftliche Lage in Deutschland?
2. Wie wird die wirtschaftliche Lage in Deutschland in einem Jahr sein?

Die Inhalte dieser Fragen werden als response Variable Y1 und Y2 definiert. Die beiden Zeitbegriffe in den Fragen “heutige“ und “in einem Jahr“ sind stets zum Zeitpunkt der Befragung zu verstehen.

Die Einschätzung nahmen die Befragten auf einer fünfteiligen Skala vor, und zwar:

²⁵Alle inhaltlichen Ausführungen zum ALLBUS in diesem Beispiel beziehen sich auf: „Allgemeine Bevölkerungsumfrage der Sozialwissenschaften, ALLBUS 1980-94“, Codebuch, ZA-Nr. 1795, herausgegeben vom Zentralarchiv für Empirische Sozialforschung Köln und vom Zentrum für Umfragen, Methoden und Analysen Mannheim (ZUMA). Die vorgenannten Institutionen tragen keine Verantwortung für die Verwendung der Daten in diesem Skript.

für die Frage 1

- 1 - sehr gut
- 2 - gut
- 3 - teils gut / teils schlecht
- 4 - schlecht
- 5 - sehr schlecht

für die Frage 2

- 1 - wesentlich besser als heute
- 2 - etwas besser
- 3 - gleichbleibend
- 4 - etwas schlechter
- 5 - wesentlich schlechter.

Die beiden response Variablen sind somit kategoriale Variable mit $G = 5$ Ausprägungen ($g = 1, \dots, 5$). Das ordinale Skalenniveau der response Variablen soll hier vernachlässigt werden.

Von den Befragten, die in Privathaushalten lebende Personen mit einem Alter von mindestens 18 Jahre zum Befragungszeitpunkt sind, werden nur deutsche Staatsbürger in die Analyse einbezogen.

Die beobachteten Häufigkeitsverteilungen der 2 response Variablen sind in der Tabelle 4.1 und die grafische Darstellung der Häufigkeitsprofile in Abb. 4.1 enthalten.

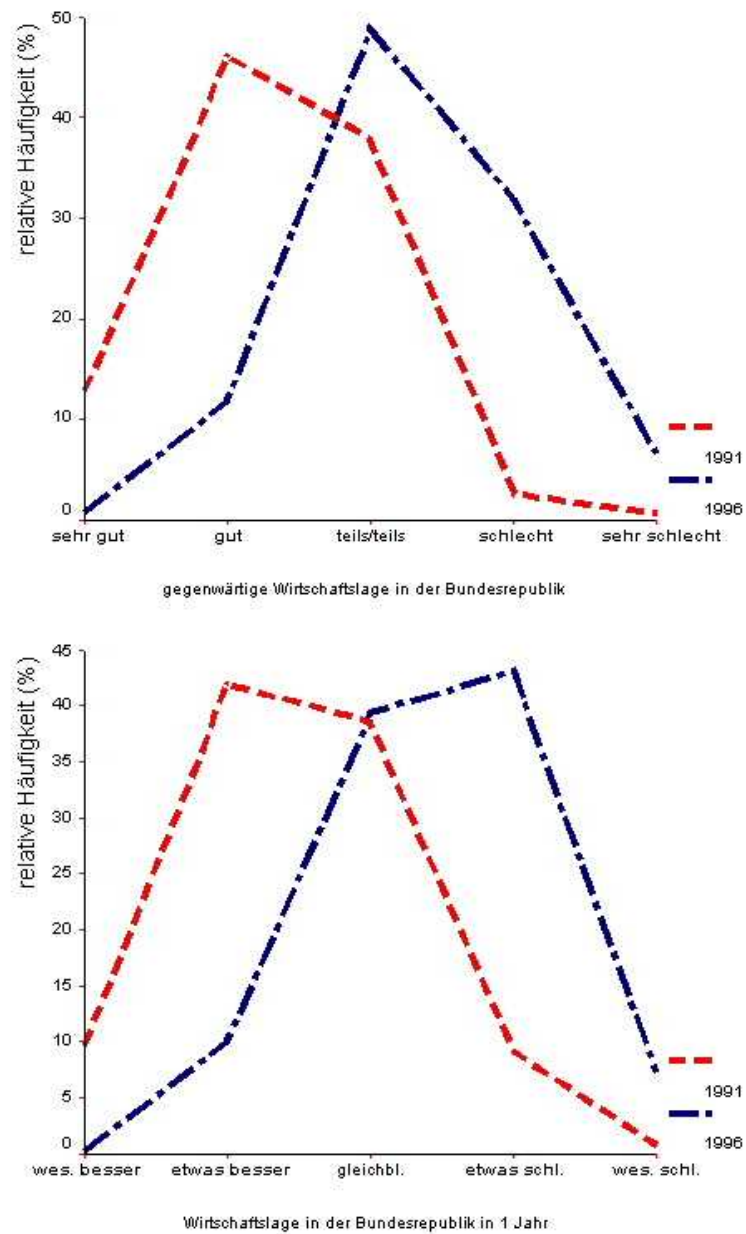
Tabelle 4.1: Häufigkeitsverteilungen der response Variablen; relative Häufigkeiten in %

Y1			Y2		
Kategorien	1991	1996	Kategorien	1991	1996
sehr gut	12,8	0,8	wesentlich besser	9,6	0,3
gut	46,1	11,8	etwas besser	42,0	10,0
teils teils	37,8	48,9	gleichbleibend	38,5	39,3
schlecht	2,6	31,9	etwas schlechter	9,1	43,2
sehr schlecht	0,7	6,6	wesentlich schlechter	0,8	7,3

Aus der Tabelle 4.1 und Abb. 4.1 ist ersichtlich, dass sowohl für die “Wirtschaftslage in der Bundesrepublik Deutschland“ (Y1) als auch für die “künftige Wirtschaftslage der Bundesrepublik“ (Y2) eine deutliche Verschlechterung der Einschätzung 1996 gegenüber 1991 eingetreten ist.

Neben den Fragen zur subjektiven Perzeption der Wirtschaftslage können aus dem Demographiebereich der ALLBUS-Datei objektive Charakteristika der befragten Personen entnommen werden, auf deren Hintergrund die Befragten die Einschätzung vornehmen und die eine Zusammenhangsanalyse zwischen subjektiven und objektiven Indikatoren

Abbildung 4.1: Häufigkeitsprofile je response Variable über die Jahre



ermöglichen. Für dieses Beispiel werden ausgewählt²⁶:

X1 Erhebungsgebiet

1 - alte Bundesländer (West) 2 - neue Bundesländer (Ost), Referenzkategorie

X2 Geschlecht

1 - Mann 2 - Frau (Referenzkategorie)

Für die weitere Analyse wird stets die letzte (5.) Kategorie der response Variablen als Referenzkategorie gewählt. Für diese Kategorie wird $\beta_5 = \mathbf{0}$ gesetzt, so dass $\eta_{5k} = \beta_5^T \mathbf{x}_k = 0$ und $\exp(\beta_5^T \mathbf{x}_k) = 1$ sind. Der Vergleich jeder anderen Bewertungskategorie wird somit stets zu “sehr schlecht“ für Y1 bzw. “wesentlich schlechter“ für Y2 geführt, d.h., die odds beinhalten $\pi_g(\mathbf{x}_k)/\pi_5(\mathbf{x}_k)$. Für die anderen Kategorien der response Variablen sind die unbekannten Parameter β_{gj} ($g = 1, \dots, 4; j = 1, \dots, J$) zu schätzen. Sie beinhalten den Effekt der jeweiligen erklärenden Variablen X_j auf die Wahrscheinlichkeit des Eintretens einer Bewertungskategorie der response Variablen Y1 bzw. Y2.

Die simultane Schätzung der Parameter wird mit der Prozedur “Regression Multinomial Logistic“ von SPSS unter Windows Release 10.0.7²⁷ durchgeführt. Diese Prozedur erfordert die Einzelauflistung aller Fälle.

Geprüft wird auf einem Signifikanzniveau von $\alpha = 0,05$.

Schätzung mit der erklärenden Variablen X1

Zunächst sollen Kontingenztafeln einen Überblick über die Beobachtungen geben, wobei neben der absoluten Häufigkeit die bedingte relative Häufigkeit, gegeben eine Ausprägung von X_1 , in den Tabellen 4.2 und 4.3 enthalten ist.

Wird nur die dichotome Variable X1 (Erhebungsgebiet) und die Indikator-Kodierung mit der letzten Kategorie als Bezugskategorie verwendet, ergibt sich eine Kontrast-Variable $\mathbf{x}^T = (x_0, x_1) = (1, 1)$ für die Befragten aus den alten Bundesländern, wobei vereinba-

²⁶Für eine ausführliche Analyse der Perzeption der Wirtschaftslage mit weiteren erklärenden X-Variablen siehe Rönz, B. (1999), Modelling the perception of current and prospective economic situation, Statistics Research Report No. 99.002, Centre for Mathematics and its Applications, School of Mathematical Sciences, The Australian National University, Canberra

²⁷Siehe SPSS Regression Models 9.0. Eine simultane Schätzung der Parameter ist auch mit der Prozedur “Regression Logit“ von SYSTAT Version 7.0.1. möglich.

rungsgemäß $x_0 = 1$ für die Konstante des Modells steht.

Tabelle 4.2: Kontingenztabelle Y1 - X1 für 1991 und 1996

			Erhebungsgebiet 1991		Total	Erhebungsgebiet 1996		Total
			West	Ost		West	Ost	
Wirt- schafts- lage	sehr gut	Count	213	167	380	20	6	26
		% within X1	14,5%	11,1%	12,8%	,9%	,5%	,8%
	gut	Count	760	606	1366	268	119	387
		% within X1	51,9%	40,4%	46,1%	12,4%	10,7%	11,8%
	teils teils	Count	449	671	1120	1028	571	1599
		% within X1	30,7%	44,7%	37,8%	47,5%	51,5%	48,9%
	schlecht	Count	37	41	78	704	338	1042
		% within X1	2,5%	2,7%	2,6%	32,5%	30,5%	31,9%
	sehr schlecht	Count	5	16	21	143	74	217
		% within X1	,3%	1,1%	,7%	6,6%	6,7%	6,6%
Total	Count	1464	1501	2965	2163	1108	3271	
	% within X1	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	

Tabelle 4.3: Kontingenztabelle Y2 - X1 für 1991 und 1996

			Erhebungsgebiet 1991		Total	Erhebungsgebiet 1996		Total
			West	Ost		West	Ost	
Wirt- schafts- lage in 1 Jahr	wesentlich	Count	75	205	280	4	5	9
	besser	% within X1	5,3%	13,8%	9,6%	,2%	,5%	,3%
	etwas besser	Count	450	771	1221	190	131	321
		% within X1	31,6%	52,0%	42,0%	8,9%	12,0%	10,0%
	gleichbleibend	Count	684	435	1119	815	447	1262
		% within X1	48,0%	29,4%	38,5%	38,4%	41,0%	39,3%
	etwas schlechter	Count	200	65	265	961	427	1388
		% within X1	14,0%	4,4%	9,1%	45,2%	39,2%	43,2%
wesentlich schlechter	Count	16	6	22	155	80	235	
	% within X1	1,1%	,4%	,8%	7,3%	7,3%	7,3%	
Total		Count	1425	1482	2907	2125	1090	3215
		% within X1	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

Als link Funktion erhält man:

$$g[\pi_g(\mathbf{x})] = \eta_g = \log \left[\frac{\pi_g(\mathbf{x}_k)}{\pi_G(\mathbf{x}_k)} \right] = \beta_g^T \mathbf{x} = \beta_{g0} + \beta_{g1}x_1, \quad g = 1, \dots, 4 \quad (4.4)$$

Ausführlich notiert ergibt sich z.B. für die response Variable Y1:

$$g[\pi_{sehr\ gut}(\mathbf{x})] = \log \left[\frac{\pi_{sehr\ gut}(\mathbf{x})}{\pi_{sehr\ schlecht}(\mathbf{x})} \right] = \beta_{10} + \beta_{11}x_1$$

$$\begin{aligned}
g[\pi_{gut}(\mathbf{x})] &= \log \left[\frac{\pi_{gut}(\mathbf{x})}{\pi_{sehr\ schlech}(\mathbf{x})} \right] = \beta_{20} + \beta_{21}x_1 \\
g[\pi_{teils\ teils}(\mathbf{x})] &= \log \left[\frac{\pi_{teils\ teils}(\mathbf{x})}{\pi_{sehr\ schlech}(\mathbf{x})} \right] = \beta_{30} + \beta_{31}x_1 \\
g[\pi_{schlecht}(\mathbf{x})] &= \log \left[\frac{\pi_{schlecht}(\mathbf{x})}{\pi_{sehr\ schlech}(\mathbf{x})} \right] = \beta_{40} + \beta_{41}x_1.
\end{aligned}$$

Für diese response Variable Y1 des Jahres 1991 sollen wichtige Informationen des SPSS-Output gezeigt werden.

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	121,077			
Final	47,997	73,08	4	,000

In dieser Tabelle steht unter Chi-Square der Wert der Differenz der Deviancen des Modells mit nur einer Konstanten und des Modells mit der erklärenden Variablen X1, d.h. ΔD gemäß (3.37): $\Delta D = 73,08$. Die Anzahl der Freiheitsgrade für ΔD gemäß (3.38) ist $f = 4$, da gegenüber dem Modell mit nur einer Konstanten vier weitere Parameter b_{g1} ($g = 1, \dots, 4$) simultan zu schätzen sind. Der Wert von ΔD wird mit dem Quantil der Chi-Quadrat-Verteilung für das vorgegebene Signifikanzniveau $\alpha = 0,05$ verglichen. Aus der Tabelle der Chi-Quadrat-Verteilung findet man für $f = 4$ Freiheitsgrade $\chi_{0,05;4}^2 = 9,49$. H_0 , dass alle Parameter für X1 gleich Null sind, wird somit auf dem 5%-Niveau abgelehnt, da $\Delta D > \chi_{0,05;4}^2$ ist. Zu diesem Ergebnis gelangt man auch durch den Vergleich des vorgegebenen Signifikanzniveaus $\alpha = 0,05$ mit der Ausgabe Sig. in der obigen Tabelle: Da $\text{Sig.} = ,000 < \alpha = 0,05$ ist, wird H_0 abgelehnt. Damit ist das Modell mit der Variablen X1 signifikant besser als das Modell, das nur die Konstante enthält.

Pseudo R-Square

Cox and Snell	,024
Nagelkerke	,027
McFadden	,011

Cox & Snell entspricht der Formel (3.39) und Nagelkerke der Formel (3.40). McFadden's R^2 ist definiert als:

$$R_{MF}^2 = \frac{l_0 - l_1}{l_0} \quad (4.13)$$

worin l_0 der Kern²⁸ der log-Likelihood-Funktion für das Logit Modell mit nur einer Konstanten und l_1 der Kern der log-Likelihood-Funktion für das interessierende Logit Modell mit einer bestimmten Anzahl erklärenden X-Variablen sind. Alle Pseudo R-Square zeigen an, dass der Anteil der erklärten Variation der response Variablen durch das multinomiale Logit Modell nur sehr gering ist.

Parameter Estimates

Wirtschafts- lage heute		B	Std. Error	Wald	df	Sig.	Exp(B)
sehr gut	Intercept	2,345	,262	80,320	110	,000	
	[X1_91=1]	1,406	,523	7,241		,007	4,081
	[X1_91=2]	0 ^a	,	,		,	,
gut	Intercept	3,634	,253	205,893	110	,000	
	[X1_91=1]	1,390	,515	7,274		,007	4,013
	[X1_91=2]	0 ^a	,	,		,	,
teils teils	Intercept	3,736	,253	218,143	110	,000	
	[X1_91=1]	,761	,516	2,178		,140	2,141
	[X1_91=2]	0 ^a	,	,		,	,
schlecht	Intercept	,941	,295	10,190	110	,001	
	[X1_91=1]	1,060	,560	3,583		,058	2,888
	[X1_91=2]	0 ^a	,	,		,	,

a. This parameter ist set zero because it is redundant.

Die Konstanten b_{g0} sind die log odds (logits) für die Befragten aus den neuen Bundesländern (Ost). So beinhaltet z.B. b_{10} den Logarithmus des Verhältnisses der Wahrscheinlichkeit, dass ein Befragter aus den neuen Bundesländern die heutige (im Jahre 1991) Wirtschaftslage der Bundesrepublik Deutschland mit sehr gut einschätzt, zur Wahrscheinlichkeit, dass ein Befragter aus den neuen Bundesländern die heutige (im Jahre 1991) Wirtschaftslage der Bundesrepublik Deutschland mit sehr schlecht einschätzt:

$$g[\hat{\pi}_1(\mathbf{x}_{Ost})] = \log \left[\frac{\hat{\pi}_{sehr\ gut}(\mathbf{x}_{Ost})}{\hat{\pi}_{sehr\ schlecht}(\mathbf{x}_{Ost})} \right] = b_{10} = 2,345$$

²⁸Die log-Likelihood-Funktion kann angegeben werden als die Summe einer multinomialen Konstanten, die nicht von den Parametern abhängt, und einem Kern, der von den Parametern abhängt. Der Wert -2log-likelihood unter SPSS enthält beides, die Konstante und den Kern (vgl. SPSS Regression Models 9.0, S. 70)

Dies lässt sich für dieses einfache Modell auch aus der Kontingenztabelle 4.2 entnehmen: $\log(167/16) = 2,345$. Der Wald-Test zeigt an, dass alle Konstanten signifikant zum 5% Niveau sind.

Die positiven Konstanten beinhalten, dass für die Befragten aus den neuen Bundesländern die Wahrscheinlichkeit einer optimistischeren Einschätzung der Wirtschaftslage (Kategorien 1, ..., 4 von Y1) größer als die Wahrscheinlichkeit für eine sehr schlechte Bewertung ist. Nach (4.2) ergibt sich z.B.:

$$\begin{aligned}\hat{\pi}_{sehr\ gut}(\mathbf{x}_{Ost}) &= \frac{e^{b_{10}}}{1 + \sum_{g=1}^{G-1} e^{b_{g0}}} = \frac{e^{2,345}}{1 + e^{2,345} + e^{3,634} + e^{3,736} + e^{0,941}} = \frac{10,4333}{93,7897} \\ &= 0,1112 \\ \hat{\pi}_{sehr\ schlecht}(\mathbf{x}_{Ost}) &= \frac{e^{b_{10}}}{1 + \sum_{g=1}^{G-1} e^{b_{g0}}} = \frac{1}{1 + e^{2,345} + e^{3,634} + e^{3,736} + e^{0,941}} = \frac{1}{93,7897} \\ &= 0,01066.\end{aligned}$$

Die Parameter b_{g1} für die Befragten aus den alten Bundesländern beinhalten eine Aussage über die Beziehung zwischen den logits und der Variablen X1 (Erhebungsgebiet) und entsprechen den Logarithmen der odds ratios. Die Parameter b_{11} und b_{21} (für sehr gute bzw. gute Einschätzung) sind zum 5% Niveau signifikant, b_{31} (für die Einschätzung teils teils) ist nicht signifikant und b_{41} (für die Bewertung schlecht) ist zum 10% Niveau signifikant (siehe die Ergebnisse des Wald-Tests).

Für die geschätzten link Funktionen resultiert:

$$g(\hat{\pi}_{sehr\ gut}(\mathbf{x})) = 2,345 + 1,406(x_1 = 1) = 3,751$$

$$g(\hat{\pi}_{gut}(\mathbf{x})) = 3,634 + 1,390(x_1 = 1) = 5,024$$

$$g(\hat{\pi}_{teils\ teils}(\mathbf{x})) = 3,736 + 0,761(x_1 = 1) = 4,497$$

$$g(\hat{\pi}_{schlecht}(\mathbf{x})) = 0,941 + 1,060(x_1 = 1) = 2,001.$$

Wählt man z.B. die 1. Kategorie von Y1, so ergibt sich nach (4.12)

$$g[\hat{\pi}_{sehr\ gut}(\mathbf{x}_{West})] - g[\hat{\pi}_{sehr\ gut}(\mathbf{x}_{Ost})] = \log \left[\frac{\frac{\hat{\pi}_{sehr\ gut}(\mathbf{x}_{West})}{\hat{\pi}_{sehr\ schlecht}(\mathbf{x}_{West})}}{\frac{\hat{\pi}_{sehr\ gut}(\mathbf{x}_{Ost})}{\hat{\pi}_{sehr\ schlecht}(\mathbf{x}_{Ost})}} \right] = b_{11},$$

und somit $b_{11} = 3,751 - 2,345 = 1,406$.

Auch dieser geschätzte Parameter b_{11} lässt sich für dieses einfache Modell aus der Kontingenztafel 4.2 entnehmen:

$$\log \left[\frac{\frac{213}{5}}{\frac{167}{16}} \right] = \log \left[\frac{42,6}{10,4375} \right] = \log(4,08144) = 1,406449 = b_{11}.$$

Für das odds ratio erhält man:

$$\exp(b_{11}) = \exp(1,406449) = 4,081,$$

was auch in der Tabelle “Parameter Estimates “ des SPSS-Output unter Exp(B) zu finden ist.

Da der geschätzte Parameter b_{11} positiv ist, beinhaltet er, dass Befragte aus den alten Bundesländern (West) wahrscheinlicher (und zwar aufgrund der Stichprobe des Jahres 1991 ca. viermal wahrscheinlicher) sind als Befragte aus den neuen Bundesländern (Ost), die die heutige (im Jahre 1991) Wirtschaftslage der Bundesrepublik Deutschland mit sehr gut anstatt mit sehr schlecht einzuschätzen.

Für die Befragten aus den alten Bundesländern ergibt sich die Wahrscheinlichkeit für die 1. Kategorie nach (4.2) zu:

$$\begin{aligned} \pi_{sehr\ gut}(\mathbf{x}_{West}) &= \frac{e^{\boldsymbol{\beta}_g^T \mathbf{x}_{West}}}{1 + \sum_{g=1}^{G-1} e^{\boldsymbol{\beta}_g^T \mathbf{x}_{West}}} = \frac{e^{3,751}}{1 + e^{3,751} + e^{5,024} + e^{4,497} + e^{2,001}} \\ &= \frac{42,5636}{292,7257} = 0,145. \end{aligned}$$

Der paarweise Vergleich zwischen anderen Kategorien r und s, die nicht Referenzkategorie sind, kann gemäß (4.5) geführt werden. Soll z.B. der Vergleich zwischen den Kategorien “gut“ und “teils teils“ für die Befragten aus den alten Bundesländern erfolgen, so ergibt sich für die log odds (logit):

$$\begin{aligned} \log \left[\frac{\hat{\pi}_{gut}(\mathbf{x}_{West})}{\hat{\pi}_{teils\ teils}(\mathbf{x}_{West})} \right] &= g[\hat{\pi}_{gut}(\mathbf{x}_{West})] - g[\hat{\pi}_{teils\ teils}(\mathbf{x}_{West})] = 5,024 - 4,497 \\ &= 0,527 \end{aligned}$$

und für die odds:

$$\frac{\hat{\pi}_{gut}(\mathbf{x}_{West})}{\hat{\pi}_{teils\ teils}(\mathbf{x}_{West})} = \frac{e^{\mathbf{b}_{gut}^T \mathbf{x}_{West}}}{e^{\mathbf{b}_{teils\ teils}^T \mathbf{x}_{West}}} = \frac{e^{5,024}}{e^{4,497}} = \frac{152,0182}{89,7475} = 1,6938.$$

Für die Befragten aus den alten Bundesländern ist eine gute Einschätzung der heutigen (im Jahre 1991) Wirtschaftslage ca. 1,7 mal wahrscheinlicher als die Bewertung teils teils.

Eine zusammenfassende Übersicht über die Schätzungen für die response Variablen Y1 und Y2 der Jahre 1991 und 1996 gibt die Tabelle 4.4.

Tabelle 4.4: Schätzungen für die Einschätzung der Wirtschaftslage und Erhebungsgebiet

(X1)				
Parameter	Y1		Y2	
	1991	1996	1991	1996
Kategorie 1				
b_{10}	2,345	-2,512	3,531	-2,773
b_{11}	1,406	0,545	-1,986	-0,885
Kategorie 2				
b_{20}	3,634	0,475	4,856	0,493
b_{21}	1,390	0,153	-1,519	-0,290
Kategorie 3				
b_{30}	3,736	2,043	4,284	1,721
b_{31}	0,761	-0,071	-0,528	-0,061
Kategorie 4				
b_{40}	0,941	1,519	2,383	1,675
b_{41}	<u>1,061</u>	0,075	0,143	0,150

Die fett gedruckten Schätzwerte sind zum 5% Niveau und die unterstrichenen Werte zum 10% Niveau signifikant.

Bezüglich der gegenwärtigen Wirtschaftslage in Deutschland (Y1) verdeutlichen die 1996 gegenüber 1991 stark gefallen Konstanten der Kategorien 1 bis 3 und die gestiegene Konstante der Kategorie 4 die deutliche Verschiebung zu einer pessimistischeren Einschätzung der Befragten der neuen Bundesländer. Fallende Konstanten über die Zeit implizieren auch eine Verschlechterung in der Bewertung der Befragten der alten Bundesländer, da das Sinken der Konstanten nicht durch hohe positive Parameter b_{g1} kompensiert wird.

Während (wie bereits ausgeführt) im Jahre 1991 signifikante Parameter b_{11} und b_{21} anzeigen, dass die Befragten in den alten Bundesländern die gegenwärtige Wirtschaftslage in Deutschland wesentlich zufriedener bewerteten als die Befragten in den neuen Bundesländern, hat 1996 die Variable Erhebungsgebiet X1 keinen wesentlichen Einfluss auf die Bewertung der gegenwärtigen Wirtschaftslage (nicht signifikante Parameter b_{g1}), d.h., die odds bei den Befragten aus den alten und den neuen Bundesländern differieren nicht bemerkenswert.

Hinsichtlich der kurzfristigen künftigen Wirtschaftslage in der Bundesrepublik Deutschland (Y2) blicken die Befragten der neuen Bundesländer 1991 deutlich optimistischer in die Zukunft als die Befragten der alten Bundesländer (signifikante negative Parameter b_{11} und b_{21} der Kategorien 1 und 2 von Y2), während es 1996 wiederum zu einer Annäherung der Einschätzungen kommt (kleine negative und nicht signifikante Parameter b_{g1}). Insgesamt sind zwischen den alten und neuen Bundesländern 1996 noch Unterschiede in der Perzeption der gegenwärtigen und künftigen Wirtschaftslage der Bundesrepublik festzustellen, aber die Tendenz der Annäherung ist unübersehbar.

Schätzung mit der erklärenden Variablen X2

Da das Geschlecht (X2) ebenfalls eine dichotome Variable ist, ergibt sich bei Verwendung der Indikator-Kodierung mit der letzten Kategorie als Bezugskategorie wiederum nur eine Kontrast-Variable $\mathbf{x}^T = (x_0, x_2) = (1, 1)$ für die männlichen Befragten, da die Kategorie Frau von X2 als Referenzkategorie gewählt wurde. Als link Funktion resultiert nach (4.4):

$$g[\pi_g(\mathbf{x})] = \eta_g = \log \left[\frac{\pi_g(\mathbf{x})}{\pi_G(\mathbf{x})} \right] = \boldsymbol{\beta}_g^T \mathbf{x} = \beta_{0g} + \beta_{g2}x_2, \quad g = 1, \dots, 4.$$

Eine zusammenfassende Übersicht über die Schätzungen ist in Tabelle 4.5 enthalten. Für die gegenwärtige und künftige Wirtschaftslage in Deutschland (Y1, Y2) ist anhand der Konstanten b_{g0} eine drastische Verschlechterung der Einschätzung bei den Frauen 1996 gegenüber 1991 ablesbar. Da die Konstanten auch in der link Funktion für die Männer enthalten sind und ihre sinkende Tendenz nicht durch hohe positive Parameter b_{g2} kompensiert werden, ist tendenziell auch eine Verschlechterung in den Einschätzungen der männlichen Befragten festzustellen.

Ein wesentlicher Einfluss der Variablen Geschlecht (X2) ist nur im Jahre 1991 auf die Einschätzung der gegenwärtigen Wirtschaftslage der Bundesrepublik Deutschland festzustellen.

Tabelle 4.5: Schätzungen für die Einschätzung der Wirtschaftslage und Geschlecht (X2)

Parameter	Y1		Y2	
	1991	1996	1991	1996
Kategorie 1				
b_{10}	2,468	-2,312	2,355	-3,689
b_{12}	1,053	0,356	0,411	0,736
Kategorie 2				
b_{20}	3,771	0,444	3,850	0,088
b_{22}	1,010	0,259	0,366	0,414
Kategorie 3				
b_{30}	3,784	2,033	3,842	1,715
b_{32}	0,555	-0,075	0,201	-0,072
Kategorie 4				
b_{40}	1,142	1,560	2,342	1,781
b_{42}	0,500	0,019	0,209	-0,009

Die fett gedruckten Schätzwerte sind zum 5% Niveau signifikant.

Da die zugehörigen geschätzten Parameter b_{12} und b_{22} positiv und signifikant zum 5% Niveau sind, lassen sie die Schlussfolgerungen zu, dass männliche Befragte wahrscheinlicher als weibliche Befragte sind, die die gegenwärtige Wirtschaftslage der Bundesrepublik mit sehr gut bzw. gut anstatt mit sehr schlecht einschätzen: Männer sind rund 2,9 mal wahrscheinlicher als Frauen ($\exp(b_{12}) = 2,867$), eine sehr gute statt einer sehr schlechten Einschätzung abzugeben, und Männer sind rund 2,7 mal wahrscheinlicher als Frauen ($\exp(b_{12}) = 2,746$), eine gute statt einer sehr schlechten Einschätzung abzugeben. Im Jahre 1991 bewerteten die Männer die gegenwärtige Wirtschaftslage somit tendenziell optimistischer als die Frauen. In allen anderen Fällen kann die Nullhypothese, dass die Parameter gleich Null sind, nicht abgelehnt werden, d.h., es konnte kein wesentlicher Unterschied in den Einschätzungen von Männer und Frauen beobachtet werden.

Schätzung mit den erklärenden Variablen X1 und X2

Die link Funktion für das multiple Logit Modell mit den beiden erklärenden Variablen Erhebungsgebiet (X1) und Geschlecht (X2) ist gemäß (4.4):

$$g[\pi_g(\mathbf{x})] = \eta_g = \log \left[\frac{\pi_g(\mathbf{x})}{\pi_G(\mathbf{x})} \right] = \beta_g^T \mathbf{x} = \beta_{0g} + \beta_{g1}x_1 + \beta_{g2}x_2, \quad g = 1, \dots, 4.$$

mit $x_1 = 1$ für die Befragten aus den alten Bundesländern und $x_2 = 1$ für die männlichen Befragten.

Beide X-Variablen zusammen haben 1991 und 1996 einen wesentlichen Einfluss auf die Einschätzung der jeweiligen Wirtschaftslage. Bezüglich des separaten Einflusses ist dies jedoch differenzierter.

- Jede X-Variable übt einen signifikanten Einfluss (5% Niveau) auf die Einschätzung der gegenwärtigen Wirtschaftslage (Y1) im Jahre 1991 aus.

Likelihood ratio Test

Effect	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept	88,533	0,000	0	,
X1_91	160,798	72,265	4	,000
X2_91	129,546	41,013	4	,000

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced Model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

- Der Einfluss des Erhebungsgebietes (X1) auf die Einschätzung der künftigen Wirtschaftslage (Y2) im Jahre 1991 ist signifikant zum 5% Niveau, der Einfluss des Geschlechts (X2) nur zum 10% Niveau.

Likelihood ratio Test

Effect	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept	94,682	,000	0	,
X1_91	376,971	282,289	4	,000
X2_91	103,132	8,450	4	,076

- Der Einfluss des Erhebungsgebietes (X1) auf die Einschätzung der künftigen Wirtschaftslage (Y2) im Jahre 1996 ist nicht signifikant zum 5% Niveau, der Einfluss des Geschlechts (X2) nur zum 10% Niveau.

Likelihood ratio Test

Effect	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept	90,736	,000	0	,
X1_96	97,071	6,334	4	,176
X2_96	100,175	9,439	4	,051

- Beide X-Variablen üben einen signifikanten (5% Niveau) separaten Einfluss auf die Einschätzung der künftigen Wirtschaftslage (Y2) im Jahre 1996 aus.

Likelihood ratio Test

Effect	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept	93,946	,000	0	,
X1_96	110,213	16,266	4	,003
X2_96	110,584	16,637	4	,002

Eine zusammenfassende Übersicht über die Schätzungen ist in Tabelle 4.6 enthalten.

Tabelle 4.6: Schätzungen für die Einschätzung der Wirtschaftslage in Abhängigkeit vom Erhebungsgebiet (X1) und Geschlecht (X2)

Parameter	Y1		Y2	
	1991	1996	1991	1996
Kategorie 1				
b_{10}	1,916	-2,696	3,318	-3,201
b_{11}	1,408	0,541	-2,005	-0,900
b_{12}	1,055	0,351	0,483	0,754
Kategorie 2				
b_{20}	3,228	0,343	4,673	<u>0,271</u>
b_{21}	1,391	0,150	-1,536	<u>-0,298</u>
b_{22}	1,012	0,257	0,423	0,420
Kategorie 3				
b_{30}	3,543	2,078	4,194	1,753
b_{31}	0,762	-0,070	-0,537	-0,059
b_{32}	0,556	-0,075	0,220	-0,071
Kategorie 4				
b_{40}	0,770	1,510	2,299	1,680
b_{41}	<u>1,061</u>	0,076	0,135	0,150
b_{42}	0,502	0,018	0,205	-0,012

Die fett gedruckten Schätzwerte sind zum 5% Niveau und die unterstrichenen Werte zum 10% Niveau signifikant.

Die Interpretation der geschätzten Parameter ist analog zu denen der einfachen Modelle, jedoch nunmehr unter der Bedingung, dass die andere Variable konstant gehalten wird. Mittels der geschätzten logits können anschließend die Wahrscheinlichkeiten und die Residuen für die Kategorien der response Variablen durch Einsetzen der verschiedenen Werte der X-Variablen geschätzt werden. Auf die Einzeldarstellung soll hier verzichtet werden.

Im Vergleich der Tabellen 4.4 bis 4.6 ist ersichtlich, dass sich die Parameterschätzungen in den multiplen Modellen gegenüber den einfachen Modellen nicht deutlich verändert haben. Gravierende Veränderungen der Parameterschätzungen zwischen den einfachen Modellen und dem multiplen Modell können ein Hinweis auf Interaktionen zwischen den X-Variablen sein. Interaktionseffekte würden bedeuten, dass der Effekt der Variablen Geschlecht (X2) nicht gleich ist für die Befragten aus den alten bzw. neuen Bundesländern. Um dies zu prüfen, müsste die Interaktion zwischen X1 und X2 in das Modell eingeschlossen werden (full factorial model).

5 Modellierung multinomialer Daten

(log-lineare Modelle)

Bei der Untersuchung von Zusammenhängen multinomialer Variablen (nominalskalierter, ordinalskalierter Variablen sowie metrisch gruppierter Variablen) wird ihre gemeinsame Wahrscheinlichkeitsverteilung herangezogen. Die gemeinsamen Wahrscheinlichkeiten der Grundgesamtheit sind jedoch im allgemeinen unbekannt. Ziel der Modellierung ist es, diese Wahrscheinlichkeiten zu schätzen und gleichzeitig Hypothesen über Beziehungen zwischen den Variablen zu prüfen.²⁹ Bei der Hypothesenprüfung steht die symmetrische Fragestellung im Vordergrund, d.h., es werden Effekte der Variablen auf die gemeinsamen Wahrscheinlichkeiten (response) geschätzt, so dass nicht zwischen abhängigen und erklärenden Variablen unterschieden wird.³⁰

Für die weitere Betrachtung sei zunächst davon ausgegangen, dass zwei Variablen in die Analyse einbezogen werden.

²⁹Siehe u.a. Dobson, A.J. (1991), Kap. 9; Fahrmeir, L., Hamerle, A. (1984), Kap. 10; Christensen, R. (1990); Santner, Th.J., Duffy, D.E. (1989); Bishop, Y.M.M., Fienberg, S.E., Holland, P.W. (1975); Langeheine, R. (1980).

³⁰Beim multinomialen Logit Modell ist eine der betrachteten Variablen die response Variable, und es werden die Effekte erklärender Variablen auf die Wahrscheinlichkeit des Eintretens der verschiedenen Kategorien der response Variablen geschätzt.

5.1 Die zweidimensionale Kontingenztafel

Es sei allgemein angenommen, dass eine Variable A mit J Variablenausprägungen ($j = 1, \dots, J$) und eine Variable B mit K Variablenausprägungen ($k = 1, \dots, K$) gegeben sind und diese Variablen an n statistischen Einheiten ($i = 1, \dots, n$) beobachtet wurden. Es gibt $J \cdot K$ mögliche Paare von Variablenausprägungen dieser beiden Variablen (A_j, B_k). Eine geeignete Darstellungsform der gemeinsamen Häufigkeitsverteilung (sample joint frequency distribution) ist eine zweidimensionale Kontingenztafel, deren allgemeiner Aufbau die Tabelle 5.1 zeigt.

Die Zeilen der Kontingenztafel entsprechen den Variablenausprägungen der Variablen A und die Spalten den Variablenausprägungen der Variablen B. Innerhalb der Kontingenztafel gibt es $J \cdot K$ Zellen.

Tabelle 5.1: Zweidimensionale Kontingenztafel

Variable A	Variable B					Randverteilung von A
	B_1	\dots	B_k	\dots	B_K	
A_1	y_{11}	\dots	y_{1k}	\dots	y_{1K}	y_{1+}
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
A_j	y_{j1}	\dots	y_{jk}	\dots	y_{jK}	y_{j+}
\vdots	\vdots	\dots	\vdots	\dots	\vdots	\vdots
A_J	y_{J1}	\dots	y_{Jk}	\dots	y_{JK}	y_{J+}
Randverteilung von B	y_{+1}	\dots	y_{+k}	\dots	y_{+K}	$y_{++} = n$

Die Zellen nehmen die absoluten oder relativen Häufigkeiten des Auftretens eines Paares von Variablenausprägungen (A_j, B_k) auf. In der Tabelle 5.1 wurden die absoluten Häufigkeiten angegeben: $y(A_j, B_k) = y_{jk}$ ($j = 1, \dots, J$; $k = 1, \dots, K$). Sie beinhaltet die Anzahl der Elemente in der Stichprobe mit der Variablenausprägung A_j der Variablen A und der Variablenausprägung B_k der Variablen B (absolute Zellhäufigkeiten). Die letzte Spalte gibt die Randverteilung der Variablen A und die letzte Zeile die Randverteilung der Variablen B an (sample marginal frequency distribution). Dabei beinhaltet:

$$y_{j+} = \sum_{k=1}^K y_{jk}; \quad j = 1, \dots, J \quad (5.1)$$

die Anzahl der Elemente in der Stichprobe mit der Variablenausprägung A_j der Variablen A (row total proportions) und

$$y_{+k} = \sum_{j=1}^J y_{jk}; \quad k = 1, \dots, K \quad (5.2)$$

die Anzahl der Elemente in der Stichprobe mit der Variablenausprägung B_k der Variablen B (column total proportions).

$$y_{++} = \sum_{j=1}^J \sum_{k=1}^K y_{jk} = \sum_{j=1}^J y_{j+} = \sum_{k=1}^K y_{+k} = n \quad (5.3)$$

ist die Gesamtzahl der beobachteten Elemente (Stichprobenumfang).

Analog kann die Kontingenztabelle mit den relativen Häufigkeiten $f(A_j, B_k) = p_{jk}$ angegeben werden:

$$\begin{aligned} f(A_j, B_k) &= p_{jk} = \frac{y_{jk}}{n} \\ f(A_j) &= p_{j+} = \frac{y_{j+}}{n} \\ f(B_k) &= p_{+k} = \frac{y_{+k}}{n} \\ p_{++} &= \sum_{j=1}^J \sum_{k=1}^K p_{jk} = \sum_{j=1}^J p_{j+} = \sum_{k=1}^K p_{+k} = 1. \end{aligned} \quad (5.4)$$

Auf Basis dieser Kontingenztabelle können bestimmt werden:

- bedingte relative Häufigkeitsverteilungen

- die bedingte relative Häufigkeitsverteilung der Variablen A für eine gegebene Ausprägung B_k der Variablen B (conditional frequency distribution for the rows given column k, column proportions):

$$f(A_j|B_k) = \frac{f(\{A = A_j\} \cap \{B = B_k\})}{f(B = B_k)} = \frac{y_{jk}}{y_{+k}}, \quad j = 1, \dots, J. \quad (5.5)$$

- die bedingte relative Häufigkeitsverteilung der Variablen B für eine gegebene Ausprägung A_j der Variablen A (conditional frequency distribution for the columns given row j, row proportions):

$$f(B_k|A_j) = \frac{f(\{A = A_j\} \cap \{B = B_k\})}{f(A = A_j)} = \frac{y_{jk}}{y_{j+}}, \quad k = 1, \dots, K. \quad (5.6)$$

- erwartete absolute Häufigkeit (expected frequencies)

Es ist diejenige absolute Häufigkeit in jeder Zelle der Kontingenztabelle, die unter der Annahme der stochastischen Unabhängigkeit der beiden Variablen A und B zu erwarten ist. Sind die Variablen A und B stochastisch unabhängig, dann gilt nach dem Multiplikationssatz für unabhängige Ereignisse:

$$P(\{A = A_j\} \cap \{B = B_K\}) = P(A = A_j) \cdot P(B = B_K) = \pi_{j+} \cdot \pi_{+k} = \pi_{jk}. \quad (5.7)$$

Die erwarteten absoluten Häufigkeiten (m_{jk}) ergeben sich zu

$$m_{jk} = n \cdot \pi_{j+} \cdot \pi_{+k} = n \cdot \pi_{jk}. \quad (5.8)$$

Da die Randwahrscheinlichkeiten π_{j+} und π_{+k} der Grundgesamtheit unbekannt sind, werden sie mittels der beobachteten Randhäufigkeiten geschätzt:

$$f(A_j) = p_{j+} = \frac{y_{j+}}{n}; \quad f(B_k) = p_{+k} = \frac{y_{+k}}{n}. \quad (5.9)$$

Für die geschätzten erwarteten absoluten Häufigkeiten folgt dann:

$$\hat{m}_{jk} = n \cdot p_{j+} \cdot p_{+k} = \frac{y_{j+} y_{+k}}{n}. \quad (5.10)$$

- Residuen

Ein Residuum beinhaltet die Abweichungen zwischen der beobachteten und der unter Unabhängigkeit erwarteten Häufigkeit einer Zelle. Dabei sind drei Arten von Residuen zu unterscheiden:

- unstandardisiertes Residuum

Es ist die Abweichung zwischen der beobachteten absoluten Häufigkeit und der bei Unabhängigkeit erwarteten absoluten Häufigkeit einer Zelle:

$$r_{jk} = y_{jk} - \hat{m}_{jk}. \quad (5.11)$$

- standardisiertes Residuum

Das standardisierte Residuum einer Zelle ergibt sich als unstandardisiertes Residuum dividiert durch die positive Quadratwurzel aus der erwarteten Häufigkeit:

$$rs_{jk} = \frac{y_{jk} - \hat{m}_{jk}}{\sqrt{\hat{m}_{jk}}} = \frac{r_{jk}}{\sqrt{\hat{m}_{jk}}} \quad (5.12)$$

- korrigiertes standardisiertes Residuum

Das korrigierte standardisierte (adjusted) Residuum einer Zelle beinhaltet das unstandardisierte Residuum dividiert durch eine Schätzung des Standardfehlers. Die Schätzung des Standardfehlers erfolgt dabei unter Berücksichtigung der Zeilenhäufigkeit y_{j+} , der Spaltenhäufigkeit y_{+k} sowie der Gesamtzahl n der Beobachtungen.

$$ra_{jk} = \frac{y_{jk} - \hat{m}_{jk}}{\sqrt{\hat{m}_{jk} \cdot \left(1 - \frac{y_{j+}}{n}\right) \left(1 - \frac{y_{+k}}{n}\right)}} = \frac{r_{jk}}{\sqrt{\hat{m}_{jk} \cdot \left(1 - \frac{y_{j+}}{n}\right) \left(1 - \frac{y_{+k}}{n}\right)}} \quad (5.13)$$

5.2 Stichprobenmodelle für multinomiale Daten

Zur Erzeugung einer $J \times K$ Kontingenztabelle mit n Beobachtungen steht eine Vielzahl von Stichprobenmodellen (Stichprobenerhebungskonzepte, sampling models) zur Verfügung. Die am häufigsten verwendeten werden nachfolgend erläutert.

5.2.1 Multinomiales Stichprobenmodell

Die Multinomialverteilung³¹ ergibt sich als Verallgemeinerung der Binomialverteilung, indem die Beschränkung auf zwei mögliche Realisationen bei der Durchführung des Zufallsexperiments aufgehoben wird.

Eine Grundgesamtheit weise eine Variable A mit J ($J \geq 2$) unterschiedlichen, sich gegenseitig ausschließenden Kategorien A_j ($j = 1, \dots, J$) auf. Die Wahrscheinlichkeit, dass ein Element der Grundgesamtheit die Kategorie j besetzt, sei π_j , wobei $\sum_j \pi_j = 1$ gilt.

³¹Siehe u.a. Dobson, A.J. (1991), S. 126 ff.; Fahrmeir, L., Hamerle, A. (1984), S. 33 ff.; Christensen, R. (1990), S. 14 ff.; Santner, Th.J., Duffy, D.E. (1989), S. 16 ff.

Ein Zufallsexperiment bestehe aus n unabhängigen Versuchen (Ziehen mit Zurücklegen bzw. (unendlich) große Grundgesamtheit), d.h., es wird eine einfache Zufallsstichprobe vom Umfang n gezogen. Bei jedem Versuch wird festgestellt, welche Kategorie der Variablen das gezogene Stichprobenelement aufweist, wobei vorausgesetzt wird, dass bei jedem Versuch eine der J Kategorien auftritt. π_j ist dann die Wahrscheinlichkeit des Auftretens der j -ten Kategorie bei jedem Versuch. Die Anzahl der Stichprobenelemente mit der Kategorie A_j bei den n Versuchen sei mit Y_j ($j = 1, \dots, J$) bezeichnet, wobei es sich vor der Durchführung der Versuche um Zufallsvariablen handelt. Die Wahrscheinlichkeit, dass in einer Stichprobe von n unabhängigen Beobachtungen genau y_1 Elemente zur Kategorie A_1 , y_2 Elemente zur Kategorie A_2 , \dots und y_J Elemente zur Kategorie A_J gehören ist

$$\begin{aligned} P(Y_1 = y_1, \dots, Y_J = y_J | n; \pi_1, \dots, \pi_J) &= f(y_1, \dots, y_J | n; \pi_1, \dots, \pi_J) \\ &= \frac{n!}{y_1! \cdot y_2! \cdot \dots \cdot y_J!} \cdot \pi_1^{y_1} \cdot \pi_2^{y_2} \cdot \dots \cdot \pi_J^{y_J} \\ &= \frac{n!}{\prod_{j=1}^J y_j!} \cdot \prod_{j=1}^J \pi_j^{y_j} \end{aligned} \quad (5.14)$$

für $y_j = 0, 1, \dots, n$; $j = 1, \dots, J$; $\sum_j y_j = n$; $\sum_j \pi_j = 1$ und $P(Y_1 = y_1, \dots, Y_J = y_J | n; \pi_1, \dots, \pi_J) = f(y_1, \dots, y_J | n; \pi_1, \dots, \pi_J) = 0$ sonst.

Die Verteilung (5.14) des Zufallsfaktors $\mathbf{Y}^T = (Y_1, \dots, Y_J)$ heisst Multinomialverteilung (oder Polymialverteilung) mit den Parametern $n; \pi_1, \dots, \pi_J$, symbolisiert mit $\mathbf{Y} \sim M(n; \boldsymbol{\pi})$, wobei $\boldsymbol{\pi}^T = (\pi_1 \dots \pi_J)$ ist.

$n!/(y_1! \cdot y_2! \cdot \dots \cdot y_J!)$ heisst Multinomialkoeffizient und gibt die Anzahl der unterschiedlichen Möglichkeiten an, n Elemente auf J Gruppen zu verteilen. Dies ist gleich der Anzahl der Permutationen (Anordnung aller n Elemente) mit Wiederholung.

Für den Erwartungswert der Multinomialverteilung ergibt sich:

$E(Y_j) = n\pi_j$, so dass

$$E(\mathbf{Y}) = (n\pi_1, \dots, n\pi_J)^T = n\boldsymbol{\pi},$$

für die Varianz

$$\text{Var}(Y_j) = \sigma_j^2 = n\pi_j(1 - \pi_j),$$

für die Kovarianz

$$\text{Cov}(Y_j, Y_{j*}) = \sigma_{jj*} = -n\pi_j\pi_{j*} \quad (j \neq j*)$$

und für die Varianz-Kovarianz-Matrix von \mathbf{Y}

$$\mathbf{\Sigma}(\mathbf{Y}) = n(\text{diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}^T).$$

Für die Varianz-Kovarianz-Matrix von \mathbf{Y} erhält man z.B. mit $J = 3$:

$$\begin{aligned} \mathbf{\Sigma}(\mathbf{Y}) &= n \left[\begin{pmatrix} \pi_1 & 0 & 0 \\ 0 & \pi_2 & 0 \\ 0 & 0 & \pi_3 \end{pmatrix} - \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{pmatrix} \begin{pmatrix} \pi_1 & \pi_2 & \pi_3 \end{pmatrix} \right] \\ &= n \left[\begin{pmatrix} \pi_1 & 0 & 0 \\ 0 & \pi_2 & 0 \\ 0 & 0 & \pi_3 \end{pmatrix} - \begin{pmatrix} \pi_1^2 & \pi_1\pi_2 & \pi_1\pi_3 \\ \pi_2\pi_1 & \pi_2^2 & \pi_2\pi_3 \\ \pi_3\pi_1 & \pi_3\pi_2 & \pi_3^2 \end{pmatrix} \right] \\ &= n \begin{pmatrix} \pi_1(1 - \pi_1) & -\pi_1\pi_2 & -\pi_1\pi_3 \\ -\pi_2\pi_1 & \pi_2(1 - \pi_2) & -\pi_2\pi_3 \\ -\pi_3\pi_1 & -\pi_3\pi_2 & \pi_3(1 - \pi_3) \end{pmatrix} \end{aligned}$$

Bei der Durchführung des Zufallsexperiments wird nun neben der Variablen A zusätzlich eine zweite Variable B mit K Kategorien B_k ($k = 1, \dots, K$) beobachtet. Die beiden Variablen werden dabei gemeinsam beobachtet.

Der Grundgesamtheit wird eine Zufallsstichprobe entnommen, wobei der Stichprobenumfang n fest vorgegeben wird, d.h., es gilt die Bedingung:

$$n = y_{++} = \sum_{j=1}^J \sum_{k=1}^K y_{jk}. \quad (5.15)$$

Damit sind in der Kontingenztafel (siehe Tabelle 5.1) alle Zellhäufigkeiten Y_{jk} und alle Randhäufigkeiten Y_{j+} und Y_{+k} ($j = 1, \dots, J$; $k = 1, \dots, K$) Zufallsvariable. Nur die rechte untere Ecke der Kontingenztafel mit $y_{++} = n$ ist keine Zufallsvariable. Die

Zellwahrscheinlichkeiten sind $P(A_j \cap B_k) = \pi_{jk}$ und die Randwahrscheinlichkeiten:

$$P(A_j) = \pi_{j+} = \sum_{k=1}^K \pi_{jk}; \quad j = 1, \dots, J \quad (5.16)$$

$$P(B_k) = \pi_{+k} = \sum_{j=1}^J \pi_{jk}; \quad k = 1, \dots, K$$

Es gelten die Bedingungen:

$$\sum_{j=1}^J \sum_{k=1}^K \pi_{jk} = 1; \quad \sum_{j=1}^J \pi_{j+} = 1; \quad \sum_{k=1}^K \pi_{+k} = 1. \quad (5.17)$$

Der Zufallsvektor $\mathbf{Y}^T = (Y_{11}, \dots, Y_{JK})$ ist multinomialverteilt mit den Parametern n und $\pi_{11}, \dots, \pi_{JK}$, d.h., die gemeinsame Verteilung der Y_{jk} ist gegeben mit:

$$f(\mathbf{y}|n, \boldsymbol{\pi}) = \frac{n!}{\prod_{j=1}^J \prod_{k=1}^K y_{jk}!} \cdot \prod_{j=1}^J \prod_{k=1}^K \pi_{jk}^{y_{jk}} \quad (5.18)$$

mit $\mathbf{y}^T = (y_{11}, \dots, y_{JK})$ und $\boldsymbol{\pi}^T = (\pi_{11}, \dots, \pi_{JK})$ bzw. in der Kurz-Notation $\mathbf{Y} \sim M(n; \boldsymbol{\pi})$. Unter Verwendung einer fortlaufenden Numerierung der Zellen mit $q = 1, \dots, Q$ und $Q = J \cdot K$ erhält man:

$$f(\mathbf{y}|n, \boldsymbol{\pi}) = \frac{n!}{\prod_{q=1}^Q y_q!} \cdot \prod_{q=1}^Q \pi_q^{y_q}. \quad (5.19)$$

Erwartungswerte, Varianzen und Kovarianzen der einzelnen Komponenten von \mathbf{y} sind gegeben mit:

$$E(Y_{jk}) = n\pi_{jk}, \quad Var(Y_{jk}) = n\pi_{jk}(1 - \pi_{jk}) \text{ für } j = 1, \dots, J \text{ und } k = 1, \dots, K;$$

$$Cov(Y_{jk}, Y_{j^*k^*}) = -n\pi_{jk}\pi_{j^*k^*} \text{ für } j \neq j^*, k \neq k^*, j, j^* = 1, \dots, J; k, k^* = 1, \dots, K.$$

Die Maximum-Likelihood-Schätzer der Zellwahrscheinlichkeiten π_{jk} sind die korrespondierenden relativen Stichprobenhäufigkeiten $p_{jk} = y_{jk}/n$.

Eine wertvolle Eigenschaft der Multinomialverteilung ist, dass die Summe von multinomialverteilten Zufallsvariablen ebenfalls multinomialverteilt ist. Die Parameter werden analog aufsummiert, um die Parameter der Verteilung der Summe zu erhalten. Daraus

folgt:

Die Zeilen- bzw. Spaltensummen sind ebenfalls multinomialverteilt, d.h. $(Y_{1+}, \dots, Y_{J+}) \sim M(n; \pi_{1+}, \dots, \pi_{j+})$ und $(Y_{+1}, \dots, Y_{+k}) \sim M(n; \pi_{+1}, \dots, \pi_{+k})$ gemäß (5.14), wenn man die entsprechenden y und π einsetzt.

Das multinomiale Stichprobenmodell ist somit durch folgende Eigenschaften charakterisiert:

1. Der Grundgesamtheit wird eine Zufallsstichprobe mit fest vorgegebenem Stichprobenumfang entnommen.
2. Alle Zellhäufigkeiten Y_{jk} und alle Randhäufigkeiten Y_{j+} und Y_{+k} ($j = 1, \dots, j$; $k = 1, \dots, K$) sind Zufallsvariable.
3. Die Zufallsvektoren $\mathbf{Y}^T = (Y_{11}, \dots, Y_{JK})$, (Y_{1+}, \dots, Y_{j+}) und (Y_{+1}, \dots, Y_{+K}) sind jeweils multinomialverteilt.

5.2.2 Produkt-multinomiales Stichprobenmodell

Bei diesem Stichprobenmodell wird eine der beiden Variablen als ein Faktor mit verschiedenen Faktorstufen aufgefasst, so dass entsprechend der Anzahl der Faktorstufen verschiedene, sich gegenseitig ausschließende Segmente der Grundgesamtheiten gegeben sind. Aus jedem Segment wird unabhängig voneinander eine Stichprobe mit festgelegtem Umfang gezogen und die Anzahl des Auftretens der Kategorien der zweiten Variablen beobachtet. Es variiert somit nur noch eine Variable, da je Segment die Ausprägung (Kategorie, Faktorstufe) der anderen Variablen feststeht.

Als Konsequenz für die Kontingenztafel ergibt sich, dass die Zeilensummen $y_{j+} = n_j$ bzw. die Spaltensummen $y_{+k} = n_k$ gegeben sind. Der Gesamtstichprobenumfang n wird nur indirekt über deren Summe ermittelt. Die weiteren Ausführungen beziehen sich auf feste Zeilensummen, d.h., die Variable A ist die Faktorvariable. Analoge Überlegungen gelten bei festgelegten Spaltensummen.

Es bezeichne $\pi_{jk|j+}$ die bedingte Wahrscheinlichkeit in der Grundgesamtheit:

$\pi_{jk|j+} = P(\text{Element weist Kategorie } k \text{ der Variablen B auf} \mid \text{Element entstammt dem})$

Segment mit der Faktorstufe j der Variablen A) $= \pi_{jk}/\pi_{j+}$.

Dann ist die Verteilung jeder Zeile j der Kontingenztabelle eine Multinomialverteilung mit den Parametern y_{j+} und $\pi_{j1|j+}, \dots, \pi_{jK|j+}$:

$(Y_{j1}, \dots, Y_{jK})^T \sim M(y_{j+}; \pi_{j1|j+}, \dots, \pi_{jK|j+})$ bzw. gemäß (5.14):

$$\begin{aligned} f(y_{j1}, \dots, y_{jK}|y_{j+}) &= \frac{y_{j+}!}{y_{j1}! \cdot \dots \cdot y_{jK}!} \cdot \pi_{j1|j+}^{y_{j1}} \cdot \dots \cdot \pi_{jK|j+}^{y_{jK}} \\ &= \frac{y_{j+}!}{\prod_{k=1}^K y_{jk}!} \cdot \prod_{k=1}^K \pi_{jk|j+}^{y_{jk}} \end{aligned} \quad (5.20)$$

für $y_{jk} = 0, 1, \dots, y_{j+}$; $\sum_k y_{jk} = y_{j+}$; $\sum_k \pi_{jk|j+} = 1$, $j = 1, \dots, J$.

Für die gemeinsame Verteilung aller Y_{jk} folgt:

$$\begin{aligned} f(y_{j1}, \dots, y_{jK}|y_{1+}, \dots, y_{J+}) &= \prod_{j=1}^J f(y_{j1}, \dots, y_{jK}|y_{j+}) \\ &= \prod_{j=1}^J \frac{y_{j+}!}{\prod_{k=1}^K y_{jk}!} \cdot \prod_{k=1}^K \pi_{jk|j+}^{y_{jk}}. \end{aligned} \quad (5.21)$$

Die Verteilung (5.21) wird als Produkt-Multinomialverteilung bezeichnet, da die gemeinsame Verteilung aller J Zeilen der Kontingenztabelle das Produkt der individuellen Zeilen-Verteilungen ist.

Erwartungswerte, Varianzen und Kovarianzen können für jede der J Multinomialverteilungen berechnet werden, wobei die anderen Zeilen (d.h. Multinomialverteilungen) ignoriert werden. Kovarianzen zwischen Zufallsvariablen verschiedener Zeilen (d.h. Multinomialverteilungen) sind wegen der Annahme der Unabhängigkeit der Stichproben aus den Segmenten der Grundgesamtheit gleich Null.

5.2.3 Poisson-Stichprobenmodell

Beim multinomialen Stichprobenmodell ist der Stichprobenumfang n fest vorgegeben.

Beim Poisson-Stichprobenmodell ist der Stichprobenumfang n dagegen nicht festgelegt.

Die Erhebung erfolgt in der Weise, dass innerhalb einer festgelegten Zeitspanne (Kontinuum) die aufgetretenen Ereignisse entsprechend ihrer Zugehörigkeit zu den Kategorienkombinationen erfasst werden; n ist damit offen. Jede Häufigkeit in der Kontingenztafel, einschließlich Y_{++} , ist eine Zufallsvariable. Es wird vorausgesetzt, dass die Häufigkeiten jeder Kategorienkombination (Zelle der Kontingenztafel) Realisationen von unabhängigen Poisson-Prozessen sind:

- das Eintreffen der Ereignisse soll unabhängig sein, d.h., die Wahrscheinlichkeit, dass in einem festen Intervall Ereignisse auftreten, soll unabhängig davon sein, was sich vor Beginn dieses Intervalls abgespielt hat;
- die Intensität des Stromes der eintretenden Ereignisse soll konstant sein, d.h., verschiebt man die feste Länge des Intervalls, so soll die mittlere Anzahl λ der in das Intervall fallenden Ereignisse unabhängig von der Länge und Lage des Intervalls sein (Stationarität);
- in einem beliebig kleinen Intervall soll die Wahrscheinlichkeit, dass mehr als ein Ereignis eintritt, gleich Null sein.

Unter dieser Voraussetzung folgt jedes Y_{jk} ($j = 1, \dots, J; k = 1, \dots, K$) einer Poisson-Verteilung mit dem Parameter $\lambda_{jk} = E(Y_{jk})$, d.h. $Y_{jk} \sim P(\lambda_{jk})$:

$$f(y_{jk}|\lambda_{jk}) = \frac{\lambda_{jk}^{y_{jk}}}{y_{jk}!} \cdot e^{-\lambda_{jk}}, \quad y_{jk} = 0, 1, \dots \quad (5.22)$$

Wegen $E(Y_{jk}) = \lambda_{jk}$ gibt dieser Parameter die im Mittel erwartete Häufigkeit je Zelle an. $Var(Y_{jk})$ ist ebenfalls gleich λ_{jk} . Aufgrund der Reproduktivitätseigenschaft der Poisson-Verteilung gilt für die Summe der Y_{jk} :

$$\sum_j \sum_k Y_{jk} = Y_{++} \sim P(\lambda_{++})$$

d.h., der Stichprobenumfang Y_{++} ist ebenfalls poisson-verteilt mit dem Erwartungswert $\lambda_{++} = \sum_j \sum_k \lambda_{jk}$. Analoges gilt für die Randhäufigkeiten.

Die gemeinsame Verteilung aller Y_{jk} ergibt sich als das Produkt der einzelnen Verteilungen zu

$$f(y_{11}, \dots, y_{JK} | \lambda_{11}, \dots, \lambda_{JK}) = \prod_{j=1}^J \prod_{k=1}^K \frac{\lambda_{jk}^{y_{jk}}}{y_{jk}!} \cdot e^{-\lambda_{jk}}. \quad (5.23)$$

5.3 Hypothesenformulierung

Für das **multinomiale Stichprobenmodell** ist von Interesse, ob die beiden Variablen die Wahrscheinlichkeiten π_{jk} der Grundgesamtheiten unabhängig beeinflussen oder ob auch Wechselwirkungen (Interaktionen) zwischen ihnen die Struktur der Grundgesamtheit bestimmen. Die Hypothese lautet somit auf Unabhängigkeit der Variablen. Gemäß dem Unabhängigkeitssatz der Wahrscheinlichkeitstheorie $P(A_j \cap B_k) = P(A_j)P(B_k)$ lautet die Hypothese:

$$H_0 : \pi_{jk} = \pi_{j+} \cdot \pi_{+k} \quad \text{für alle } j \text{ und } k, \quad (5.24)$$

wobei π_{j+} und π_{+k} die marginalen Wahrscheinlichkeiten der Zeilen- bzw. Spaltenvariablen sind und $\sum_j \pi_{j+} = \sum_k \pi_{+k} = 1$ gilt. Für gegebenen Stichprobenumfang n sind die erwarteten Zellhäufigkeiten (im weiteren mit m_{jk} symbolisiert):

$$m_{jk} = E(Y_{jk}) = n\pi_{jk}.$$

Bei Gültigkeit der H_0 folgt:

$$m_{jk}^0 = n\pi_{j+}\pi_{+k},$$

wobei die hochgestellte 0 die Gültigkeit von H_0 symbolisiert.

Da die Randverteilungen ebenfalls multinomialverteilt sind mit

$$m_{j+} = E(Y_{j+}) = n\pi_{j+} \quad \text{bzw.} \quad m_{+k} = E(Y_{+k}) = n\pi_{+k},$$

lässt sich die Nullhypothese (5.24) äquivalent als

$$H_0 : m_{jk}^0 = m_{j+}m_{+k}/n \quad \text{für alle } j \text{ und } k, \quad (5.25)$$

schreiben. Verwendet man den ML-Schätzer von π_{j+}

$$p_{j+} = y_{j+}/n$$

und analog als eine Schätzung von π_{+k}

$$p_{+k} = y_{+k}/n$$

so folgt als Schätzung für m_{jk}^0 :

$$\hat{m}_{jk}^0 = y_{j+}y_{+k}/n. \quad (5.26)$$

Die geschätzten erwarteten Zellhäufigkeiten \hat{m}_{jk}^0 erfüllen die Bedingungen, dass

- die Randverteilungen der beiden Variablen gleich den beobachteten Häufigkeiten sind und
- $\sum_j \sum_k \hat{m}_{jk}^0 = (1/n) \sum_j \sum_k y_{j+}y_{+k} = (1/n) \sum_j y_{j+} \cdot \sum_k y_{+k} = (1/n)n \cdot n = n$ ist.

Die geschätzten erwarteten Zellhäufigkeiten \hat{m}_{jk}^0 werden im allgemeinen von den beobachteten Zellhäufigkeiten abweichen.

Beim **produkt-multinomialen Stichprobenmodell** wird vor allem die Homogenitätshypothese geprüft. In einer zweidimensionalen Kontingenztafel mit festen Zeilensummen y_{j+} sind äquivalente Formulierungen der Nullhypothese:

- Die Faktorausprägungen der Variablen A haben keinen Einfluss auf die Zellwahrscheinlichkeit der Variablen B.
- Die Multinomialverteilungen der J Segmente der Variablen A der Grundgesamtheit sind identisch.
- Die Zellwahrscheinlichkeiten sind in jeder Spalte k gleich.

Zur Vereinfachung der Notation wird im weiteren $\pi_{jk|j+} = \pi_{jk}$ gesetzt, wobei man sich jedoch stets in Erinnerung rufen muss, dass die inhaltliche Bedeutung jetzt eine andere ist: $\pi_{jk} = P(B_k|A_j)$ ist beim produkt-multinomialen Stichprobenmodell eine bedingte Wahrscheinlichkeit und es gilt $\sum_k \pi_{jk} = 1$, während beim multinomialen Stichprobenmodell $\pi_{jk} = P(A_j \cap B_k)$ und $\sum_j \sum_k \pi_{jk} = 1$ ist.

Die Formulierung der Nullhypothese beim produkt-multinomialen Stichprobenmodell lautet:

$H_0 : \pi_{1k} = \pi_{2k} = \dots = \pi_{Jk}$ für alle k (d.h. für jede Spalte der Kontingenztafel). (5.27)

Jede Zeile der Kontingenztafel folgt (wie bereits gezeigt) einer Multinomialverteilung mit dem Erwartungswert

$$m_{jk} = E(Y_{jk}) = y_{j+}\pi_{jk}.$$

Bei Gültigkeit von H_0 sind die π_{jk} für alle $j = 1, \dots, J$ gleich. Wegen

$$m_{j+} = \sum_k m_{jk} = \sum_k y_{j+}\pi_{jk} = y_{j+} \sum_k \pi_{jk} = y_{j+}$$

und

$$m_{+k} = \sum_j m_{jk} = \sum_j y_{j+}\pi_{jk} = \pi_{jk} \sum_j y_{j+} = \pi_{jk} n$$
 ³²

kann H_0 auch äquivalent mit den erwarteten Zellhäufigkeiten geschrieben werden:

$$H_0 : m_{jk}^0 = m_{j+}(m_{+k}/n). \quad (5.28)$$

Eine Schätzung des gemeinsamen Wertes der (Spalten-)Wahrscheinlichkeit π_{jk} (pooled estimate) ergibt sich unter H_0 zu

$$p_{jk}^0 = y_{+k}/n,$$

so dass sich eine Schätzung der erwarteten Zellhäufigkeit zu

$$\hat{m}_{jk}^0 = y_{j+}(y_{+k}/n) \quad (5.29)$$

ergibt, wobei y_{j+} durch das Stichprobenmodell festgelegt ist und der Gesamtstichprobenumfang mit in die Schätzung eingeht.

Auch hier gilt es, die Abweichungen zwischen den geschätzten erwarteten Zellhäufigkeiten \hat{m}_{jk}^0 und den beobachteten Zellhäufigkeiten mittels einer geeigneten Teststatistik zu überprüfen.

³² $\sum_j y_{j+}\pi_{jk} = \pi_{jk} \sum_j y_{j+}$ gilt, weil unter H_0 die π_{jk} für alle j gleich und somit eine Konstante unabhängig vom Laufindex j sind.

5.4 Das log-lineare Modell

Mit den formulierten Hypothesen wurde unterstellt, dass die beiden Variablen A und B bestimmte Effekte auf die Zellhäufigkeiten aufweisen, die nunmehr geschätzt werden sollen. Im Abschnitt 5.3 wurde mit (5.25) und (5.28) gezeigt, dass sich für das multinomiale und das produkt-multinomiale Stichprobenmodell³³ die Nullhypothese als ein multiplikatives Modell

$$m_{jk} = m_{j+}m_{+k}/m_{++} \quad (5.30)$$

mit bestimmten Nebenbedingungen formulieren ließ.

Logarithmieren ergibt:

$$\log m_{jk} = \log m_{j+} + \log m_{+k} - \log m_{++} \quad (5.31)$$

d.h., bei Gültigkeit der Nullhypothese lässt sich der Logarithmus der erwarteten Häufigkeit einer Zelle darstellen als die Summe dreier Terme, wobei der erste nur von der Zeile, der zweite nur von der Spalte und der dritte nur von der Gesamtbesetzung abhängt.

Die allgemeine Notation des Modells wird jedoch nicht so gewählt, wie sie mit (5.31) gegeben ist. Man wählt vielmehr eine Parametrisierung, die zu guten Schätzeigenschaften führt und vor allem eine einfache Verallgemeinerung auf höherdimensionale Probleme (mit mehr als zwei Variablen) ermöglicht, indem man die Form

$$g(\pi_{jk}) = \eta_{jk} = \log m_{jk} = \beta + \beta_{A(j)} + \beta_{B(k)} \quad (5.32)$$

wählt. (5.32) wird als log-lineares Modell bezeichnet, dass sich bei Gültigkeit der Nullhypothese ergibt.

Im log-linearen Modell bei Gültigkeit von H_0 ist die link Funktion η_{jk} eine Linearkombination von Parametern: einer Konstanten β , eines Haupteffektes der Variablen A, enthalten im Parameter $\beta_{A(j)}$, und eines Haupteffektes der Variablen B, enthalten im Parameter $\beta_{B(k)}$.

Das log-lineare Modell hat jedoch zu viele Parameter, so dass zur Erlangung einer eindeutigen Lösung Restriktionen über die Parameter eingeführt werden, die denen der

³³Für die weiteren Betrachtungen erfolgt eine Beschränkung auf diese beiden Stichprobenmodelle.

Varianzanalyse gleichen. Eine Parametrisierung kann in der Form der sum-to-zero Restriktion, die durch die Deviation- oder Effekt-Kodierung realisiert wird, oder in Form der corner-point Restriktion, die durch die Indikator- oder Dummy-Kodierung realisiert wird, erfolgen.

Wählt man die sum-to-zero Restriktion, so muss bei der Wahl der letzten Kategorie als Referenzkategorie gelten:

$$\sum_j \beta_{A(j)} = \sum_k \beta_{B(k)} = 0. \quad (5.33)$$

Wegen dieser Restriktion muss für die Parameter $\beta_{A(J)}$ und $\beta_{B(K)}$ gelten:

$$\beta_{A(j)} = -\sum_{j=1}^{J-1} \beta_{A(j)}, \quad \beta_{B(k)} = -\sum_{k=1}^{K-1} \beta_{B(k)}. \quad (5.34)$$

Für die Parameter des log-linearen Modells bei Gültigkeit der Nullhypothese H_0 erhält man durch Summenbildung von (5.32) unter Berücksichtigung von (5.33):

$$\begin{aligned} \sum_{j=1}^J \sum_{k=1}^K \log m_{jk} &= \sum_{j=1}^J \sum_{k=1}^K \beta + \sum_{j=1}^J \sum_{k=1}^K \beta_{A(j)} + \sum_{j=1}^J \sum_{k=1}^K \beta_{B(k)} \\ &= JK\beta + K \sum_{j=1}^J \beta_{A(j)} + J \sum_{k=1}^K \beta_{B(k)} = JK\beta, \\ \beta &= \frac{1}{JK} \sum_{j=1}^J \sum_{k=1}^K \log m_{jk}. \end{aligned} \quad (5.35)$$

Die Konstante β ist das arithmetische Mittel der Logarithmen aller erwarteten Zellhäufigkeiten.

$$\begin{aligned} \sum_{k=1}^K \log m_{jk} &= \sum_{k=1}^K \beta + \sum_{k=1}^K \beta_{A(j)} + \sum_{k=1}^K \beta_{B(k)} \\ &= K\beta + K\beta_{A(j)}, \\ \beta_{A(j)} &= \frac{1}{K} \sum_{k=1}^K \log m_{jk} - \beta. \end{aligned} \quad (5.36)$$

Die Parameter $\beta_{A(j)}$ beinhalten die Differenz zwischen dem Mittelwert der Logarithmen der erwarteten Zellhäufigkeiten der Zeile j und dem Gesamtmittelwert.

$$\begin{aligned}\sum_{j=1}^J \log m_{jk} &= \sum_{j=1}^J \beta + \sum_{j=1}^J \beta_{A(j)} + \sum_{j=1}^J \beta_{B(k)} \\ &= J\beta + J\beta_{B(k)}, \\ \beta_{B(k)} &= \frac{1}{J} \sum_{j=1}^J \log m_{jk} - \beta.\end{aligned}\tag{5.37}$$

Die Parameter $\beta_{B(k)}$ beinhalten die Differenz zwischen dem Mittelwert der Logarithmen der erwarteten Zellhäufigkeiten der Spalte k und dem Gesamtmittelwert.

Da die erwarteten Zellhäufigkeiten m_{jk} unbekannt sind, müssen sie aus den Daten geschätzt werden.

Aufgrund der unterschiedlichen Stichprobenmodelle, die mit bestimmten Nebenbedingungen verbunden sind, müssen diese Nebenbedingungen auch beim log-linearen Modell berücksichtigt werden.³⁴

1. Für das multinomiale Stichprobenmodell ist die Nullhypothese (5.30) äquivalent zum log-linearen Modell (5.32) zusammen mit (5.33) und der zusätzlichen Nebenbedingung

$$y_{++} = \sum_{j=1}^J \sum_{k=1}^K e^{\beta + \beta_{A(j)} + \beta_{B(k)}}.\tag{5.38}$$

Diese zusätzliche Nebenbedingung sichert die Forderung $m_{++} = y_{++} = n$, d.h., die Summe der erwarteten Zellhäufigkeiten ist identisch mit dem vorgegebenen Stichprobenumfang.

2. Für das produkt-multinomiale Stichprobenmodell ist die Hypothese (5.30) äquivalent zum log-linearen Modell (5.32) zusammen mit (5.33) und der zusätzlichen Nebenbedingung

$$y_{j+} = \sum_{k=1}^K e^{\beta + \beta_{A(j)} + \beta_{B(k)}} = e^{\beta} \cdot e^{\beta_{A(j)}} \sum_{k=1}^K e^{\beta_{B(k)}}, \quad j = 1, \dots, J.\tag{5.39}$$

³⁴Vgl. Fahrmeir, L., Hamerle, A. (1984), S. 480 f.

5 Modellierung multinomialer Daten (log-lineare Modelle)

Diese zusätzliche Nebenbedingung sichert die Forderung $m_{j+} = y_{j+}$ für alle j , d.h., die erwarteten Zeilenrandsummen sind identisch mit den vorgegebenen Umfängen der Stichproben aus den Segmenten der Grundgesamtheit.

Wenn das log-lineare Modell gültig ist, dann ergibt sich aus (5.32):

$$m_{jk} = e^{\beta + \beta_{A(j)} + \beta_{B(k)}} \quad (5.40)$$

Wegen

$$m_{j+} = \sum_{k=1}^K m_{jk} = \sum_{k=1}^K e^{\beta} e^{\beta_{A(j)}} e^{\beta_{B(k)}} = e^{\beta} e^{\beta_{A(j)}} \sum_{k=1}^K e^{\beta_{B(k)}} \quad (5.41)$$

ist beim produkt-multinomialen Stichprobenmodell die Forderung $m_{j+} = y_{j+}$ erfüllt, denn m_{j+} aus (5.41) ist gleich y_{j+} aus (5.39).

Weiterhin ist:

$$m_{j+} = \sum_{k=1}^J m_{jk} = \sum_{k=1}^J e^{\beta + \beta_{A(j)} + \beta_{B(k)}} = e^{\beta} e^{\beta_{B(k)}} \sum_{k=1}^J e^{\beta_{A(j)}}, \quad (5.42)$$

$$m_{++} = \sum_{j=1}^J \sum_{k=1}^K m_{jk} = \sum_{j=1}^J \sum_{k=1}^K e^{\beta} e^{\beta_{A(j)}} e^{\beta_{B(k)}}. \quad (5.43)$$

Mit (5.43) ist die Forderung $m_{++} = y_{++} = n$ beim multinomialen Stichprobenmodell erfüllt, denn m_{++} aus (5.43) ist gleich y_{++} aus (5.38).

Ein und derselbe Datensatz kann somit unter verschiedenen Gesichtspunkten betrachtet werden, je nachdem welche Hypothese geprüft werden soll bzw. welches log-lineare Modell verwendet wird.

Wenn die Struktur der Daten auch durch Wechselwirkungen zwischen den beiden Variablen bestimmt wird, muss dies im log-linearen Modell berücksichtigt werden. Die Aufnahme von Parametern $\beta_{AB(jk)}$, die solche Interaktionseffekte beinhalten, führt zum vollständigen oder saturierten log-linearen Modell:

$$g(\pi_{jk}) = \eta_{jk} = \log m_{jk} = \beta + \beta_{A(j)} + \beta_{B(k)} + \beta_{AB(jk)}, \quad (5.44)$$

hinter dem die triviale Hypothese steht, dass sich die beobachteten Zellhäufigkeiten durch sich selbst erklären. Im log-linearen Modell gehören wegen der symmetrischen

Fragestellung sowohl zur Variablen A als auch zur Variablen B Haupteffekte $\beta_{A(j)}$ bzw. $\beta_{B(k)}$ und das Verschwinden der Interaktionseffekte $\beta_{AB(jk)}$ bedeutet Unabhängigkeit bzw. Homogenität zwischen A und B.

Die Restriktionen (5.33) müssen für das saturierte Modell erweitert werden, so dass

$$\sum_j \beta_{A(j)} = \sum_k \beta_{B(k)} = 0, \quad \sum_j \beta_{AB(jk)} = \sum_k \beta_{AB(jk)} = 0 \text{ für alle } j, k \quad (5.45)$$

gilt. Entsprechend dem Stichprobenmodell sind die o.g. zusätzlichen Nebenbedingungen erforderlich.

Da die Parameter β , $\beta_{A(j)}$ und $\beta_{B(k)}$ gemäß (5.35) bis (5.37) bestimmt werden, ergeben sich die Parameter $\beta_{AB(jk)}$ als Differenz:

$$\beta_{AB(jk)} = \log m_{jk} - (\beta + \beta_{A(j)} + \beta_{B(k)}). \quad (5.46)$$

Die Restriktionen (5.45) reduzieren die Anzahl der unabhängigen Parameter des Modells:

- Für den Parameter β ist in (5.45) keine Restriktion angegeben, er ist frei wählbar, also ist die Anzahl der unabhängigen Parameter gleich 1.
- Die Werte der Parameter $\beta_{A(j)}$ sind für jede der J Kategorien der Variablen A verschieden. Die Restriktion $\sum_j \beta_{A(j)} = 0$ reduziert die Anzahl der unabhängigen Parameter auf J - 1, denn (siehe (5.34)) der Parameter $\beta_{A(J)}$ der Referenzkategorie ergibt sich als negative Summe der J - 1 anderen Parameter $\beta_{A(j)}$ ($j=1, \dots, J-1$).
- Die Werte der Parameter $\beta_{B(k)}$ sind für jede der K Kategorien der Variablen B verschieden. Die Restriktion $\sum_k \beta_{B(k)} = 0$ reduziert die Anzahl der unabhängigen Parameter auf K - 1, denn (siehe (5.34)) der Parameter $\beta_{B(K)}$ der Referenzkategorie ergibt sich als negative Summe der K - 1 anderen Parameter $\beta_{B(k)}$ ($k=1, \dots, K-1$).
- Die Werte der Parameter $\beta_{AB(jk)}$ sind für jede der J · K Kategorienkombinationen der Variablen A und B verschieden, d.h., sie bilden ein J · K großes Feld von Parametern. Die Restriktion $\sum_j \beta_{AB(jk)} = 0$ reduziert in jeder Spalte die Anzahl der unabhängigen Parameter auf J - 1, denn der Parameter $\beta_{AB(Jk)}$ ergibt sich als negative Summe der anderen Parameter $\beta_{AB(jk)}$ ($j = 1, \dots, J - 1$). Die Restriktion

5 Modellierung multinomialer Daten (log-lineare Modelle)

$\sum_k \beta_{AB(jk)} = 0$ reduziert in jeder Zeile die Anzahl der unabhängigen Parameter auf $K - 1$, denn der Parameter $\beta_{AB(jK)}$ ergibt sich als negative Summe der anderen Parameter $\beta_{AB(jk)}$ ($k = 1, \dots, K - 1$). Somit resultieren $(J - 1)(K - 1) = J \cdot K - J - K + 1$ unabhängige Parameter $\beta_{AB(jk)}$.

Tabelle 5.2 fasst diese Überlegungen zur Anzahl der unabhängigen Parameter der einzelnen Terme zusammen.

Tabelle 5.2: Anzahl der unabhängigen Parameter der Terme des log-linearen Modells

(5.44)

Term	Anzahl der unabhängigen Parameter
β	1
$\beta_{A(j)}$	$J - 1$
$\beta_{B(k)}$	$K - 1$
$\beta_{AB(jk)}$	$J \cdot K - J - K + 1$
Gesamt	$J \cdot K$

Die Summe $J \cdot K$ der unabhängigen Parameter im saturierten Modell stimmt mit der Anzahl der Zellen überein. Die Spezifizierung aller $(J - 1)(K - 1)$ Parameter $\beta_{AB(jk)} = 0$ führt wieder zum Unabhängigkeitsmodell.

Notiert man das saturierte Modell (5.44) ausführlich (was hier der Übersichtlichkeit halber mit $J = 2$ und $K = 3$ erfolgen soll)

$$\begin{aligned}
 \log m_{11} &= \beta + \beta_{A(1)} && + \beta_{B(1)} && + \beta_{AB(11)} \\
 \log m_{12} &= \beta + \beta_{A(1)} && + \beta_{B(2)} && + \beta_{AB(12)} \\
 \log m_{13} &= \beta + \beta_{A(1)} && + \beta_{B(3)} && + \beta_{AB(13)} \\
 \log m_{21} &= \beta && + \beta_{A(2)} + \beta_{B(1)} && + \beta_{AB(21)} \\
 \log m_{22} &= \beta && + \beta_{A(2)} + \beta_{B(2)} && + \beta_{AB(22)} \\
 \log m_{23} &= \beta && + \beta_{A(2)} + \beta_{B(3)} && + \beta_{AB(23)},
 \end{aligned}$$

so sieht man, dass zu jedem Parameter der Gleichungen eine dummy-Variable X eingeführt werden kann, die nur 0 oder 1 enthält:

$$\begin{pmatrix} \log m_{11} \\ \log m_{12} \\ \log m_{13} \\ \log m_{21} \\ \log m_{22} \\ \log m_{23} \end{pmatrix} = \beta \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \beta_{A(1)} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \beta_{A(2)} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \beta_{B(1)} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \beta_{B(2)} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + \beta_{B(3)} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

$$+\beta_{AB(11)} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \beta_{AB(12)} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \beta_{AB(13)} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \beta_{AB(21)} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \beta_{AB(22)} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + \beta_{AB(23)} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Berücksichtigt man nun die Restriktionen (5.45) und nimmt man als Referenzkategorie jeweils die letzte Kategorie (d.h. allgemein bei der Variablen A die J-te Kategorie, bei der Variablen B die K-te Kategorie), dann werden die zu diesen Kategorien gehörenden Terme aus dem Modell herausgenommen, denn sie ergeben sich gemäß (5.33). In den X-Variablen bei den verbleibenden Termen wird dies durch eine Kodierung mit -1 berücksichtigt.

Die Kodierung erfolgt somit in der folgenden Weise:

für die Variable A:

$$x_j = \begin{cases} 1 & \text{für Kategorie j der Variablen A} \\ -1 & \text{für Kategorie J der Variablen A} \\ 0 & \text{für alle anderen Kategorien der Variablen A;} \end{cases} \quad (5.47)$$

für die Variable B:

$$x_k = \begin{cases} 1 & \text{für Kategorie k der Variablen B} \\ -1 & \text{für Kategorie K der Variablen B} \\ 0 & \text{für alle anderen Kategorien der Variablen B.} \end{cases} \quad (5.48)$$

Die Kodierung für die X-Variablen bei den Interaktionstermen erfolgt durch Multiplikation der Werte der X-Variablen bei den jeweiligen Haupteffekt-Termen.

Dadurch wird gesichert, dass in dem Modell nur die Anzahl von unabhängigen Parametern laut Tabelle 5.2 enthalten ist (für das Beispiel $1+1+2+2 = 6$).

Für das obige Beispiel ergibt sich gemäß dieser sum-to-zero Kodierung:

$$\begin{pmatrix} \log m_{11} \\ \log m_{12} \\ \log m_{13} \\ \log m_{21} \\ \log m_{22} \\ \log m_{23} \end{pmatrix} = \beta \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \beta_{A(1)} \begin{pmatrix} 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{pmatrix} + \beta_{B(1)} \begin{pmatrix} 1 \\ 0 \\ -1 \\ 1 \\ 0 \\ -1 \end{pmatrix} \\ + \beta_{B(2)} \begin{pmatrix} 1 \\ 1 \\ -1 \\ 0 \\ 1 \\ -1 \end{pmatrix} + \beta_{AB(11)} \begin{pmatrix} 1 \\ 0 \\ -1 \\ -1 \\ 0 \\ 1 \end{pmatrix} + \beta_{AB(12)} \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \\ -1 \\ 1 \end{pmatrix}$$

Unter Verwendung der Vektoren

$$\boldsymbol{\eta}^T = (\eta_{11}, \dots, \eta_{JK}), \mathbf{log} \mathbf{m} = (\log m_{11}, \dots, \log m_{JK}), \boldsymbol{\beta}^T = (\beta, \beta_{A(1)}, \dots, \beta_{AB(JK)})$$

und durch Zusammenfassung der X-Variablen in der Design-Matrix \mathbf{X} erhält man:

$$\boldsymbol{\eta} = \mathbf{log} \mathbf{m} = \mathbf{X}\boldsymbol{\beta} \quad (5.49)$$

bzw. bei fortlaufender Numerierung der Zellen mit $q = 1, \dots, Q$

$$\eta_q = \log m_q = \mathbf{x}_q^T \boldsymbol{\beta}, \quad (5.50)$$

wobei \mathbf{x}_q^T die q-te Zeile der Matrix \mathbf{X} ist. Diese Matrixnotation ist vor allem für multi-dimensionale Kontingenztabellen günstig.

Die Modelle zu den Hypothesen im Abschnitt 5.3 erhält man durch Weglassen bestimmter Terme und Berücksichtigung der entsprechenden Nebenbedingungen, wodurch aber die grundsätzliche Form von (5.49) bzw. (5.50) erhalten bleibt.

Die Multinomialverteilung gehört ebenso wie die Binomialverteilung und die Poisson-Verteilung zur exponentiellen Familie von Verteilungen, denn es ist

$$f(y; \pi) = \exp \left\{ \log(n!) - \sum_{j=1}^J \log(y_j!) + \sum_{j=1}^J y_j \log \pi_j \right\}. \quad (5.51)$$

Somit kann zusammengefasst werden:

Das log-lineare Modell ist ein verallgemeinertes lineares Modell mit (5.49) bzw. (5.50) als

link Funktion. Alle einbezogenen Variablen sind erklärende Variablen; die abhängigen Variablen sind die Zelhäufigkeiten.

In log-linearen Modellen sind die Terme höherer Ordnung als Differenzen von den Termen niedrigerer Ordnung definiert. So repräsentiert z.B.

- $\beta_{A(j)}$ den zusätzlichen Effekt der j-ten Kategorie (Zeile j) der Variablen A über den durchschnittlichen Effekt β hinaus;
- $\beta_{AB(jk)}$ repräsentiert den zusätzlichen Effekt über den durchschnittlichen Effekt plus den Effekt der j-ten Kategorie der Variablen A (Zeile) plus den Effekt der k-ten Kategorie der Variablen B (Spalte) hinaus.

Dabei geht man im allgemeinen von hierarchischen Modellen aus. Ein hierarchisches Modell ist dadurch gekennzeichnet, dass Terme höherer Ordnung nur in das Modell eingeschlossen werden, wenn bereits alle korrespondierenden Terme niedrigerer Ordnung im Modell enthalten sind. So wird der Interaktionsterm $\beta_{AB(jk)}$ nur in das Modell aufgenommen, wenn bereits die Haupteffekte $\beta_{A(j)}$ und $\beta_{B(k)}$ im Modell enthalten sind. Umgekehrt, wenn ein Term niedrigerer Ordnung im Modell nicht enthalten ist, so sind auch alle korrespondierenden Terme höherer Ordnung nicht im Modell enthalten.

Bei der Analyse von Kontingenztabelle richtet sich das Hauptinteresse vor allem auf die Assoziation zwischen zwei Variablen, weshalb in log-linearen Modellen die Interaktionsterme zwischen zwei oder mehreren Variablen im Mittelpunkt von Hypothesen stehen.

Entsprechend dem Stichprobenmodell sind in den Formeln für die Berechnung der erwarteten Zelhäufigkeiten bestimmte Terme festgelegt, z.B. n beim multinomialen Stichprobenmodell bzw. y_{j+} beim produkt-multinomialen Stichprobenmodell in einer zweidimensionalen Kontingenztabelle. Das hat zur Konsequenz, dass die korrespondierenden Parameter stets im log-linearen Modell enthalten sein müssen. Die Tabelle 5.3 enthält eine Übersicht über die log-linearen Modelle für eine zweidimensionale Kontingenztabelle.³⁵ Zu beachten ist, dass Unterschiede hinsichtlich der einzuschliessenden Terme und

³⁵Nach Dobson, A.J. (1991), S. 131.

vor allem Unterschiede in der Interpretation bestehen, obwohl das gleiche Modell auf jede der angegebenen Verteilungen anwendbar ist.

Tabelle 5.3: Log-lineare Modelle für eine zweidimensionale Kontingenztafel

Log-lineares Modell	Poisson-Verteilung	Multinomialverteilung	Produkt-multinomial Verteilung mit festen Zeilensummen
Saturiertes Modell $\beta + \beta_{A(j)} + \beta_{B(k)} + \beta_{AB(jk)}$ mit JK unabhängigen Parametern	$E(Y_{jk}) = \lambda_{jk}$	$E(Y_{jk}) = n\pi_{jk}$ mit $\sum_j \sum_k \pi_{jk} = 1$	$E(Y_{jk}) = y_{j+} \pi_{jk}$ mit $\sum_k \pi_{jk} = 1$ für $j = 1, \dots, J$
$\beta + \beta_{A(j)} + \beta_{B(k)}$ mit J+K-1 unabhängigen Parametern	Unabhängigkeitshypothese $E(Y_{jk}) = \lambda_j \lambda_k$	Unabhängigkeitshypothese $E(Y_{jk}) = n\pi_{j+} \pi_{+k}$ mit $\sum_j \pi_{j+} = \sum_k \pi_{+k} = 1$	Homogenitätshypothese $E(Y_{jk}) = y_{j+} \pi_{+k}$ mit $\sum_k \pi_{+k} = 1$
einzuschließende Terme		β , da n fest vorgegeben	$\beta + \beta_{A(j)}$, da y_{j+} fest vorgegeben

5.5 Schätzung log-linearer Modelle

Die Schätzung der erwarteten Zellhäufigkeiten erfolgt durch die Maximierung der log-Likelihood-Funktion unter Berücksichtigung der Nebenbedingungen. Ausgehend von (5.14) können die Likelihood-Funktion und die log-Likelihood-Funktion der multinomialverteilten Zufallsvariablen in der folgenden Form geschrieben werden³⁶:

$$L(\boldsymbol{\pi}; \mathbf{y}) = \frac{n!}{y_1! \cdots y_{J-1}! (n - \sum_{j=1}^{J-1} y_j)!} \cdot \prod_{j=1}^{J-1} \pi_j^{y_j} \left(1 - \sum_{j=1}^{J-1} \pi_j \right)^{n - \sum_{j=1}^{J-1} y_j} \quad (5.52)$$

$$l(\boldsymbol{\pi}; \mathbf{y}) = \log(n!) - \sum_{j=1}^{J-1} \log(y_j!) - \log \left[(n - \sum_{j=1}^{J-1} y_j)! \right] \quad (5.53)$$

$$+ \sum_{j=1}^{J-1} y_j \log \pi_j + \left(n - \sum_{j=1}^{J-1} y_j \right) \cdot \log \left(1 - \sum_{j=1}^{J-1} \pi_j \right).$$

³⁶Siehe Fahrmeir, L., Hamerle, A. (1984), S. 34, 55, 60.

Differenzieren nach π_j und Nullsetzen führt zu:

$$\begin{aligned} \frac{\partial l(\boldsymbol{\pi}; \mathbf{y})}{\partial \pi_j} &= y_j \cdot \frac{1}{\pi_j} + \left(n - \sum_{j=1}^{J-1} y_j \right) \cdot \frac{1}{1 - \sum_{j=1}^{J-1} \pi_j} \cdot (-1) \\ &= \frac{y_j}{\pi_j} - \frac{n - \sum_{j=1}^{J-1} y_j}{1 - \sum_{j=1}^{J-1} \pi_j} = \frac{y_j}{\pi_j} - \frac{y_J}{\pi_J} \doteq 0 \end{aligned} \quad (5.54)$$

für $j = 1, \dots, J-1$. Aus diesen Gleichungen erhält man: $y_j/\hat{\pi}_j = c$ mit $c = y_J/\hat{\pi}_J$ und schließlich $\hat{\pi}_j = y_j/c$ ($j = 1, \dots, J$). Für die Summation über $j = 1, \dots, J$ resultiert:

$$1 = \sum_j \hat{\pi}_j = \sum_j y_j/c = n/c$$

und somit $c = n$. Als ML-Schätzer ergibt sich damit:

$$\hat{\pi}_j = y_j/n, \quad j = 1, \dots, J, \quad (5.55)$$

d.h., die ML-Schätzer der Wahrscheinlichkeiten sind die beobachteten relativen Häufigkeiten. Der ML-Schätzer für die erwarteten Häufigkeiten ist:

$$\hat{m}_j = n \cdot \hat{\pi}_j = y_j, \quad j = 1, \dots, J. \quad (5.56)$$

Mit diesem Ergebnis können die ML-Schätzer bei den verschiedenen Stichprobenmodellen ermittelt werden.

Multinomiales Stichprobenmodell:

Ausgehend von (5.18)

$$f(y|n, \pi) = \frac{n!}{\prod_{j=1}^J \prod_{k=1}^K y_{jk}!} \cdot \prod_{j=1}^J \prod_{k=1}^K \pi_{jk}^{y_{jk}} \quad (5.18)$$

ist die log-Likelihood-Funktion

$$l(\pi; y) = \log(n!) - \sum_{j=1}^J \sum_{k=1}^K \log(y_{jk}!) + \sum_{j=1}^J \sum_{k=1}^K y_{jk} \log \pi_{jk}. \quad (5.57)$$

5 Modellierung multinomialer Daten (log-lineare Modelle)

Zur Maximierung der log-Likelihood-Funktion braucht nur der letzte Term auf der rechten Seite betrachtet werden. Nach (5.55) wird das Maximum bei $\pi_{jk} = \hat{\pi}_{jk}$ mit

$$\hat{\pi}_{jk} = y_{jk}/n, \quad j = 1, \dots, J; \quad k = 1, \dots, K \quad (5.58)$$

erreicht. Entsprechend gilt nach (5.56):

$$\hat{m}_{jk} = n \cdot \hat{\pi}_{jk} = y_{jk}, \quad j = 1, \dots, J; \quad k = 1, \dots, K. \quad (5.59)$$

Der ML-Schätzer der Zellwahrscheinlichkeit ist die beobachtete relative Zellhäufigkeit und der ML-Schätzer für die erwartete Zellhäufigkeit ist die beobachtete Zellhäufigkeit. Für die beiden Randverteilungen erhält man analog gemäß (5.55) und (5.56):

$$\begin{aligned} \hat{\pi}_{j+} &= y_{j+}/n, \quad j = 1, \dots, J & \hat{\pi}_{+k} &= y_{+k}/n, \quad k = 1, \dots, K; \\ \hat{m}_{j+} &= n \cdot \hat{\pi}_{j+} = y_{j+}, \quad j = 1, \dots, J & \hat{m}_{+k} &= n \cdot \hat{\pi}_{+k} = y_{+k}, \quad k = 1, \dots, K. \end{aligned}$$

Berücksichtigt man nun die Nullhypothese (5.24), so folgt

$$\hat{\pi}_{jk}^0 = \hat{\pi}_{j+} \hat{\pi}_{+k} = (y_{j+}/n)(y_{+k}/n) \text{ und } \hat{m}_{jk}^0 = n \cdot \hat{\pi}_{jk}^0 = y_{j+} y_{+k} / n. \quad (5.60)$$

Produkt-multinomialles Stichprobenmodell:

Ausgehend von (5.21)

$$f(y_{j1}, \dots, y_{jk} | y_{1+}, \dots, y_{J+}) = \prod_{j=1}^J \frac{y_{j+}!}{K^{y_{j+}}} \cdot \prod_{k=1}^K \pi_{jk|j+}^{y_{jk}} \quad (5.21)$$

ist die log-Likelihood-Funktion

$$l(\pi; y) = \sum_{j=1}^J \log(y_{j+}!) - \sum_{j=1}^J \sum_{k=1}^K \log(y_{jk}!) + \sum_{j=1}^J \sum_{k=1}^K y_{jk} \log \pi_{jk}. \quad (5.61)$$

Auch hier braucht zur Maximierung der log-Likelihood-Funktion nur der letzte Term auf der rechten Seite betrachtet werden. Nach (5.55) wird das Maximum bei $\pi_{jk} = \hat{\pi}_{jk}$ mit

$$\hat{\pi}_{jk} = y_{jk}/y_{j+}, \quad j = 1, \dots, J; \quad k = 1, \dots, K \quad (5.62)$$

erreicht. Entsprechend gilt nach (5.56):

$$\hat{m}_{jk} = y_{j+} \cdot \hat{\pi}_{jk} = y_{jk}, \quad j = 1, \dots, J; \quad k = 1, \dots, K. \quad (5.63)$$

Beim produkt-multinomialen Stichprobenmodell ist der ML-Schätzer der Zellwahrscheinlichkeit die beobachtete bedingte relative Zellhäufigkeit und der ML-Schätzer für die erwartete Zellhäufigkeit die beobachtete Zellhäufigkeit.

Berücksichtigt man nun die Nullhypothese (5.27), wobei $\hat{\pi}_{1k} = \dots = \hat{\pi}_{Jk} = \hat{\pi}_k$ gesetzt wird, so folgt gemäß (5.55) und (5.56):

$$\hat{\pi}_{jk}^0 = \hat{\pi}_k^0 = y_{+k}/n \text{ und } \hat{m}_{jk}^0 = y_{j+} \cdot \hat{\pi}_k^0 = y_{j+}y_{+k}/n. \quad (5.64)$$

Für das saturierte Modell ergibt sich die geschätzte link Funktion unter Berücksichtigung von (5.59) bzw. (5.63) zu

$$\hat{\eta}_{jk} = \log \hat{m}_{jk} = \log y_{jk} = b + b_{A(j)} + b_{B(k)} + b_{AB(jk)}, \quad (5.65)$$

wobei die geschätzten Parameter mit b symbolisiert werden.

Bei Gültigkeit der Nullhypothese ergibt sich die geschätzte link Funktion unter Berücksichtigung von (5.60) bzw. (5.64) zu:

$$\hat{\eta}_{jk} = \log \hat{m}_{jk} = \log(y_{j+}y_{+k}/n) = b + b_{A(j)} + b_{B(k)} \quad (5.66)$$

bzw.

$$\hat{\eta}_{jk} = \log \hat{m}_{jk} = \log y_{j+} + \log y_{+k} - \log n = b + b_{A(j)} + b_{B(k)}.$$

Verwendung der sum-to-zero Restriktion:

Bei Verwendung der sum-to-zero Restriktion erhält man mittels Summenbildung und unter Beachtung der Restriktionen (5.45) die Schätzungen der Parameter für das saturierte Modell:

$$\begin{aligned} \sum_{j=1}^J \sum_{k=1}^K \log y_{jk} &= \sum_{j=1}^J \sum_{k=1}^K (b + b_{A(j)} + b_{B(k)} + b_{AB(jk)}) \\ &= JKb + K \sum_{j=1}^J b_{A(j)} + J \sum_{k=1}^K b_{B(k)} + \sum_{j=1}^J \sum_{k=1}^K b_{AB(jk)} = JKb \\ b &= \frac{1}{JK} \sum_{j=1}^J \sum_{k=1}^K \log y_{jk}. \end{aligned} \quad (5.67)$$

Die geschätzte Konstante b ist das arithmetische Mittel der Logarithmen aller beobachteten Zellhäufigkeiten.

$$\begin{aligned}
 \sum_{k=1}^K \log y_{jk} &= \sum_{k=1}^K (b + b_{A(j)} + b_{B(k)} + b_{AB(jk)}) \\
 &= Kb + Kb_{A(j)} + \sum_{k=1}^K b_{B(k)} + \sum_{k=1}^K b_{AB(jk)} = Kb + Kb_{A(j)} \\
 b_{A(j)} &= \frac{1}{K} \sum_{k=1}^K \log y_{jk} - b.
 \end{aligned} \tag{5.68}$$

Der geschätzte Parameter $b_{A(j)}$ beinhaltet die Differenz zwischen dem Mittelwert der Logarithmen der beobachteten Zellhäufigkeiten der Zeile j und dem Gesamtmittelwert.

$$\begin{aligned}
 \sum_{j=1}^J \log y_{jk} &= \sum_{j=1}^J (b + b_{A(j)} + b_{B(k)} + b_{AB(jk)}) \\
 &= Jb + \sum_{j=1}^J b_{A(j)} + Jb_{B(k)} + \sum_{j=1}^J b_{AB(jk)} = Jb + Jb_{B(k)} \\
 b_{B(k)} &= \frac{1}{J} \sum_{j=1}^J \log y_{jk} - b.
 \end{aligned} \tag{5.69}$$

Der geschätzte Parameter $b_{B(k)}$ beinhaltet die Differenz zwischen dem Mittelwert der Logarithmen der beobachteten Zellhäufigkeiten der Spalte k und dem Gesamtmittelwert.

$$b_{AB(jk)} = \log y_{jk} - (b + b_{A(j)} + b_{B(k)}). \tag{5.70}$$

Der geschätzte Interaktionsparameter $b_{AB(jk)}$ beinhaltet die Differenz zwischen dem Logarithmus der beobachteten Zellhäufigkeiten der Kategorie j von A und der Kategorie k von B (Zelle jk) und der Summe der geschätzten Konstanten, des geschätzten Parameters $b_{A(j)}$ für den Haupteffekt von A und des geschätzten Parameters $b_{B(k)}$ für den Haupteffekt von B.

Bei Verwendung der sum-to-zero Restriktion erhält man mittels Summenbildung und unter Beachtung der Restriktionen (5.33) die Schätzungen der Parameter für das Modell unter H_0 :

$$\sum_{j=1}^J \sum_{k=1}^K (\log y_{j+} + \log y_{+k} - \log n) = \sum_{j=1}^J \sum_{k=1}^K (b + b_{A(j)} + b_{B(k)})$$

$$\begin{aligned}
K \sum_{j=1}^J \log y_{j+} + J \sum_{k=1}^K \log y_{+k} - JK \log n &= JKb + K \sum_{j=1}^J b_{A(j)} + J \sum_{k=1}^K b_{B(k)} \\
b &= \frac{1}{J} \sum_{j=1}^J \log y_{j+} + \frac{1}{K} \sum_{k=1}^K \log y_{+k} - \log n
\end{aligned} \tag{5.71}$$

Die geschätzte Konstante b ist die Differenz zwischen der Summe der arithmetischen Mittel der Logarithmen der beobachteten Zellhäufigkeiten der Spalten und der Logarithmen der beobachteten Zellhäufigkeiten der Zeilen und dem Logarithmus des Stichprobenumfanges.

$$\begin{aligned}
\sum_{k=1}^K (\log y_{j+} + \log y_{+k} - \log n) &= \sum_{k=1}^K (b + b_{A(j)} + b_{B(k)}) \\
K \log y_{j+} + J \sum_{k=1}^K \log y_{+k} - K \log n &= Kb + Kb_{A(j)} + \sum_{k=1}^K b_{B(k)} \\
b_{A(j)} &= \log y_{j+} + \frac{1}{K} \sum_{k=1}^K \log y_{+k} - \log n - b \\
&= \log y_{j+} + \frac{1}{K} \sum_{k=1}^K \log y_{+k} - \log n - \frac{1}{J} \sum_{j=1}^J \log y_{j+} - \frac{1}{K} \sum_{k=1}^K \log y_{+k} + \log n \\
b_{A(j)} &= \log y_{j+} - \frac{1}{J} \sum_{j=1}^J \log y_{j+}
\end{aligned} \tag{5.72}$$

Der geschätzte Parameter $b_{A(j)}$ beinhaltet die Differenz zwischen dem Logarithmus der beobachteten Randhäufigkeit der Kategorie j und dem Mittelwert der Logarithmen der beobachteten Randhäufigkeiten der Variablen A .

$$\begin{aligned}
\sum_{j=1}^J (\log y_{j+} + \log y_{+k} - \log n) &= \sum_{j=1}^J (b + b_{A(j)} + b_{B(k)}) \\
\sum_{j=1}^J \log y_{j+} + J \log y_{+k} - J \log n &= Jb + \sum_{j=1}^J b_{A(j)} + Jb_{B(k)}
\end{aligned}$$

$$\begin{aligned}
 b_{B(K)} &= \frac{1}{J} \sum_{j=1}^J \log y_{j+} + \log y_{+K} - \log n - b \\
 &= \frac{1}{J} \sum_{j=1}^J \log y_{j+} + \log y_{+K} - \log n - \frac{1}{J} \sum_{j=1}^J \log y_{j+} - \frac{1}{K} \sum_{k=1}^K \log y_{+k} + \log n \\
 b_{B(k)} &= \log y_{+k} - \frac{1}{K} \sum_{k=1}^K \log y_{+k}
 \end{aligned} \tag{5.73}$$

Der geschätzte Parameter $b_{B(k)}$ beinhaltet die Differenz zwischen dem Logarithmus der beobachteten Randhäufigkeit der Kategorie k und dem Mittelwert der Logarithmen der beobachteten Randhäufigkeiten der Variablen B.

Verwendung der corner-point Restriktion:

Wählt man die corner-point Restriktion, wird jeder Kategorie der Variablen A und B eine Dummy-Variable (0;1-Variable, Scheinvariable) zugeordnet. Die betrachtete Kategorie erhält in der Dummy-Variablen eine Eins und in allen anderen Dummy-Variablen eine Null. Die Werte für die Referenzkategorie werden in allen Dummy-Variablen Null gesetzt. Damit sind die Parameter der Referenzkategorie gleich Null gesetzt.

Für die Parameter des saturierten Modells erhält man:

$$b = \log y_{JK}. \tag{5.74}$$

Die geschätzte Konstante b entspricht dem Logarithmus der beobachteten Zellhäufigkeit der Referenzkategorien von A und B (Zelle JK).

$$b_{A(j)} = \log y_{jK} - \log y_{JK} = \log y_{jK} - b, \quad j = 1, \dots, J - 1. \tag{5.75}$$

Der geschätzte Parameter $b_{A(j)}$ ist die Differenz zwischen den Logarithmen der beobachteten Zellhäufigkeiten der j-ten Kategorie und der Referenzkategorie J von A, gegeben die Referenzkategorie K von B (Zellen jK und JK) und gibt den Effekt der Variablen A an, der beim Übergang von ihrer Referenzkategorie J zur j-ten Kategorie entsteht.

$$b_{B(k)} = \log y_{Jk} - \log y_{JK} = \log y_{Jk} - b, \quad k = 1, \dots, K - 1. \tag{5.76}$$

Der geschätzte Parameter $b_{B(k)}$ ist die Differenz zwischen den Logarithmen der beobachteten Zellhäufigkeiten der k-ten Kategorie und der Referenzkategorie K von B, gegeben

die Referenzkategorie J von A (Zellen Jk und JK) und gibt den Effekt der Variablen B an, der beim Übergang von ihrer Referenzkategorie K zur k-ten Kategorie entsteht.

$$b_{AB(jk)} = \log y_{jk} - \log y_{jK} - b_{B(k)}, \quad j = 1, \dots, J-1, \quad k = 1, \dots, K-1. \quad (5.77)$$

Die Interaktionsparameter $b_{AB(jk)}$ geben die über den Haupteffekt der Kategorie k von B hinausgehende Wirkung an, wenn für eine gegebene Kategorie j von A der Übergang von der Referenzkategorie K zur k-ten Kategorie von B erfolgt.

Für die Parameter des Modells bei Gültigkeit der Nullhypothese H_0 erhält man:

$$b = \log y_{J+} + \log y_{+K} - \log n. \quad (5.78)$$

Die geschätzte Konstante b ist die Differenz zwischen der Summe der Logarithmen der beobachteten Randhäufigkeiten der Referenzkategorien von A und B und dem Logarithmus des Stichprobenumfanges n.

$$b_{A(j)} = \log y_{j+} - \log y_{J+}, \quad j = 1, \dots, J-1. \quad (5.79)$$

Der geschätzte Parameter $b_{A(j)}$ ist die Differenz zwischen den Logarithmen der beobachteten Randhäufigkeiten der j-ten Kategorie und der Referenzkategorie J von A.

$$b_{B(k)} = \log y_{+k} - \log y_{+K}, \quad k = 1, \dots, K-1. \quad (5.80)$$

Der geschätzte Parameter $b_{B(k)}$ ist die Differenz zwischen den Logarithmen der beobachteten Randhäufigkeiten der k-ten Kategorie und der Referenzkategorie K von B.

In höherdimensionalen Kontingenztabellen kann ein solches direktes Ergebnis nicht in jedem Fall erreicht werden. Es wird dann ein iteratives Verfahren (iterative proportional fitting) verwendet, bei dem die geschätzten erwarteten Zellhäufigkeiten solange angepasst werden, bis sie sich mit einer vorgegebenen Genauigkeit zu den geforderten Randsummen aufsummieren. ³⁷

Um die Bedeutung der verschiedenen Parameterschätzungen vergleichen zu können, werden standardisierte Schätzungen berechnet, die sich durch die Division der Parameterschätzung durch ihre Standardfehler ergeben. Die asymptotischen Varianzen der

³⁷Siehe u.a. Bishop, Fienberg, Holland (1991), S. 83 ff.

geschätzten Parameter sind:

$$\begin{aligned}
 \text{Var}(b) &= \left(\frac{1}{JK}\right)^2 \sum_{j=1}^J \sum_{k=1}^K \left(\frac{1}{y_{jk}}\right) - \frac{1}{n} \\
 \text{Var}(b_{A(j)}) &= \left(\frac{1}{JK}\right)^2 \sum_{j=1}^J \sum_{k=1}^K \left(\frac{1}{y_{jk}}\right) + \left(\frac{J-1}{JK^2}\right) \sum_{k=1}^K \left(\frac{1}{y_{jk}}\right) \\
 \text{Var}(b_{B(k)}) &= \left(\frac{1}{JK}\right)^2 \sum_{j=1}^J \sum_{k=1}^K \left(\frac{1}{y_{jk}}\right) + \left(\frac{K-1}{KJ^2}\right) \sum_{j=1}^J \left(\frac{1}{y_{jk}}\right) \\
 \text{Var}(b_{AB(jk)}) &= \left(\frac{1}{JK}\right)^2 \sum_{j=1}^J \sum_{k=1}^K \left(\frac{1}{y_{jk}}\right) + \left(\frac{J-1}{JK^2}\right) \sum_{k=1}^K \left(\frac{1}{y_{jk}}\right) \\
 &\quad + \left(\frac{K-1}{KJ^2}\right) \sum_{j=1}^J \left(\frac{1}{y_{jk}}\right) + \frac{(J-1)(K-1)}{JK} \left(\frac{1}{y_{jk}}\right)
 \end{aligned} \tag{5.81}$$

$$\tag{5.82}$$

5.6 Hypothesenprüfung

Hat man für ein gegebenes log-lineares Modell die erwarteten Zellhäufigkeiten geschätzt, so stellt sich als nächstes die Frage, wie gut die Anpassung durch dieses Modell erfolgte. Zur Prüfung der Anpassung (goodness-of-fit) wird im allgemeinen einer der folgenden Tests verwendet.

Pearson's Chi-Quadrat-Statistik:

Pearson's Chi-Quadrat-Statistik³⁸ setzt jedoch voraus, dass

- die Beobachtungen unabhängig sind,
- die Aufteilung der Variablenausprägungen vollständig ist, d.h., jede statistische Einheit (Fall) gehört zu einem Paar von Variablenausprägungen (A_j, B_k) ,
- große Stichproben vorliegen,

³⁸Zur Pearson's Chi-Quadrat-Statistik siehe u.a. Rönz, B., Strohe, H.G. (1994), S. 67 ff.; Bamberg, G., Baur, F. (1996), S. 198 ff.; Schlittgen, R. (1990), S. 384 ff.; Büning, H., Trenkler, G. (1994), S. 220 ff.; Schwarze, J. (1990), S. 249 ff.; Bosch, K. (1992), S. 384 ff.; Backhaus, K., Erichson, B., Plinke, W., Weiber, R. (2000), S. 234 ff.; Hartung, J., Elpelt, B., Klöser, K.-H. (1993), S. 413 ff.

- die Zufallsstichprobe aus einer multinomialverteilten Grundgesamtheit stammt.

Die Teststatistik ist:

$$\chi^2 = \sum_{j=1}^J \sum_{k=1}^K \frac{(y_{jk} - \hat{m}_{jk}^0)^2}{\hat{m}_{jk}^0}. \quad (5.83)$$

Zur Überprüfung der Nullhypothese werden die beobachteten mit den unter H_0 erwarteten Zellhäufigkeiten verglichen. Wenn H_0 wahr ist, werden y_{jk} und \hat{m}_{jk}^0 weitgehend übereinstimmen, d.h., es treten nur kleine Differenzen $y_{jk} - \hat{m}_{jk}^0$ auf, womit die Teststatistik χ^2 einen kleinen Wert annimmt. Wenn H_0 nicht gültig ist, werden die Abweichungen $y_{jk} - \hat{m}_{jk}^0$ groß sein und die Teststatistik χ^2 nimmt einen großen Wert an. Daraus folgt unmittelbar, dass die Nullhypothese für “zu große“ Werte von χ^2 abgelehnt wird. Was als große Abweichung anzusehen ist, hängt wiederum von der Größe der Zellhäufigkeiten ab: Eine Abweichung von 5 ist bei $y_{jk} = 15$ und $\hat{m}_{jk}^0 = 10$ als groß und bei $y_{jk} = 1015$ und $\hat{m}_{jk}^0 = 1010$ als klein anzusehen. Zur Korrektur dieses Problems wird durch die erwarteten Zellhäufigkeiten dividiert, d.h., \hat{m}_{jk}^0 im Nenner von (5.83) ist ein Skalierungsfaktor zur Berücksichtigung der relativen Größe der Abweichungen.

Die Komponenten von Pearson’s Chi-Quadrat-Statistik sind die quadrierten standardisierten Residuen (5.12), so dass festgestellt werden kann, welche Zellen den größten Beitrag zum Wert von χ^2 leisten.

In großen Stichproben kann über die Stichprobenhäufigkeiten y_{j+}/n und y_{+k}/n (aufgrund des zentralen Grenzwertsatzes) angenommen werden, dass sie normalverteilt sind. Die Verteilung von Pearson’s χ^2 -Teststatistik folgt dann einer Verteilung der Summe der Quadrate von standardisierten normalverteilten Zufallsvariablen. Damit ist für große Stichproben die Teststatistik χ^2 unter H_0 approximativ chi-quadrat-verteilt mit f Freiheitsgraden. Die Nullhypothese wird zum vorgegebenen Signifikanzniveau α abgelehnt, wenn $\chi^2 > \chi_{1-\alpha;f}^2$ ist. $\chi_{1-\alpha;f}^2$ ist das Quantil der Chi-Quadrat-Verteilung der Ordnung $1 - \alpha$, das aus der Tabelle der Chi-Quadrat-Verteilung entnommen werden kann. Bei der Ermittlung der Anzahl der Freiheitsgrade ist zu beachten, dass nicht die wahren, sondern die aus der Stichprobe geschätzten Randverteilungen verwendet werden. Da sich die geschätzten Randwahrscheinlichkeiten zu Eins addieren, werden bei der Randverteilung

von A nur $J - 1$ und bei der Randverteilung von B nur $K - 1$ Schätzungen benötigt, da sich der jeweilige letzte Randwert automatisch ergibt. Gleiches gilt für die aus der Stichprobe geschätzten relativen Zellwahrscheinlichkeiten, bei denen nur $J \cdot K - 1$ notwendig sind. Die Anzahl der Freiheitsgrade ergibt sich als: $f = \text{Anzahl der Zellen} - \text{Anzahl der geschätzten Parameter}$.

Für das log-lineare Modell (5.32) und die zugrundeliegende Unabhängigkeitshypothese ergibt sich die Anzahl der Freiheitsgrade zu:

$$f = JK - (J - 1) - (K - 1) - 1 = JK - J - K + 1 = (J - 1)(K - 1).$$

Wählt man dagegen z.B. ein log-lineares Modell der Form $\eta_{jk} = \log m_{jk} = \beta + \beta_{A(j)}$, beträgt die Anzahl der Freiheitsgrade $f = JK - 1 - (J - 1) = JK - J$.

Der Test sollte nicht verwendet werden,

- wenn erwartete Zellhäufigkeiten kleiner als 1 sind,
- wenn mehr als 20% der Zellen erwartete Häufigkeiten kleiner als 5 aufweisen.

Tritt ein solcher Fall ein, sollte geprüft werden, ob bei den Variablen A und/oder B Kategorien zusammengefasst werden können, wodurch diese Voraussetzungen von Pearson's Chi-Quadrat-Statistik möglicherweise erfüllt werden können.

Deviance (log-likelihood ratio statistic):

Die Teststatistik³⁹ ergibt sich gemäß (2.97) zu:

$$D = 2[l(\hat{\mathbf{m}}_{max}; \mathbf{y}) - l(\hat{\mathbf{m}}; \mathbf{y})]. \quad (5.84)$$

Mit $\hat{m}_{jk} = y_{jk}$ für das saturierte Modell (siehe (5.59) und (5.63)) und $\hat{m}_{jk}^0 = y_{j+}y_{+k}/n$ für die Unabhängigkeits- bzw. Homogenitätshypothese (siehe (5.60) und (5.64)) erhält man nach Einsetzen in die jeweilige log-Likelihood-Funktion und Differenzenbildung:

$$D = 2 \sum_{j=1}^J \sum_{k=1}^K y_{jk} (\log \hat{m}_{jk} - \log \hat{m}_{jk}^0)$$

³⁹Siehe u.a. Rönz, B., Strohe, H.G. (1994), S. 219; Büning, H., Trenkler, G. (1994), S. 36; Berry, D.A., Lindgren, B.W. (1990), S. 509 ff., 579 ff., 608 f.

$$\begin{aligned}
&= 2 \sum_{j=1}^J \sum_{k=1}^K y_{jk} \log \left(\frac{\hat{m}_{jk}}{\hat{m}_{jk}^0} \right) \\
&= 2 \sum_{j=1}^J \sum_{k=1}^K y_{jk} \log \left(\frac{ny_{jk}}{y_{j+}y_{+k}} \right).
\end{aligned} \tag{5.85}$$

Für große Stichproben ist D unter H_0 asymptotisch chi-quadrat-verteilt mit f Freiheitsgraden. Die Nullhypothese wird verworfen, wenn $D > \chi_{1-\alpha;f}^2$ ist, mit $\chi_{1-\alpha;f}^2$ als Quantil der Chi-Quadrat-Verteilung der Ordnung $1 - \alpha$.

Auch bei log-linearen Modellen sollte eine Modellprüfung vorgenommen werden, indem die Residuen $y_{jk} - \hat{m}_{jk}^0$ bzw. standardisierten Residuen

$$rs_{jk} = \frac{y_{jk} - \hat{m}_{jk}^0}{\sqrt{\hat{m}_{jk}^0}}$$

inspiziert werden.

5.7 Beispiel

Zur Demonstration der Logik der log-linearen Analyse und der Interpretation der Modellparameter soll der bivariate Zusammenhang zwischen der Wahlbeteiligung an der letzten Bundestagswahl und der politischen Interessiertheit untersucht werden.⁴⁰ Im ALLBUS 1996⁴¹ findet man entsprechende Befragungsergebnisse.

Die Wahlbeteiligung (Variable A) wurde über die berichtete Wahlbeteiligung bei der letzten Bundestagswahl erfasst und weist die beiden Kategorien $A_1 = \text{ja}$ und $A_2 = \text{nein}$ auf.

Die Antwort zur Frage „Wie stark interessieren Sie sich für Politik?“ konnten die Befragten auf einer Ordinalskala von 1 bis 5 geben. Zur Vereinfachung der Darstellung werden die Ausprägungen umkodiert, so dass das politische Interesse der Befragten (Variable B) die Kategorien $B_1 = \text{stark}$, $B_2 = \text{mittel}$ und $B_3 = \text{wenig}$ aufweist. Der ordinale Charakter dieser Variablen soll im weiteren nicht berücksichtigt werden, d.h., sie wird als

⁴⁰Das Beispiel ist angelehnt an: Kühl, S.M. (1997).

⁴¹Siehe Fußnote 25.

nominalskalierte Variable behandelt. Die auswertbare Anzahl von Befragten (Stichprobenumfang) beträgt $n = 3513$.

Für die Analyse wird vom multinomialen Stichprobenmodell ausgegangen, da von einem fest vorgegebenem Stichprobenumfang ausgegangen werden kann. Aufgrund dessen sind alle Zellhäufigkeiten Y_{jk} (response Variablen) und alle Randhäufigkeiten Y_{j+} und Y_{+k} ($j = 1, 2$; $k = 1, 2, 3$) Zufallsvariable und die Zufallsvektoren $\mathbf{Y}^T = (Y_{11}, \dots, Y_{23})$, (Y_{1+}, Y_{2+}) und (Y_{+1}, Y_{+2}, Y_{+3}) sind jeweils multinomialverteilt gemäß (5.18) bzw. (5.14). Tabelle 5.4 enthält die Stichprobenrealisationen dieser Zufallsvariablen.

Tabelle 5.4: Kontingenztabelle Wahlbeteiligung und politisches Interesse

Variable A (Wahlbeteiligung)	Variable B (politisches Interesse)			Randverteilung A
	B_1 (stark)	B_2 (mittel)	B_3 (wenig)	
A_1 (ja)	822	1264	737	2823
A_2 (nein)	103	240	347	690
Randverteilung B	925	1504	1084	3513

Da die Wahrscheinlichkeiten π_{jk} bzw. die Zellhäufigkeiten m_{jk} der Grundgesamtheit (wahlberechtigte Bevölkerung der Bundesrepublik Deutschland) unbekannt sind, werden sie aus den Daten unter Berücksichtigung der Effekte der Variablen Wahlbeteiligung und politisches Interesse geschätzt.

Die Schätzung der Modellparameter wird zum einen leicht überprüfbar dargestellt und zum anderen unter Verwendung von SPSS für Windows Release 10.0.7 durchgeführt. Statistische Tests erfolgen auf einem Signifikanzniveau von $\alpha = 0,05$.

Ausgehend von der Tabelle 5.4 sind in der Tabelle 5.5 die Logarithmen der beobachteten Zell- und Randhäufigkeiten auf 6 Dezimalstellen gerundet angegeben.

Tabelle 5.5: $\log y_{jk}$ für die Stichprobe

Variable A (Wahlbeteiligung)	Variable B (politisches Interesse)			Randverteilung A
	B_1 (stark)	B_2 (mittel)	B_3 (wenig)	
A_1 (ja)	6,711740	7,142037	6,602588	7,945555
A_2 (nein)	4,634729	5,480639	5,849325	6,536692
Randverteilung B	6,829794	7,315884	6,988413	8,164226

Weiterhin sind:

$$\begin{array}{lll} \sum_j \sum_k \log y_{jk} = 36,4210575 & \sum_k \log y_{1k} = 20,456365 & \sum_k \log y_{2k} = 15,964693 \\ \sum_j \log y_{j1} = 11,346469 & \sum_j \log y_{j2} = 12,622676 & \sum_j \log y_{j3} = 12,451913 \\ \sum_j \log y_{j+} = 14,482247 & \sum_k \log y_{+k} = 21,134091. & \end{array}$$

Zur Schätzung der log-linearen Modelle muss die Design-Matrix \mathbf{X} in (5.49) spezifiziert werden. Für die Variable A (Wahlbeteiligung), die zwei Kategorien aufweist, wird eine Kontrast-Variable gebildet. Für die Variable B (politisches Interesse), die drei Kategorien aufweist, sind zwei Kontrast-Variablen festzulegen. Die spezielle Konstruktion der Kontrast-Variablen der Design-Matrix hängt von der verwendeten Restriktion ab: Die sum-to-zero Restriktion wird durch die Deviation- oder Effekt-Kodierung und die corner-point-Restriktion durch die Indikator- oder Dummy-Kodierung umgesetzt. Für beide Variablen wird die letzte Kategorie als Referenzkategorie gewählt.

Mit Hilfe der Kontrast-Variablen kann das log-lineare Modell formuliert werden, wobei eine Konstante berücksichtigt wird, der eine Kontrast-Variable zugeordnet wird, die nur Einsen enthält. Mit

$$\hat{\eta}^T = (\hat{\eta}_{11}, \hat{\eta}_{12}, \hat{\eta}_{13}, \hat{\eta}_{21}, \hat{\eta}_{22}, \hat{\eta}_{23}) \text{ und} \\ (\mathbf{log} \hat{\mathbf{m}})^T = (\log \hat{m}_{11}, \hat{m}_{12}, \hat{m}_{13}, \hat{m}_{21}, \hat{m}_{22}, \hat{m}_{23})$$

lautet die geschätzte link Funktion gemäß (5.49):

$$\hat{\eta} = \mathbf{log} \hat{\mathbf{m}} = \mathbf{Xb}$$

Die letztendliche Spezifikation der Design-Matrix hängt vom Parametervektor \mathbf{b} ab, dessen Komponenten wiederum davon bestimmt werden, ob ein log-lineares Modell bei Gültigkeit der Nullhypothese oder das saturierte Modell geschätzt wird.

Die Nullhypothese H_0 lautet: Die Variablen Wahlbeteiligung und politisches Interesse wirken unabhängig auf die Zellwahrscheinlichkeiten.

5.7.1 Modell unter der Nullhypothese

Der Vektor \mathbf{b} und die Design-Matrix \mathbf{X} sind für das Modell unter H_0 konkret für dieses Beispiel:

$$\mathbf{b}^T = (b \ b_{A(1)} \ b_{B(1)} \ b_{B(2)}),$$

bei Dummy-Kodierung

$$X = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

bei Effekt-Kodierung

$$X = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 0 \\ 1 & -1 & 0 & 1 \\ 1 & -1 & -1 & -1 \end{pmatrix}$$

Damit ergeben sich die link Funktionen:

- bei Dummy-Kodierung:

Wahlbeteiligung pol. Interesse

1	1	$\log \hat{m}_{11} = b \cdot 1 + b_{A(1)} \cdot 1 + b_{B(1)} \cdot 1 + b_{B(2)} \cdot 0$
1	2	$\log \hat{m}_{12} = b \cdot 1 + b_{A(1)} \cdot 1 + b_{B(1)} \cdot 0 + b_{B(2)} \cdot 1$
1	3	$\log \hat{m}_{13} = b \cdot 1 + b_{A(1)} \cdot 1 + b_{B(1)} \cdot 0 + b_{B(2)} \cdot 0$
2	1	$\log \hat{m}_{21} = b \cdot 1 + b_{A(1)} \cdot 0 + b_{B(1)} \cdot 1 + b_{B(2)} \cdot 0$
2	2	$\log \hat{m}_{22} = b \cdot 1 + b_{A(1)} \cdot 0 + b_{B(1)} \cdot 0 + b_{B(2)} \cdot 1$
2	3	$\log \hat{m}_{23} = b \cdot 1 + b_{A(1)} \cdot 0 + b_{B(1)} \cdot 0 + b_{B(2)} \cdot 0$

- bei Effekt-Kodierung:

Wahlbeteiligung pol. Interesse

1	1	$\log \hat{m}_{11} = b \cdot 1 + b_{A(1)} \cdot 1 + b_{B(1)} \cdot 1 + b_{B(2)} \cdot 0$
1	2	$\log \hat{m}_{12} = b \cdot 1 + b_{A(1)} \cdot 1 + b_{B(1)} \cdot 0 + b_{B(2)} \cdot 1$
1	3	$\log \hat{m}_{13} = b \cdot 1 + b_{A(1)} \cdot 1 + b_{B(1)} \cdot (-1) + b_{B(2)} \cdot (-1)$
2	1	$\log \hat{m}_{21} = b \cdot 1 + b_{A(1)} \cdot (-1) + b_{B(1)} \cdot 1 + b_{B(2)} \cdot 0$
2	2	$\log \hat{m}_{22} = b \cdot 1 + b_{A(1)} \cdot (-1) + b_{B(1)} \cdot 0 + b_{B(2)} \cdot 1$
2	3	$\log \hat{m}_{23} = b \cdot 1 + b_{A(1)} \cdot (-1) + b_{B(1)} \cdot (-1) + b_{B(2)} \cdot (-1)$

Für das Modell unter H_0 erhält man bei Verwendung der **Dummy-Kodierung** nach (5.78) bis (5.80):

$$b = 6,536692 + 6,988413 - 8,164226 = 5,360879$$

$$b_{A(1)} = 7,945555 - 6,536692 = 1,408863$$

$$b_{A(2)} = 0$$

$$b_{B(1)} = 6,829794 - 6,988413 = -0,158619$$

$$b_{B(2)} = 7,315884 - 6,988413 = 0,327471$$

$$b_{B(3)} = 0$$

$b_{A(2)}$ und $b_{B(3)}$ sind durch die corner-point-Restriktion bei Wahl der letzten Kategorie als Referenzkategorie gleich Null gesetzt.

Die Konstante entspricht dem Logarithmus der geschätzten erwarteten Häufigkeit für diejenige Zelle, bei der alle Kontrast-Variablen den Wert Null aufweisen. Im Beispiel ist dies die letzte Zelle der Kontingenztabelle für Wahlbeteiligung = nein und politisches Interesse = wenig. Die geschätzte erwartete Zellhäufigkeit dieser Zelle ergibt sich zu $\exp(5,3609) = 212,91$.

Bei der Dummy-Kodierung erfolgt die Interpretation der Parameter der Kategorien der Variablen A bzw. B stets zur Referenzkategorie.

Der Parameter $b_{A(1)} = 1,4089$ bezieht sich auf die erste Kategorie “ja“ der Variablen Wahlbeteiligung. Dieser Parameter besagt, um welchen Wert die logarithmierten erwarteten Zellhäufigkeiten ansteigen, wenn statt der Referenzkategorie “nein“ die erste Kategorie der Wahlbeteiligung zutrifft. Hierbei handelt es sich um einen partiellen Effekt der ersten Kategorie der Variablen Wahlbeteiligung bei Kontrolle der Kategorien der Variablen politisches Interesse. $\exp(b_{A(1)})$ gibt dann für eine gegebene Kategorie von B (politisches Interesse) an, mit welchem Faktor die erwartete Zellhäufigkeit beim Übergang von den Nichtwählern (Referenzkategorie) zu den Wählern multipliziert werden muss. Es werden rund 4,1 mal mehr Befragte erwartet, die zur Wahl gehen, als Befragte, die nicht wählen ($\exp(1,4089) = 4,091452$). So erhält man z.B. für die Befragten mit wenig politischem Interesse (Referenzkategorie von B) beim Übergang von den Nichtwählern zu den Wählern eine erwartete Zellhäufigkeit von rund $\exp(5,3609 + 1,4089) = \exp(5,3609) \cdot \exp(1,4089) = 212,916 \cdot 4,091452 = 871,1$.

Der Parameter $b_{(1)} = -0,1586$ bezieht sich auf die erste Kategorie “stark“ der Variablen politisches Interesse. Dieser Parameter besagt, um welchen Wert die logarithmierten erwarteten Zellhäufigkeiten sinken, wenn statt der Referenzkategorie “wenig“ die erste

Kategorie der Variablen politisches Interesse zutrifft. Hierbei handelt es sich um einen partiellen Effekt der ersten Kategorie der Variablen politisches Interesse bei Kontrolle der Kategorien der Variablen Wahlbeteiligung. $\exp(b_{b(1)})$ gibt für eine gegebene Kategorie von A (Wahlbeteiligung) an, mit welchem Faktor die erwartete Zellhäufigkeit beim Übergang von den politisch wenig Interessierten (Referenzkategorie) zu den politisch stark Interessierten multipliziert werden muss ($\exp(-0,1586) = 0,85334$). So erhält man z.B. für die Nichtwähler (Referenzkategorie von A) beim Übergang von den politisch wenig Interessierten zu den politisch stark Interessierten eine erwartete Zellhäufigkeit von rund:

$$\exp(5,3609 - 0,1586) = \exp(5,3609) \cdot \exp(-0,1586) = 212,916 \cdot 0,85334 = 181,68.$$

Der Parameter $b_{B(2)}$ bezieht sich auf die zweite Kategorie “mittel“ der Variablen politisches Interesse. Der Parameter besagt, um welchen Wert die logarithmierten erwarteten Zellhäufigkeiten ansteigen, wenn statt der Referenzkategorie “wenig“ die zweite Kategorie der Variablen politisches Interesse zutrifft. Hierbei handelt es sich um einen partiellen Effekt der zweiten Kategorie der Variablen politisches Interesse bei Kontrolle der Kategorien der Variablen Wahlbeteiligung. $\exp(b_{b(2)})$ gibt für eine gegebene Kategorie von A (Wahlbeteiligung) an, mit welchem Faktor die erwartete Zellhäufigkeit beim Übergang von den politisch wenig Interessierten (Referenzkategorie) zu den politisch mittelmäßig Interessierten multipliziert werden muss ($\exp(0,3275) = 1,38749$). So erhält man z.B. für die Nichtwähler (Referenzkategorie von A) beim Übergang von den politisch wenig Interessierten zu den politisch mittelmäßig Interessierten eine erwartete Zellhäufigkeit von rund:

$$\exp(5,3609 + 0,3275) = \exp(5,3609) \cdot \exp(0,3275) = 212,916 \cdot 1,38749 = 295,41.$$

Einsetzen der geschätzten Parameter in die oben angegebenen link Funktionen ergibt nachstehende Schätzwerte der link Funktionen:

$$\log \hat{m}_{11} = 5,360879 + 1,408863 \cdot 1 - 0,158619 \cdot 1 + 0,327471 \cdot 0 = 6,611123$$

$$\log \hat{m}_{12} = 5,360879 + 1,408863 \cdot 1 - 0,158619 \cdot 0 + 0,327471 \cdot 1 = 7,097213$$

$$\log \hat{m}_{13} = 5,360879 + 1,408863 \cdot 1 - 0,158619 \cdot 0 + 0,327471 \cdot 0 = 6,769742$$

$$\log \hat{m}_{21} = 5,360879 + 1,408863 \cdot 0 - 0,158619 \cdot 1 + 0,327471 \cdot 0 = 5,20226$$

$$\log \hat{m}_{22} = 5,360879 + 1,408863 \cdot 0 - 0,158619 \cdot 0 + 0,327471 \cdot 1 = 5,68835$$

$$\log \hat{m}_{23} = 5,360879 + 1,408863 \cdot 0 - 0,158619 \cdot 0 + 0,327471 \cdot 0 = 5,360879$$

Für das Modell unter H_0 erhält man bei Verwendung der **Effekt-Kodierung** nach (5.71) bis (5.73):

$$b = 14,482247/2 + 21,134091/3 - 8,164226 = 6,1215945$$

$$b_{A(1)} = 7,945555 - 14,482247/2 = 0,7044315$$

$$b_{A(2)} = 6,536692 - 14,482247/2 = -0,7044315$$

$$b_{B(1)} = 6,829794 - 21,134091/3 = -0,214903$$

$$b_{B(2)} = 7,315884 - 21,134091/3 = 0,271187$$

$$b_{B(3)} = 6,988413 - 21,134091/3 = -0,056284$$

$b_{A(2)}$ und $b_{B(3)}$ können auch über die Restriktionen (5.33) ermittelt werden.

Die Konstante $b = 6,1215945$ gibt den mittleren Wert der logarithmierten erwarteten Zellhäufigkeiten wieder. Werden die link Funktionen über alle Zellen aufsummiert, bleibt auf der rechten Seite der Gleichung nur die Summe der Konstanten (= sechsfache der Konstanten) übrig, so dass resultiert: $\sum_j \sum_k \log \hat{m}_{jk}/6 = 36,7296/6 = 6,1216$. Die im Mittel erwartete Zellhäufigkeit ist damit: $\exp(6,1215945) = 455,59$.

Aus der Tatsache, dass die Summe der Werte einer Kontrast-Variablen Null ist, folgt für die Interpretation des zugehörigen Parameters, dass dieser die Abweichung vom Mittelwert erfasst.

Der Parameter $b_{A(1)} = 0,7044315$ für die erste Kategorie (ja) der Variablen Wahlbeteiligung besagt, dass die logarithmierten erwarteten Häufigkeiten der Zellen, bei denen die Variable Wahlbeteiligung die Kategorie "ja" aufweist, um 0,7044315 über dem Durchschnitt der logarithmierten erwarteten Zellhäufigkeiten liegen bzw. die erwarteten Häufigkeiten dieser Zellen rund das 2-fache des Durchschnitts der erwarteten Zellhäufigkeiten sind, wobei jedoch für die Variable politisches Interesse kontrolliert wird:

$$6,1215945 + 0,7044315 = 6,826026; \quad \exp(0,7044315) = 2,0227$$

$$\begin{aligned}\exp(6, 1215945 + 0, 7044315) &= \exp(6, 1215945) \cdot \exp(0, 7044315) \\ &= 455, 59 \cdot 2, 0227 = 921, 52.\end{aligned}$$

Der Parameter $b_{A(2)} = -0, 7044315$ für die zweite Kategorie (nein) der Variablen Wahlbeteiligung besagt, dass die logarithmierten erwarteten Häufigkeiten der Zellen, bei denen die Variable Wahlbeteiligung die Kategorie “nein“ aufweist, um 0,7044315 unter dem Durchschnitt der logarithmierten erwarteten Zellhäufigkeiten liegen bzw. die erwarteten Zellhäufigkeiten dieser Zellen nur rund die Hälfte des Durchschnitts der erwarteten Zellhäufigkeiten betragen, wobei für die Variable politisches Interesse kontrolliert wird:

$$\begin{aligned}6, 1215945 - 0, 7044315 &= 5, 417163; & \exp(-0, 7044315) &= 0, 49439; \\ \exp(6, 1215945 - 0, 7044315) &= \exp(6, 1215945) \cdot \exp(-0, 7044315) \\ &= 455, 59 \cdot 0, 49439 = 225, 24.\end{aligned}$$

Der Parameter $b_{B(1)} = -0, 214903$ für die erste Kategorie (stark) der Variablen politisches Interesse besagt, dass die logarithmierten erwarteten Häufigkeiten der Zellen, bei denen die Variable politisches Interesse die Kategorie “stark“ aufweist, um 0,214903 unter dem Durchschnitt der logarithmierten erwarteten Zellhäufigkeiten liegen bzw. die erwarteten Zellhäufigkeiten dieser Zellen nur rund 0,8 des Durchschnitts der erwarteten Zellhäufigkeiten betragen, wobei für die Variable Wahlbeteiligung kontrolliert wird:

$$\begin{aligned}6, 1215945 - 0, 214903 &= 5, 9066915; & \exp(-0, 214903) &= 0, 80662; \\ \exp(6, 1215945 - 0, 214903) &= \exp(6, 1215945) \cdot \exp(-0, 214903) \\ &= 455, 59 \cdot 0, 80662 = 367, 49.\end{aligned}$$

Der Parameter $b_{B(2)} = 0, 271187$ für die zweite Kategorie (mittel) der Variablen politisches Interesse besagt, dass die logarithmierten erwarteten Häufigkeiten der Zellen, bei denen die Variable politisches Interesse die Kategorie “mittel“ aufweist, um 0,271187 über dem Durchschnitt der logarithmierten erwarteten Zellhäufigkeiten liegen bzw. die erwarteten Zellhäufigkeiten dieser Zellen das 1,3-fache des Durchschnitts der erwarteten Zellhäufigkeiten betragen, wobei für die Variable Wahlbeteiligung kontrolliert wird:

$$6, 1215945 + 0, 271187 = 6, 3927815; \quad \exp(0, 271187) = 1, 3115;$$

$$\begin{aligned}\exp(6,1215945 + 0,271187) &= \exp(6,1215945) \cdot \exp(0,271187) \\ &= 455,59 \cdot 1,3115 = 597,52.\end{aligned}$$

Der Parameter $b_{B(3)} = -0,056284$ für die dritte Kategorie (wenig) der Variablen politisches Interesse besagt, dass die logarithmierten erwarteten Häufigkeiten der Zellen, bei denen die Variable politisches Interesse die Kategorie “wenig“ aufweist, um 0,056284 unter dem Durchschnitt der logarithmierten erwarteten Zellhäufigkeiten liegen bzw. die erwarteten Zellhäufigkeiten dieser Zellen nur rund 0,95 des Durchschnitts der erwarteten Zellhäufigkeiten betragen, wobei für die Variable Wahlbeteiligung kontrolliert wird:

$$\begin{aligned}6,1215945 - 0,056284 &= 6,0653105; & \exp(-0,056284) &= 0,94527; \\ \exp(6,1215945 - 0,056284) &= \exp(6,1215945) \cdot \exp(-0,056284) \\ &= 455,59 \cdot 0,94527 = 430,66.\end{aligned}$$

Einsetzen der geschätzten Parameter in die oben angegebenen link Funktionen ergibt nachstehende Schätzwerte der link Funktionen:

$$\begin{aligned}\log \hat{m}_{11} &= 6,1215945 + 0,7044315 \cdot 1 - 0,214903 \cdot 1 + 0,271187 \cdot 0 = 6,611123 \\ \log \hat{m}_{12} &= 6,1215945 + 0,7044315 \cdot 1 - 0,214903 \cdot 0 + 0,271187 \cdot 1 = 7,097213 \\ \log \hat{m}_{13} &= 6,1215945 + 0,7044315 \cdot 1 - 0,214903 \cdot (-1) + 0,271187 \cdot (-1) = 6,769742 \\ \log \hat{m}_{21} &= 6,1215945 + 0,7044315 \cdot (-1) - 0,214903 \cdot 1 + 0,271187 \cdot 0 = 5,20226 \\ \log \hat{m}_{22} &= 6,1215945 + 0,7044315 \cdot (-1) - 0,214903 \cdot 0 + 0,271187 \cdot 1 = 5,68835 \\ \log \hat{m}_{23} &= 6,1215945 + 0,7044315 \cdot (-1) - 0,214903 \cdot (-1) + 0,271187 \cdot (-1) = 5,360879\end{aligned}$$

Der Vergleich der log-linearen Modelle mit Dummy-Kodierung und Effekt-Kodierung zeigt gleiche Ergebnisse für die erwarteten Zellhäufigkeiten und damit auch für die Residuen sowie identische Werte für die Goodness-of-Fit Statistiken. Das liegt darin begründet, dass es sich bei den Modellen mit Dummy-Kodierung bzw. Effekt-Kodierung um Reparametrisierungen der gleichen Aussagen zur Struktur der analysierten Kontingenztafel handelt. Zu Unterschieden würde es nur dann kommen, wenn Modelle mit unterschiedlichen Hypothesen verglichen würden, z.B. das Unabhängigkeitsmodell und das saturierte Modell.

Tabelle 5.6: Geschätzte erwartete Zellhäufigkeiten für das Modell unter H_0

Variable A (Wahlbeteiligung)	Variable B (politisches Interesse)			Randver- teilung A
	B_1 (stark)	B_2 (mittel)	B_3 (wenig)	
A_1 (ja)	743,32	1208,59	871,09	2823
A_2 (nein)	181,68	295,41	212,91	690
Randverteilung B	925	1504	1084	3513

Im Vergleich zur Tabelle 5.4 ist ersichtlich, dass die beobachteten und erwarteten Randhäufigkeiten übereinstimmen.

Die Zellbeiträge zu Pearson's Chi-Quadrat-Teststatistik (5.83) sind in der Tabelle 5.7 enthalten: $(y_{jk} - \hat{m}_{jk}^0)^2 / \hat{m}_{jk}^0$.

Tabelle 5.7: Zellbeiträge zu Pearson's Chi-Quadrat-Teststatistik

Variable A (Wahlbeteiligung)	Variable B (politisches Interesse)			Summe
	B_1 (stark)	B_2 (mittel)	B_3 (wenig)	
A_1 (ja)	8,38	2,54	20,64	
A_2 (nein)	34,07	10,39	84,45	
Summe				160,43

Da für das Modell unter H_0 die vier nicht redundanten Parameter b , $b_{A(1)}$, $b_{B(1)}$ und $b_{B(2)}$ geschätzt werden müssen, beträgt die Anzahl der Freiheitsgrade: $f = 2 \cdot 3 - (1 + 1 + 2) = 2$. Für ein vorgegebenes Signifikanzniveau von $\alpha = 0,05$ findet man in der Tafel der Chi-Quadrat-Verteilung $\chi_{0,95;2}^2 = 5,99$. Wegen $\chi^2 = 160,43 > \chi_{0,95;2}^2 = 5,99$ wird die Nullhypothese, dass Abweichungen zwischen den beobachteten und den erwarteten Zellhäufigkeiten nur durch zufällige Stichprobenschwankungen hervorgerufen werden, verworfen. Es gibt also gute Gründe, an der Übereinstimmung von Modell und Daten zu zweifeln. Da die Nullhypothese des betrachteten Modells gleichzeitig die Unabhängigkeit der beiden Modellvariablen impliziert, ist diese auf dem vorgegebenen Signifikanzniveau abzulehnen.

Zum gleichen Ergebnis gelangt man über die Berechnung der Deviance nach (5.85).

Die Zellbeiträge zur Deviance

$$y_{jk} \log \left(\frac{ny_{jk}}{y_{j+}y_{+k}} \right)$$

sind in der Tabelle 5.8 enthalten.

Somit ist: $D = 2 \cdot 77,3548 = 154,709$. Da D unter H_0 ebenfalls asymptotisch chi-quadratverteilt ist, ergibt sich die gleiche Testbeantwortung.

Tabelle 5.8: Zellbeiträge zur Deviance

Variable A (Wahlbeteiligung)	Variable B (politisches Interesse)			Summe
	B_1 (stark)	B_2 (mittel)	B_3 (wenig)	
A_1 (ja)	82,7071	56,6566	-123,1933	
A_2 (nein)	-58,4557	-49,8505	169,4906	
Summe				77,3548

Sowohl bei den Zellbeiträgen zu Pearson's Chi-Quadrat-Teststatistik als auch bei den Zellbeiträgen zur Deviance wird deutlich, dass Zelle 23, d.h. (Wahlbeteiligung = nein) \cap (politisches Interesse = wenig), die größte Abweichung zwischen beobachteter und erwarteter Zellhäufigkeit aufweist.

Unter SPSS für Windows Release 10.0.7 wird für die Schätzung des log-linearen Modells mit Dummy-Kodierung die Prozedur "General Loglinear Analysis"⁴² verwendet, die unmittelbar über Dialogfelder aufgerufen werden kann. Eine Veränderung der Kodierung ist bei dieser Prozedur nur über vorab definierte und in der Datei befindliche Kontrast-Variablen möglich.

Da für dieses Beispiel das multinomiale Stichprobenmodell unterstellt wird, ist bei Distribution of Cell Counts auf Multinomial zu entscheiden.

Um das Modell unter H_0 zu schätzen, wird über das Dialogfeld "General Loglinear Analysis: Model:" ein Modell spezifiziert, das nur die Haupteffekte (main effects) der Variablen Wahlbeteiligung und politisches Interesse enthält.

Über das Dialogfeld "General Loglinear Analysis: Options:" kann der Output gestaltet

⁴²Siehe SPSS Advanced Models 9.0 (1999).

werden. In SPSS wird standardmäßig der Wert 0,5 zu allen Zellhäufigkeiten addiert, um Zellhäufigkeiten von Null zu vermeiden. Da in diesem Beispiel keine Zelle unbesetzt ist, wird in diesem Dialogfeld der Delta-Wert auf Null gesetzt, wodurch die zu analysierenden Zellhäufigkeiten unverändert bleiben.

Alle oben diskutierten Schätzergebnisse sind in diesem Output enthalten. Darüber hinaus ist noch ersichtlich, dass die Parameter $b_{A(1)}$, $b_{B(1)}$ und $b_{B(2)}$ der Haupteffekte der beiden Variablen signifikant zum 5% Niveau sind, da die 95% Konfidenzintervalle die Null nicht einschließen.

SPSS-Output für das Modell unter H_0 mit Dummy-Kodierung:

```

-----
                        GENERAL LOGLINEAR ANALYSIS
-----
Data Information
  3513 cases are accepted.
    5 cases are rejected because of missing data.
  3513 weighted cases will be used in the analysis.
    6 cells are defined.
    0 structural zeros are imposed by design.
    0 sampling zeros are encountered.

Variable Information
Factor      Levels      Value
BET          2          Wahlbeteiligung, letzte Bundestagswahl
                1 JA
                2 NEIN
POL_INT       3          Politisches Interesse
                1 stark
                2 mittel
                3 wenig

Model and Design Information
  Model: Multinomial
  Design: Constant + BET + POL_INT

Correspondence Between Parameters and Terms of the Design
Parameter  Aliased  Term
          1          Constant
          2          [BET = 1]
          3          [BET = 2]
          4          [POL_INT = 1]
          5          [POL_INT = 2]
          6          [POL_INT = 3]

Note: 'x' indicates an aliased (or a redundant) parameter.
      These parameters are set to zero.

Design Matrix

```

Factor	Value	Cell Structure	Parameter			
			1	2	4	5
BET	JA					
POL_INT	stark	1,000	1	1	1	0
POL_INT	mittel	1,000	1	1	0	1
POL_INT	wenig	1,000	1	1	0	0
BET	NEIN					
POL_INT	stark	1,000	1	0	1	0
POL_INT	mittel	1,000	1	0	0	1
POL_INT	wenig	1,000	1	0	0	0

Convergence Information

Maximum number of iterations: 20

Relative difference tolerance: ,001

Final relative difference: 8,61220E-07

Maximum likelihood estimation converged at iteration 4.

Table Information

Factor	Value	Observed		Expected	
		Count	%	Count	%
BET	JA				
POL_INT	stark	822,00 (23,40)		743,32 (21,16)	
POL_INT	mittel	1264,00 (35,98)		1208,59 (34,40)	
POL_INT	wenig	737,00 (20,98)		871,09 (24,80)	
BET	NEIN				
POL_INT	stark	103,00 (2,93)		181,68 (5,17)	
POL_INT	mittel	240,00 (6,83)		295,41 (8,41)	
POL_INT	wenig	347,00 (9,88)		212,91 (6,06)	

Table Information

Factor	Value	Resid.	Adj.	Dev.
			Resid.	Resid.
BET	JA			
POL_INT	stark	78,68	7,59	12,86
POL_INT	mittel	55,41	4,76	10,64
POL_INT	wenig	-134,09	-12,33	-15,70
BET	NEIN			
POL_INT	stark	-78,68	-7,59	-10,81
POL_INT	mittel	-55,41	-4,76	-9,99
POL_INT	wenig	134,09	12,33	18,41

Goodness-of-fit Statistics

	Chi-Square	DF	Sig.
Likelihood Ratio	154,7098	2	3,E-34
Pearson	160,4224	2	2,E-35

Parameter Estimates

Constant	Estimate
1	5,3609

Note: Constant is not a parameter under multinomial assumption.
Therefore, standard errors are not calculated.

Parameter	Estimate	SE	Z-value	Asymptotic 95% CI	
				Lower	Upper
2	1,4089	,0425	33,17	1,33	1,49

5 Modellierung multinomialer Daten (log-lineare Modelle)

3	,0000	,	,	,	,
4	-,1586	,0448	-3,54	-,25	-,07
5	,3275	,0398	8,22	,25	,41
6	,0000	,	,	,	,

Für die Schätzung des log-linearen Modells mit Effekt-Kodierung (Deviation) wird die Prozedur “Loglinear Analysis“⁴³ eingesetzt, die jedoch nur über die Syntax verfügbar ist.

Für dieses Beispiel ist folgendes Programm zu verwenden:

```
LOGLINEAR
bet(1 2) pol_int(1 3)
/CONTRAST (bet)=Deviation /CONTRAST (pol_int)=Deviation
/PRINT=FREQ RESID DESIGN ESTIM
CRITERIA ITERATION (20) CONVERGE(.001) DELTA(0)
/DESIGN bet pol_int.
```

SPSS-Output für das Modell unter H_0 mit Effekt-Kodierung:

```
***** LOG LINEAR ANALYSIS *****
Data Information
  3513 cases are accepted.
    0 cases rejected because of out-of-range factor values.
    5 cases rejected because of missing data.
  3513 weighted cases will be used in the analysis.

FACTOR Information
Factor      Level   Label
BET          2     Wahlbeteiligung, letzte Bundestagswahl
POL_INT      3     Politisches Interesse

DESIGN Information
  1 Design/Model will be processed.
Correspondence Between Effects and Columns of Design/Model 1
Starting   Ending
Column     Column   Effect Name
  1         1       BET
  2         3       POL_INT

Design Matrix
1-BET 2-POL_INT
Factor      1  2          1          2          3
  1  1      1,000    1,000    ,000
  1  2      1,000    ,000    1,000
```

⁴³Siehe dazu unter SPSS im Help-System unter dem Indexstichwort “Loglinear command syntax“.

```

1  3      1,000  -1,000  -1,000
2  1     -1,000   1,000   ,000
2  2     -1,000   ,000   1,000
2  3     -1,000  -1,000  -1,000
*** ML converged at iteration 3.
    Maximum difference between successive iterations =   ,00097.

Observed, Expected Frequencies and Residuals
    Factor      Code      OBS. count & PCT.  EXP. count & PCT.
BET      JA
POL_INT  stark      822,00 (23,40)      743,32 (21,16)
POL_INT  mittel     1264,00 (35,98)     1208,59 (34,40)
POL_INT  wenig      737,00 (20,98)      871,09 (24,80)

BET      NEIN
POL_INT  stark      103,00 ( 2,93)      181,68 ( 5,17)
POL_INT  mittel     240,00 ( 6,83)      295,41 ( 8,41)
POL_INT  wenig      347,00 ( 9,88)      212,91 ( 6,06)

Observed, Expected Frequencies and Residuals (cont.)
    Factor      Code      Residual  Std. Resid.  Adj. Resid.
BET      JA
POL_INT  stark      78,6825      2,8860      7,5869
POL_INT  mittel     55,4059      1,5937      4,7553
POL_INT  wenig     -134,0878     -4,5432     -12,3281

BET      NEIN
POL_INT  stark     -78,6825     -5,8374     -7,5869
POL_INT  mittel    -55,4059     -3,2236     -4,7553
POL_INT  wenig     134,0878      9,1894     12,3281

Goodness-of-fit Statistics
    Likelihood Ratio Chi-Square = 154,70978  DF = 2  P = ,000
    Pearson Chi-Square = 160,42225  DF = 2  P = ,000

Estimates for Parameters
BET
Parameter      Coeff.      Std. Err.      Z-Value  Lower 95 CI  Upper 95 CI
1  ,7044313795      ,02123      33,17491      ,66281      ,74605
POL_INT
Parameter      Coeff.      Std. Err.      Z-Value  Lower 95 CI  Upper 95 CI
2  -,214903071      ,02563     -8,38507     -,26514     -,16467
3  ,2711866966      ,02276     11,91377      ,22657      ,31580

```

Bei dieser Prozedur ist die Konstante nicht im Output enthalten, sie muss nach (5.71) berechnet werden. Die im Output enthaltenen Parameter entsprechen: Parameter 1 = $b_{A(1)}$, Parameter 2 = $b_{B(1)}$ und Parameter 3 = $b_{B(2)}$. Die Parameter $b_{A(2)}$ und $b_{B(3)}$ sind gemäß (5.34) zu ermitteln oder indem ein erneuter Durchlauf der Prozedur “Loglinear Analysis“ vorgenommen wird, wobei für beide Variablen als Referenzkategorie die erste Kategorie gewählt wird. Auch für das Modell unter H_0 mit Effekt Kodierung sind alle oben diskutierten Schätzergebnisse im Output enthalten. Auch bei dieser Modellschätzung sind die

Parameter $b_{A(1)}$, $b_{B(1)}$ und $b_{B(2)}$ der Haupteffekte der beiden Variablen signifikant zum 5% Niveau, da die 95% Konfidenzintervalle die Null nicht einschließen.

5.7.2 Satturiertes Modell

Da das Modell unter der Nullhypothese, das die Unabhängigkeit der Variablen Wahlbeteiligung und politisches Interesse beinhaltet, zu keiner ausreichenden Anpassung an die Daten führte, wird nun das saturierte Modell geschätzt.

Der Vektor \mathbf{b} und die Design-Matrix \mathbf{X} sind für das saturierte Modell:

$$\mathbf{b}^T = (b \quad b_{A(1)} \quad b_{B(1)} \quad b_{B(2)} \quad b_{AB(11)} \quad b_{AB(12)}),$$

bei Dummy-Kodierung	bei Effekt-Kodierung
$X = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$	$X = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 0 & -1 & 0 \\ 1 & -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{pmatrix}$

Die Kontrast-Variablen für die Interaktionsparameter erhält man durch Multiplikation der Kontrast-Variablen, die zu den jeweiligen Haupteffekten gehören. So ergibt sich z.B. die Kontrast-Variable für $b_{AB(11)}$ durch Multiplikation der Kontrast-Variablen von $b_{A(1)}$ (zweiter Spaltenvektor) und der Kontrast-Variablen von $b_{B(1)}$ (dritter Spaltenvektor).

Damit ergeben sich die link Funktionen:

- bei Dummy-Kodierung:

$$\begin{aligned} \log \hat{m}_{11} &= b \cdot 1 + b_{A(1)} \cdot 1 + b_{B(1)} \cdot 1 + b_{B(2)} \cdot 0 + b_{AB(11)} \cdot 1 + b_{AB(12)} \cdot 0 \\ \log \hat{m}_{12} &= b \cdot 1 + b_{A(1)} \cdot 1 + b_{B(1)} \cdot 0 + b_{B(2)} \cdot 1 + b_{AB(11)} \cdot 0 + b_{AB(12)} \cdot 1 \\ \log \hat{m}_{13} &= b \cdot 1 + b_{A(1)} \cdot 1 + b_{B(1)} \cdot 0 + b_{B(2)} \cdot 0 + b_{AB(11)} \cdot 0 + b_{AB(12)} \cdot 0 \\ \log \hat{m}_{21} &= b \cdot 1 + b_{A(1)} \cdot 0 + b_{B(1)} \cdot 1 + b_{B(2)} \cdot 0 + b_{AB(11)} \cdot 0 + b_{AB(12)} \cdot 0 \\ \log \hat{m}_{22} &= b \cdot 1 + b_{A(1)} \cdot 0 + b_{B(1)} \cdot 0 + b_{B(2)} \cdot 1 + b_{AB(11)} \cdot 0 + b_{AB(12)} \cdot 0 \\ \log \hat{m}_{23} &= b \cdot 1 + b_{A(1)} \cdot 0 + b_{B(1)} \cdot 0 + b_{B(2)} \cdot 0 + b_{AB(11)} \cdot 0 + b_{AB(12)} \cdot 0 \end{aligned}$$

- bei Effekt-Kodierung:

$$\begin{aligned}
\log \hat{m}_{11} &= b \cdot 1 + b_{A(1)} \cdot 1 + b_{B(1)} \cdot 1 + b_{B(2)} \cdot 0 + b_{AB(11)} \cdot 1 + b_{AB(12)} \cdot 0 \\
\log \hat{m}_{12} &= b \cdot 1 + b_{A(1)} \cdot 1 + b_{B(1)} \cdot 0 + b_{B(2)} \cdot 1 + b_{AB(11)} \cdot 0 + b_{AB(12)} \cdot 1 \\
\log \hat{m}_{13} &= b \cdot 1 + b_{A(1)} \cdot 1 + b_{B(1)} \cdot (-1) + b_{B(2)} \cdot (-1) + b_{AB(11)} \cdot (-1) + b_{AB(12)} \cdot (-1) \\
\log \hat{m}_{21} &= b \cdot 1 + b_{A(1)} \cdot (-1) + b_{B(1)} \cdot 1 + b_{B(2)} \cdot 0 + b_{AB(11)} \cdot (-1) + b_{AB(12)} \cdot 0 \\
\log \hat{m}_{22} &= b \cdot 1 + b_{A(1)} \cdot (-1) + b_{B(1)} \cdot 0 + b_{B(2)} \cdot 1 + b_{AB(11)} \cdot 0 + b_{AB(12)} \cdot (-1) \\
\log \hat{m}_{23} &= b \cdot 1 + b_{A(1)} \cdot (-1) + b_{B(1)} \cdot (-1) + b_{B(2)} \cdot (-1) + b_{AB(11)} \cdot 1 + b_{AB(12)} \cdot 1
\end{aligned}$$

Für das saturierte Modell erhält man bei Verwendung der **Dummy-Kodierung** nach (5.74) bis 5.(77):

$$\begin{aligned}
b &= 5,849325 \\
b_{A(1)} &= 6,602588 - 5,849325 = 0,753263 \\
b_{A(2)} &= 0 \\
b_{B(1)} &= 4,634729 - 5,849325 = -1,214596 \\
b_{B(2)} &= 5,480639 - 5,849325 = -0,368686 \\
b_{B(3)} &= 0 \\
b_{AB(11)} &= 6,711740 - 6,602588 - (-1,214596) = 1,323748 \\
b_{AB(12)} &= 7,142037 - 6,602588 - (-0,368686) = 0,908135 \\
b_{AB(13)} &= 0, \quad b_{AB(21)} = 0, \quad b_{AB(22)} = 0, \quad b_{AB(23)} = 0
\end{aligned}$$

$b_{A(2)}$, $b_{B(3)}$, $b_{AB(13)}$, $b_{AB(21)}$, $b_{AB(22)}$ und $b_{AB(23)}$ sind durch die corner-point-Restriktion bei Wahl der letzten Kategorie als Referenzkategorie gleich Null gesetzt.

Einsetzen der geschätzten Parameter in die oben angegebenen link Funktionen ergibt nachstehende Schätzwerte der link Funktionen:

$$\begin{aligned}
\log \hat{m}_{11} &= 5,849325 + 0,753263 \cdot 1 - 1,214596 \cdot 1 - 0,368686 \cdot 0 + 1,323748 \cdot 1 \\
&\quad + 0,908135 \cdot 0 = 6,71174 \\
\log \hat{m}_{12} &= 5,849325 + 0,753263 \cdot 1 - 1,214596 \cdot 0 - 0,368686 \cdot 1 + 1,323748 \cdot 0 \\
&\quad + 0,908135 \cdot 1 = 7,142037
\end{aligned}$$

5 Modellierung multinomialer Daten (log-lineare Modelle)

$$\begin{aligned}\log \hat{m}_{13} &= 5,849325 + 0,753263 \cdot 1 - 1,214596 \cdot 0 - 0,368686 \cdot 0 + 1,323748 \cdot 0 \\ &\quad + 0,908135 \cdot 0 = 6,602588\end{aligned}$$

$$\begin{aligned}\log \hat{m}_{21} &= 5,849325 + 0,753263 \cdot 0 - 1,214596 \cdot 1 - 0,368686 \cdot 1 + 1,323748 \cdot 0 \\ &\quad + 0,908135 \cdot 0 = 4,634729\end{aligned}$$

$$\begin{aligned}\log \hat{m}_{22} &= 5,849325 + 0,753263 \cdot 0 - 1,214596 \cdot 0 - 0,368686 \cdot 1 + 1,323748 \cdot 0 \\ &\quad + 0,908135 \cdot 0 = 5,480639\end{aligned}$$

$$\begin{aligned}\log \hat{m}_{23} &= 5,849325 + 0,753263 \cdot 0 - 1,214596 \cdot 0 - 0,368686 \cdot 0 + 1,323748 \cdot 0 \\ &\quad + 0,908135 \cdot 0 = 5,849325\end{aligned}$$

Ein Vergleich der Parameter der Haupteffekte mit den entsprechenden Parameterwerten des Unabhängigkeitsmodells weist deutlich veränderte Werte aus, die zum Teil sogar andere Vorzeichen aufweisen. Änderungen der Parameterschätzungen sind in der Regel zu beobachten, wenn die zusätzlich spezifizierten Interaktionseffekte die Vorhersagen der logarithmierten Zelhäufigkeiten merklich verbessern können.

Die Konstante $b = 5,849325$ gibt den Logarithmus der beobachteten Häufigkeit für diejenige Zelle an, bei der alle Kontrast-Variablen den Wert Null aufweisen (Wahlbeteiligung = nein und politisches Interesse = wenig).

Der Parameter $b_{A(1)} = 0,753263$ bezieht sich auf die erste Kategorie "ja" der Variablen Wahlbeteiligung. Der positive Wert weist darauf hin, dass die beobachtete Anzahl der Wähler rund zweimal höher ist als die beobachtete Anzahl der Nichtwähler (Referenzkategorie): $\exp(0,753263) = 2,12392$. Mit

$$\exp(5,849325 + 0,753263) = \exp(5,849325) \cdot \exp(0,753263) = 347 \cdot 2,12392 = 737$$

ergibt sich die beobachtete Häufigkeit der Zelle 13 (Wahlbeteiligung = ja und politisches Interesse = wenig).

Der Parameter $b_{B(1)} = -1,214596$ bezieht sich auf die Kategorie "stark" der Variablen politisches Interesse und beinhaltet, dass die beobachtete Anzahl der politisch stark Interessierten nur rund 0,3 der beobachteten Anzahl der politisch wenig Interessierten beträgt: $\exp(-1,214596) = 0,296823$. Mit

$$\exp(5,849325 - 1,214596) = \exp(5,849325) \cdot \exp(-1,214596) = 347 \cdot 0,296823 = 103$$

ergibt sich die beobachtete Häufigkeit der Zelle 21 (Wahlbeteiligung = nein und politisches Interesse = stark).

Der Parameter $b_{B(2)} = -0,368686$ bezieht sich auf die Kategorie “mittel“ der Variablen politisches Interesse und beinhaltet, dass die beobachtete Anzahl der mittelmäßig politisch Interessierten nur rund 0,7 der beobachteten Anzahl der politisch wenig Interessierten beträgt: $\exp(-0,368686) = 0,691643$. Mit $\exp(5,849325 - 0,368686) = \exp(5,849325) \cdot \exp(-0,368686) = 347 \cdot 0,691643 = 240$ ergibt sich die beobachtete Häufigkeit der Zelle 22 (Wahlbeteiligung = nein und politisches Interesse = mittel).

Hier wird ein entscheidender Unterschied in der Berechnung und der Interpretation der Parameter der Haupteffekte des Modells unter H_0 und des saturierten Modells deutlich: Während die Parameter der Haupteffekte des Modells unter H_0 logarithmierte Verhältnisse (log odds) von Randhäufigkeiten beinhalten, sind es bei den Parametern der Haupteffekte des saturierten Modells logarithmierte Verhältnisse von beobachteten Zellhäufigkeiten, wobei für die zweite Variable kontrolliert wird. Die Interpretation der Parameter der Haupteffekte des Modells unter H_0 bezieht sich stets auf die unter H_0 erwarteten Zellhäufigkeiten, die Interpretation der Parameter der Haupteffekte des saturierten Modells dagegen stets auf die beobachteten Zellhäufigkeiten.

Dieses Grundmuster wird nun beim saturierten Modell durch die Interaktionseffekte modifiziert.

Der Interaktionsparameter $b_{AB(11)} = 1,323748$ bezieht sich auf Wahlbeteiligung = ja und politisches Interesse = stark. Er besagt, dass der Anstieg der beobachteten Zellhäufigkeit von der Referenzkategorie der politisch wenig Interessierten zu denjenigen, die stark interessiert sind, in der Gruppe der Wähler höher ist als in der Gruppe der Nichtwähler. Der Anstieg der beobachteten Zellhäufigkeit beim Wechsel von den Nichtwählern zu den Wählern in der Gruppe derjenigen, die politisch stark interessiert sind, ist um den Faktor $\exp(1,323748) = 3,7575$ höher als in der Gruppe derjenigen, die politisch wenig interessiert sind. Zu diesem Ergebnis gelangt man auch, wenn man das odds ratio aus der Kontingenztafel berechnet:

$$[822/103]/[737/347] = 3,7575.$$

Der Interaktionsparameter $b_{AB(12)} = 0,908135$ bezieht sich auf Wahlbeteiligung = ja und politisches Interesse = mittel. Er besagt, dass der Anstieg der beobachteten Zellhäufigkeit von der Referenzkategorie der politisch wenig Interessierten zu denjenigen, die mittelmäßig interessiert sind, in der Gruppe der Wähler höher ist als in der Gruppe der Nichtwähler. Der Anstieg der beobachteten Zellhäufigkeit beim Wechsel von den Nichtwählern zu den Wählern in der Gruppe derjenigen, die politisch mittelmäßig interessiert sind, ist um den Faktor $\exp(0,908135) = 2,4797$ höher als in der Gruppe derjenigen, die politisch wenig interessiert sind. Zu diesem Ergebnis gelangt man auch, wenn man das odds ratio aus der Kontingenztafel berechnet:

$$[1264/240]/[737/347] = 2,4797.$$

Zusammengenommen folgt aus dem Modell, dass das Verhältnis der Wähler zu Nichtwählern bei den politisch wenig Interessierten relativ am geringsten und bei denjenigen, die politisch stark interessiert sind, am höchsten ist.

Für das saturierte Modell erhält man bei Verwendung der **Effekt-Kodierung** nach (5.67) bis (5.70):

$$\begin{aligned} b &= 36,4210575/6 = 6,07017625 \\ b_{A(1)} &= 20,456365/3 - 6,07017625 = 0,748612 \\ b_{A(2)} &= 15,964693/3 - 6,07017625 = -0,748612 \\ b_{B(1)} &= 11,346469/2 - 6,07017625 = -0,396942 \\ b_{B(2)} &= 12,622676/2 - 6,07017625 = 0,241162 \\ b_{B(3)} &= 12,451913/2 - 6,07017625 = 0,15578 \\ b_{AB(11)} &= 6,711740 - (6,07017625 + 0,748612 - 0,396942) = 0,289894 \\ b_{AB(12)} &= 7,142037 - (6,07017625 + 0,748612 + 0,241162) = 0,0820868 \\ b_{AB(13)} &= 6,602588 - (6,07017625 + 0,748612 + 0,15578) = -0,371980 \\ b_{AB(21)} &= 4,634729 - (6,07017625 - 0,748612 - 0,396942) = -0,289894 \\ b_{AB(22)} &= 5,480639 - (6,07017625 - 0,748612 + 0,241162) = -0,0820868 \end{aligned}$$

$$b_{AB(23)} = 5,849325 - (6,07017625 - 0,748612 + 0,15578) = 0,371980$$

$b_{A(2)}$, $b_{B(3)}$, $b_{AB(13)}$, $b_{AB(21)}$, $b_{AB(22)}$ und $b_{AB(23)}$ können auch über die Restriktionen (5.45) ermittelt werden.

Einsetzen der geschätzten Parameter in die oben angegebenen link Funktionen ergibt wiederum die logarithmierten beobachteten Zellhäufigkeiten.

Die Konstante ist der Mittelwert der logarithmierten beobachteten Zellhäufigkeiten.

Der Parameter $b_{A(1)} = 0,748612$ für die erste Kategorie (ja) der Variablen Wahlbeteiligung besagt, dass die logarithmierten beobachteten Häufigkeiten der Zellen, bei denen die Variable Wahlbeteiligung die Kategorie “ja“ aufweist, um 0,748612 über dem Durchschnitt der logarithmierten beobachteten Zellhäufigkeiten liegen bzw. die beobachteten Zellhäufigkeiten dieser Zellen rund das 2-fache des Durchschnitts der beobachteten Zellhäufigkeiten sind, bei Kontrolle der Variablen politisches Interesse und Berücksichtigung der Interaktionseffekte: $\exp(0,748612) = 2,1141$.

Der Parameter $b_{A(2)} = -0,748612$ für die zweite Kategorie (nein) der Variablen Wahlbeteiligung besagt, dass die logarithmierten beobachteten Häufigkeiten der Zellen, bei denen die Variable Wahlbeteiligung die Kategorie “nein“ aufweist, um 0,748612 unter dem Durchschnitt der logarithmierten beobachteten Zellhäufigkeiten liegen bzw. die beobachteten Zellhäufigkeiten dieser Zellen nur rund die Hälfte des Durchschnitts der beobachteten Zellhäufigkeiten betragen, bei Kontrolle der Variablen politisches Interesse und Berücksichtigung der Interaktionseffekte: $\exp(-0,748612) = 0,4730$.

Aufgrund der zwei Haupteffekte der Wahlbeteiligung kann man schlussfolgern, dass es überdurchschnittlich viele Personen gibt, die bei der letzten Bundestagswahl (vor 1996) gewählt haben, und dass es dagegen unterdurchschnittlich viele Personen gibt, die nicht gewählt haben.

Der Parameter $b_{B(1)} = -0,396942$ für die erste Kategorie (stark) der Variablen politisches Interesse besagt, dass die logarithmierten beobachteten Häufigkeiten der Zellen, bei denen die Variable politisches Interesse die Kategorie “stark“ aufweist, um 0,396942 unter dem Durchschnitt der logarithmierten beobachteten Zellhäufigkeiten liegen bzw. die beobachteten Häufigkeiten dieser Zellen nur rund 0,7 des Durchschnitts der beob-

achteten Zellhäufigkeiten betragen, bei Kontrolle der Variablen Wahlbeteiligung und Berücksichtigung der Interaktionseffekte: $\exp(-0,396942) = 0,6724$.

Der Parameter $b_{B(2)} = 0,241162$ für die zweite Kategorie (mittel) der Variablen politisches Interesse besagt, dass die logarithmierten beobachteten Häufigkeiten der Zellen, bei denen die Variable politisches Interesse die Kategorie “mittel“ aufweist, um 0,241162 über dem Durchschnitt der logarithmierten beobachteten Zellhäufigkeiten liegen bzw. die beobachteten Häufigkeiten dieser Zellen rund das 1,3-fache des Durchschnitts der beobachteten Zellhäufigkeiten betragen, bei Kontrolle der Variablen Wahlbeteiligung und Berücksichtigung der Interaktionseffekte: $\exp(0,241162) = 1,2727$.

Der Parameter $b_{B(3)} = 0,15578$ für die dritte Kategorie (wenig) der Variablen politisches Interesse besagt, dass die logarithmierten beobachteten Häufigkeiten der Zellen, bei denen die Variable politisches Interesse die Kategorie “wenig“ aufweist, um 0,15578 über dem Durchschnitt der logarithmierten beobachteten Zellhäufigkeiten liegen bzw. die beobachteten Häufigkeiten dieser Zellen rund das 1,2-fache des Durchschnitts der beobachteten Zellhäufigkeiten betragen, bei Kontrolle der Variablen Wahlbeteiligung und Berücksichtigung der Interaktionseffekte: $\exp(0,15578) = 1,1686$.

Die drei Haupteffekte der Variablen politisches Interesse lassen den Schluss zu, dass es unterdurchschnittlich viele Personen gibt, die politisch stark interessiert sind, und dass es dagegen überdurchschnittlich viele Personen gibt, die mittleres bzw. wenig politisches Interesse zeigen.

Die Interaktionseffekte modifizieren diese Grundaussage in folgender Weise:

Der Parameter $b_{AB(11)} = 0,289894$ für Wahlbeteiligung = ja und politisches Interesse = stark besagt, dass durch die Interaktion der beiden Variablen die beobachtete Anzahl der Wähler mit starkem politischen Interesse auf rund das 1,3-fache der mittleren, nur aufgrund der Haupteffekte der beiden Variablen geschätzten Anzahl von Wählern mit starkem politischen Interesse ansteigt:

$$\begin{aligned} & \exp(b + b_{A(1)} + b_{B(1)}) \cdot \exp(b_{AB(11)}) \\ = & \exp(6,07017625 + 0,748612 - 0,396942) \cdot \exp(0,289894) \\ = & 615,1377654 \cdot 1,336285834 = 822. \end{aligned}$$

Analog ergibt der Parameter $b_{AB(21)} = -0,289894$ für Wahlbeteiligung = nein und politisches Interesse = stark, dass durch die Interaktion der beiden Variablen die beobachtete Anzahl der Nichtwähler mit starkem politischen Interesse nur rund das 0,75-fache der mittleren, nur aufgrund der Haupteffekte der beiden Variablen geschätzten Anzahl von Nichtwählern mit starkem politischen Interesse ausmacht:

$$\begin{aligned} & \exp(b + b_{A(2)} + b_{B(1)}) \cdot \exp(b_{AB(21)}) \\ = & \exp(6,07017625 - 0,748612 - 0,396942) \cdot \exp(-0,289894) \\ = & 137,63734 \cdot 0,748342887 = 103. \end{aligned}$$

Der Parameter $b_{AB(12)} = 0,0820868$ für Wahlbeteiligung = ja und politisches Interesse = mittel besagt, dass durch die Interaktion der beiden Variablen die beobachtete Anzahl der Wähler mit mittlerem politischen Interesse auf rund das 1,1-fache der mittleren, nur aufgrund der Haupteffekte der beiden Variablen geschätzten Anzahl von Wählern mit mittlerem politischen Interesse ansteigt:

$$\begin{aligned} & \exp(b + b_{A(1)} + b_{B(2)}) \cdot \exp(b_{AB(12)}) \\ = & \exp(6,07017625 + 0,748612 + 0,241162) \cdot \exp(0,0820868) \\ = & 1164,38724 \cdot 1,08555 = 1264. \end{aligned}$$

Analog ergibt der Parameter $b_{AB(22)} = -0,0820868$ für Wahlbeteiligung = nein und politisches Interesse = mittel, dass durch die Interaktion der beiden Variablen die beobachtete Anzahl der Nichtwähler mit mittlerem politischen Interesse nur rund das 0,9-fache der mittleren, nur aufgrund der Haupteffekte der beiden Variablen geschätzten Anzahl von Nichtwählern mit mittlerem politischen Interesse ausmacht:

$$\begin{aligned} & \exp(b + b_{A(2)} + b_{B(2)}) \cdot \exp(b_{AB(22)}) \\ = & \exp(6,07017625 - 0,748612 + 0,241162) \cdot \exp(-0,0820868) \\ = & 260,5321 \cdot 0,921192 = 240. \end{aligned}$$

Der Parameter $b_{AB(13)} = -0,371980$ für Wahlbeteiligung = ja und politisches Interesse = wenig besagt, dass durch die Interaktion der beiden Variablen die beobachtete Anzahl

der Wähler mit wenigem politischen Interesse nur rund das 0,7-fache der mittleren, nur aufgrund der Haupteffekte der beiden Variablen geschätzten Anzahl von Wählern mit wenigem politischen Interesse ausmacht:

$$\begin{aligned} & \exp(b + b_{A(1)} + b_{B(3)}) \cdot \exp(b_{AB(13)}) \\ = & \exp(6,07017625 + 0,748612 + 0,15578) \cdot \exp(-0,371980) \\ = & 1069,095508 \cdot 0,6893675 = 737. \end{aligned}$$

Analog ergibt der Parameter $b_{AB(23)} = 0,371980$ für Wahlbeteiligung = nein und politisches Interesse = wenig, dass durch die Interaktion der beiden Variablen die beobachtete Anzahl der Nichtwähler mit wenigem politischen Interesse auf rund das 1,4-fache der mittleren, nur aufgrund der Haupteffekte der beiden Variablen geschätzten Anzahl von Nichtwählern mit wenigem politischen Interesse ansteigt:

$$\begin{aligned} & \exp(b + b_{A(2)} + b_{B(3)}) \cdot \exp(b_{AB(23)}) \\ = & \exp(6,07017625 - 0,748612 + 0,15578) \cdot \exp(0,371980) \\ = & 239,2106 \cdot 1,450604 = 347. \end{aligned}$$

Der Unterschied bei der Interpretation der Parameter liegt also darin, dass bei der Effekt-Kodierung jeweils Abweichungen vom Mittel gemessen werden, bei der Dummy-Kodierung dagegen Abweichungen von der Referenzkategorie.

Da ein saturiertes Modell geschätzt wird, sind die erwarteten Zellhäufigkeiten identisch mit den beobachteten Zellhäufigkeiten und Anpassungstests sind nicht relevant.

Unter Verwendung von SPSS erhält man nachstehende Outputs.

SPSS-Output für das saturierte Modell mit Dummy-Kodierung:

```

-----
                        GENERAL LOGLINEAR ANALYSIS
-----
Data Information
  3513 cases are accepted.
    5 cases are rejected because of missing data.
  3513 weighted cases will be used in the analysis.
    6 cells are defined.
    0 structural zeros are imposed by design.
    0 sampling zeros are encountered.
```

Variable Information

Factor	Levels	Value
BET	2	Wahlbeteiligung, letzte Bundestagswahl
		1 JA
		2 NEIN
POL_INT	3	Politisches Interesse
		1 stark
		2 mittel
		3 wenig

Model and Design Information

Model: Multinomial

Design: Constant + BET + POL_INT + BET*POL_INT

Correspondence Between Parameters and Terms of the Design

Parameter	Aliased	Term
1		Constant
2		[BET = 1]
3	x	[BET = 2]
4		[POL_INT = 1]
5		[POL_INT = 2]
6	x	[POL_INT = 3]
7		[BET = 1]*[POL_INT = 1]
8		[BET = 1]*[POL_INT = 2]
9	x	[BET = 1]*[POL_INT = 3]
10	x	[BET = 2]*[POL_INT = 1]
11	x	[BET = 2]*[POL_INT = 2]
12	x	[BET = 2]*[POL_INT = 3]

Note: 'x' indicates an aliased (or a redundant) parameter.
These parameters are set to zero.

Design Matrix

Factor	Value	Cell Structure	Parameter					
			1	2	4	5	7	8
BET	JA							
POL_INT	stark	1,000	1	1	1	0	1	0
POL_INT	mittel	1,000	1	1	0	1	0	1
POL_INT	wenig	1,000	1	1	0	0	0	0
BET	NEIN							
POL_INT	stark	1,000	1	0	1	0	0	0
POL_INT	mittel	1,000	1	0	0	1	0	0
POL_INT	wenig	1,000	1	0	0	0	0	0

Convergence Information

Maximum number of iterations: 20
 Relative difference tolerance: ,001
 Final relative difference: 6,21812E-13
 Maximum likelihood estimation converged at iteration 1.

Table Information

Factor	Value	Observed		Expected	
		Count	%	Count	%
BET	JA				
POL_INT	stark	822,00 (23,40)		822,00 (23,40)	
POL_INT	mittel	1264,00 (35,98)		1264,00 (35,98)	

5 Modellierung multinomialer Daten (log-lineare Modelle)

POL_INT	wenig	737,00 (20,98)	737,00 (20,98)
BET	NEIN		
POL_INT	stark	103,00 (2,93)	103,00 (2,93)
POL_INT	mittel	240,00 (6,83)	240,00 (6,83)
POL_INT	wenig	347,00 (9,88)	347,00 (9,88)

Table Information

Factor	Value	Resid.	Adj. Resid.	Dev. Resid.
BET	JA			
POL_INT	stark	,00	,00	,00
POL_INT	mittel	,00	,00	,00
POL_INT	wenig	,00	,00	,00
BET	NEIN			
POL_INT	stark	,00	,00	,00
POL_INT	mittel	,00	,00	,00
POL_INT	wenig	,00	,00	,00

Goodness-of-fit Statistics

	Chi-Square	DF	Sig.
Likelihood Ratio	,0000	0	,
Pearson	,0000	0	,

Parameter Estimates

Constant	Estimate
1	5,8493

Note: Constant is not a parameter under multinomial assumption.
Therefore, standard errors are not calculated.

Parameter	Estimate	SE	Z-value	Asymptotic 95% CI	
2	,7533	,0651	11,57	,63	,88
3	,0000	,	,	,	,
4	-1,2146	,1122	-10,82	-1,43	-,99
5	-,3687	,0840	-4,49	-,53	-,20
6	,0000	,	,	,	,
7	1,3237	,1231	10,75	1,08	1,57
8	,9081	,0959	9,47	,72	1,10
9	,0000	,	,	,	,
10	,0000	,	,	,	,
11	,0000	,	,	,	,
12	,0000	,	,	,	,

SPSS-Output für das saturierte Modell mit Effekt-Kodierung:

***** LOG LINEAR ANALYSIS *****

Data Information

3513 cases are accepted.

0 cases rejected because of out-of-range factor values.

5 cases rejected because of missing data.

3513 weighted cases will be used in the analysis.

FACTOR Information

Factor	Level	Label
BET	2	Wahlbeteiligung, letzte Bundestagswahl
POL_INT	3	Politisches Interesse

DESIGN Information

1 Design/Model will be processed.

Correspondence Between Effects and Columns of Design/Model 1

Starting Column	Ending Column	Effect Name
1	1	BET
2	3	POL_INT
4	5	BET BY POL_INT

Design Matrix

1-BET 2-POL_INT		Parameter				
Factor		1	2	3	4	5
1	1	1,000	1,000	,000	1,000	,000
1	2	1,000	,000	1,000	,000	1,000
1	3	1,000	-1,000	-1,000	-1,000	-1,000
2	1	-1,000	1,000	,000	-1,000	,000
2	2	-1,000	,000	1,000	,000	-1,000
2	3	-1,000	-1,000	-1,000	1,000	1,000

*** ML converged at iteration 3.

Maximum difference between successive iterations = ,00000.

Observed, Expected Frequencies and Residuals

Factor	Code	OBS. count & PCT.	EXP. count & PCT.
BET	JA		
POL_INT	stark	822,00 (23,40)	822,00 (23,40)
POL_INT	mittel	1264,00 (35,98)	1264,00 (35,98)
POL_INT	wenig	737,00 (20,98)	737,00 (20,98)
BET	NEIN		
POL_INT	stark	103,00 (2,93)	103,00 (2,93)
POL_INT	mittel	240,00 (6,83)	240,00 (6,83)
POL_INT	wenig	347,00 (9,88)	347,00 (9,88)

Observed, Expected Frequencies and Residuals (cont.)

Factor	Code	Residual	Std. Resid.	Adj. Resid.
BET	JA			
POL_INT	stark	,0000	,0000	,0000
POL_INT	mittel	,0000	,0000	,0000
POL_INT	wenig	,0000	,0000	,0000
BET	NEIN			
POL_INT	stark	,0000	,0000	,0000
POL_INT	mittel	,0000	,0000	,0000
POL_INT	wenig	,0000	,0000	,0000

Goodness-of-fit Statistics

Likelihood Ratio Chi-Square = ,00000 DF = 0 P = 1,000
 Pearson Chi-Square = ,00000 DF = 0 P = 1,000

Estimates for Parameters

Parameter	Coeff.	Std. Err.	Z-Value	Lower 95 CI	Upper 95 CI
1	,7486120284	,02364	31,66466	,70227	,79495
POL_INT					

Parameter	Coeff.	Std. Err.	Z-Value	Lower 95 CI	Upper 95 CI
2	-,396941567	,03833	-10,35523	-,47207	-,32181
3	,2411614901	,03118	7,73490	,18005	,30227
BET BY POL_INT					
Parameter	Coeff.	Std. Err.	Z-Value	Lower 95 CI	Upper 95 CI
4	,2898936750	,03833	7,56261	,21476	,36503
5	,0820867973	,03118	2,63281	,02098	,14320

5.8 Log-lineare Modelle für dreidimensionale Kontingenztafeln

5.8.1 Die Modelle

In die Analyse werden nunmehr drei Variable einbezogen:

- Variable A mit J Kategorien, $j = 1, \dots, J$;
- Variable B mit K Kategorien, $k = 1, \dots, K$;
- Variable C mit L Kategorien, $l = 1, \dots, L$.

Es resultiert eine dreidimensionale Kontingenztafel mit $J \cdot K \cdot L$ Zellen, die die Zellhäufigkeiten y_{jkl} aufweisen.

Bei einer dreidimensionalen Kontingenztafel ergeben sich folgende Randverteilungen:

- Mit der Summation über l erhält man eine $J \times K$ Randtafel der Variablen A und B mit den Zellhäufigkeiten y_{jk+} .
- Mit der Summation über k erhält man eine $J \times L$ Randtafel der Variablen A und C mit den Zellhäufigkeiten y_{j+l} .
- Mit der Summation über j erhält man eine $K \times L$ Randtafel der Variablen B und C mit den Zellhäufigkeiten y_{+kl} .
- Mit der Summation über k und l erhält man die Randverteilung der Variablen A mit den Randhäufigkeiten y_{j++} .

- Mit der Summation über j und l erhält man die Randverteilung der Variablen B mit den Randhäufigkeiten y_{+k+} .
- Mit der Summation über j und k erhält man die Randverteilung der Variablen C mit den Randhäufigkeiten y_{++l} .

$$\begin{aligned}
 y_{jk+} &= \sum_{l=1}^L y_{jkl}, & y_{j+l} &= \sum_{k=1}^K y_{jkl}, & y_{+kl} &= \sum_{j=1}^J y_{jkl}, \\
 y_{j++} &= \sum_{k=1}^K \sum_{l=1}^L y_{jkl} = \sum_{k=1}^K y_{jk+} = \sum_{l=1}^L y_{j+l}, \\
 y_{+k+} &= \sum_{j=1}^J \sum_{l=1}^L y_{jkl} = \sum_{j=1}^J y_{jk+} = \sum_{l=1}^L y_{+kl}, \\
 y_{++l} &= \sum_{j=1}^J \sum_{k=1}^K y_{jkl} = \sum_{j=1}^J y_{j+l} = \sum_{k=1}^K y_{+kl}, \\
 y_{+++} &= \sum_{j=1}^J \sum_{k=1}^K \sum_{l=1}^L y_{jkl} = n.
 \end{aligned} \tag{5.86}$$

Analog ist die Notation für die Zellwahrscheinlichkeiten $\pi_{jkl} = m_{jkl}/n = P(A_j \cap B_k \cap C_l)$, die geschätzten Zellwahrscheinlichkeiten $\hat{\pi}_{jkl}$, die erwarteten Zellhäufigkeiten m_{jkl} und die geschätzten erwarteten Zellhäufigkeiten \hat{m}_{jkl} .

Hieran ist sofort ersichtlich, dass es eine größere Vielfalt von möglichen Hypothesen und damit verbundener log-linearer Modelle gibt. Nachfolgend sollen einige log-lineare Modelle auf Basis des multinomialen Stichprobenmodells mit den dazugehörigen Hypothesen betrachtet werden. Dazu sei vereinbart, dass die Variable A in den Zeilen, die Variable B in den Spalten und die Variable C in den Schichten der dreidimensionalen Kontingenztafel stehen.

Die Multinomialverteilung bei Einbeziehung von 3 Variablen lautet:

$$f(y; \pi) = n! \prod_{j=1}^J \prod_{k=1}^K \prod_{l=1}^L \frac{\pi_{jkl}^{y_{jkl}}}{y_{jkl}!}. \tag{5.87}$$

Saturiertes Modell (ABC):

Wie bereits bei einer zweidimensionalen Kontingenztafel gezeigt wurde, werden im

saturierten Modell keine Nebenbedingungen an die erwarteten Zellhäufigkeiten gestellt. Allerdings sind die Restriktionen zur Vermeidung einer Überparametrisierung zu beachten. Für dieses Modell ist

$$E(Y_{jkl}) = m_{jkl} = n \cdot \pi_{jkl}. \quad (5.88)$$

Die link Funktion des saturierten Modells lautet:

$$\eta_{jkl} = \log m_{jkl} = \beta + \beta_{A(j)} + \beta_{B(k)} + \beta_{C(l)} + \beta_{AB(jk)} + \beta_{AC(jl)} + \beta_{BC(kl)} + \beta_{ABC(jkl)}. \quad (5.89)$$

Generell wird für hierarchische log-lineare Modelle höherer Dimension eine Kurzschreibweise eingeführt, um nicht jedesmal das Modell ausführlich notieren zu müssen. Diese Kurzform gibt die Terme höchster Ordnung an, die keine Marginaleffekte eines anderen Effekts sind. Diese Kurzschreibweise soll sowohl das Modell identifizieren als auch anzeigen, welche Randsummen bei den Maximum-Likelihood-Schätzungen angepasst werden müssen. Für das angegebene saturierte Modell impliziert der Einschluss der Drei-Variablen-Interaktion, dass auch alle Zwei-Variablen-Interaktionen und die Haupteffekte aller drei Variablen im Modell enthalten sind, d.h., es reicht aus, diese Drei-Variablen-Interaktion anzugeben. Somit schreibt man für das saturierte Modell kurz (ABC).

Die Anzahl der unabhängigen, zu schätzenden Parameter des saturierten Modells ist $J \cdot K \cdot L$ und setzt sich gemäß Tabelle 5.9 zusammen.

Tabelle 5.9: Anzahl der unabhängigen Parameter

Term	Anzahl der unabhängigen Parameter
β	1
$\beta_{A(j)}$	$J - 1$
$\beta_{B(k)}$	$K - 1$
$\beta_{C(l)}$	$L - 1$
$\beta_{AB(jk)}$	$J \cdot K - J - K + 1 = (J - 1)(K - 1)$
$\beta_{AC(jl)}$	$J \cdot L - J - L + 1 = (J - 1)(L - 1)$
$\beta_{BC(kl)}$	$K \cdot L - K - L + 1 = (K - 1)(L - 1)$
$\beta_{ABC(jkl)}$	$JKL - JK - JL - KL + J + K + L - 1 = (J - 1)(K - 1)(L - 1)$
Gesamt	$J \cdot K \cdot L$

Die sum-to-zero Restriktion lautet:

$$\begin{aligned}
 \sum_{j=1}^J \beta_{A(j)} &= \sum_{k=1}^K \beta_{B(k)} = \sum_{l=1}^L \beta_{C(l)} = 0; \\
 \sum_{j=1}^J \beta_{AB(jk)} &= \sum_{k=1}^K \beta_{AB(jk)} = 0; \quad \sum_{j=1}^J \beta_{AC(jl)} = \sum_{l=1}^L \beta_{AC(jl)} = 0; \\
 \sum_{k=1}^K \beta_{BC(kl)} &= \sum_{l=1}^L \beta_{BC(kl)} = 0; \quad \sum_{j=1}^J \beta_{ABC(jkl)} = \sum_{k=1}^K \beta_{ABC(jkl)} = \sum_{l=1}^L \beta_{ABC(jkl)} = 0; \\
 n = y_{+++} &= \sum_{j=1}^J \sum_{k=1}^K \sum_{l=1}^L \exp(\beta + \beta_{A(j)} + \beta_{B(k)} + \beta_{C(l)} + \beta_{AB(jk)} + \beta_{AC(jl)} \\
 &\quad + \beta_{BC(kl)} + \beta_{ABC(jkl)})
 \end{aligned} \tag{5.90}$$

Die letzte Angabe in (5.90) bezieht sich auf das multinomiale Stichprobenmodell, bei dem der Gesamtstichprobenumfang n fest vorgegeben ist. Die geschätzten Parameter der Haupteffekte ergeben sich analog zu (5.67) bis (5.69), und die geschätzten Parameter höherer Ordnung sind die Differenzen zu den geschätzten Parametern niedriger Ordnung:

$$\begin{aligned}
 b &= \frac{1}{JKL} \sum_{j=1}^J \sum_{k=1}^K \sum_{l=1}^L \log y_{jkl}, \\
 b_{A(j)} &= \frac{1}{KL} \sum_{k=1}^K \sum_{l=1}^L \log y_{jkl} - b, \\
 b_{B(k)} &= \frac{1}{JL} \sum_{j=1}^J \sum_{l=1}^L \log y_{jkl} - b, \\
 b_{C(l)} &= \frac{1}{JK} \sum_{j=1}^J \sum_{k=1}^K \log y_{jkl} - b, \\
 b_{AB(jk)} &= \frac{1}{L} \sum_{l=1}^L \log y_{jkl} - b_{A(j)} - b_{B(k)} - b, \\
 b_{AC(jl)} &= \frac{1}{K} \sum_{k=1}^K \log y_{jkl} - b_{A(j)} - b_{C(l)} - b,
 \end{aligned} \tag{5.91}$$

$$b_{BC(kl)} = \frac{1}{J} \sum_{j=1}^J \log y_{jkl} - b_{B(k)} - b_{C(l)} - b,$$

$$b_{ACB(jlk)} = \log y_{jkl} - b_{AB(jk)} - b_{AC(jl)} - b_{BC(kl)} - b_{A(j)} - b_{B(k)} - b_{C(l)} - b.$$

Bei Verwendung der corner-point Restriktion resultieren die Parameter:

$$\begin{aligned} b &= \log y_{JKL}, \\ b_{A(j)} &= \log y_{jKL} - \log y_{JKL}, \\ b_{B(k)} &= \log y_{JkL} - \log y_{JKL}, \\ b_{C(l)} &= \log y_{JKl} - \log y_{JKL}, \\ b_{AB(jk)} &= \log y_{jkl} - \log y_{jKL} - b_{B(k)}, \\ b_{AC(jl)} &= \log y_{jkl} - \log y_{jKL} - b_{C(l)}, \\ b_{BC(kl)} &= \log y_{jkl} - \log y_{jKL} - b_{C(l)}, \\ b_{ACB(jlk)} &= \log y_{jkl} - \log y_{JKL} - b_{A(j)} - b_{B(k)} - b_{C(l)} - b_{AB(jk)} - b_{AC(jl)} - b_{BC(kl)} \end{aligned} \quad (5.92)$$

für $j = 1, \dots, J - 1$, $k = 1, \dots, K - 1$ und $l = 1, \dots, L - 1$.

Im saturierten log-linearen Modell sind die ML-Schätzungen der erwarteten Zellhäufigkeiten:

$$\hat{m}_{jkl} = y_{jkl} \quad (j = 1, \dots, J; k = 1, \dots, K; l = 1, \dots, L).$$

Nachfolgend werden weitere Spezifikationen von hierarchischen log-linearen Modellen angegeben, wobei jedoch auf die Einzeldarstellung der Formeln zur Parameterschätzung verzichtet wird.

Modell der totalen Unabhängigkeit (A|B|C):

Für dieses Modell lautet die Nullhypothese

$$\pi_{jkl} = P(A_j \cap B_k \cap C_l) = P(A_j)P(B_k)P(C_l) = \pi_{j++}\pi_{+k+}\pi_{++l}. \quad (5.93)$$

Damit resultiert als Erwartungswert

$$E(Y_{jkl}) = m_{jkl} = n \cdot \pi_{j++}\pi_{+k+}\pi_{++l} = m_{j++}m_{+k+}m_{++l}/n^2. \quad (5.94)$$

Die link Funktion ist gegeben mit:

$$\eta_{jkl} = \log m_{jkl} = \beta + \beta_{A(j)} + \beta_{B(k)} + \beta_{C(l)}. \quad (5.95)$$

Die Kurzform lautet $(A|B|C)$, d.h., die Variablen A, B und C sind total unabhängig. Die Kurzform impliziert gleichzeitig, dass $\hat{m}_{j++} = y_{j++}$, $\hat{m}_{+k+} = y_{+k+}$ und $\hat{m}_{++l} = y_{++l}$ für alle j, k und l als Randsummenbedingungen zu erfüllen sind.

Dieses Modell $(A|B|C)$ weist $J + K + L - 2$ unabhängige, zu schätzende Parameter auf; die Anzahl der Freiheitsgrade beträgt $f = JKL - J - K - L + 2$.

Als ML-Schätzungen der π_{jkl} für diese Modell erhält man gemäß (5.55):

$$\hat{\pi}_{jkl}(A|B|C) = \hat{\pi}_{j++}\hat{\pi}_{+k+}\hat{\pi}_{++l} = (y_{j++}/n)(y_{+k+}/n)(y_{++l}/n)$$

und wegen $m_{jkl} = n\hat{\pi}_{jkl}$ als ML-Schätzungen für m_{jkl}

$$\hat{m}_{jkl}(A|B|C) = n\hat{\pi}_{jkl}(A|B|C)y_{j++}y_{+k+}y_{++l}/n^2. \quad (5.96)$$

Modell mit einer Variablen unabhängig von den anderen beiden:

Es wird unterstellt, dass zwei Variablen voneinander abhängig sind, aber jede dieser beiden von der dritten unabhängig ist. Deshalb wird diese Modellspezifikation auch als das Modell der partiellen Unabhängigkeit bezeichnet. Hierfür müssen drei verschiedene Modellspezifikationen unterschieden werden.

Modell $(BC|A)$:

Für dieses Modell lautet die Nullhypothese:

$$\pi_{jkl} = P(A_j)P(B_k \cap C_l) = \pi_{j++}\pi_{+kl}. \quad (5.97)$$

Damit resultiert als Erwartungswert

$$E(Y_{jkl}) = n \cdot \pi_{j++}\pi_{+kl} = m_{j++}m_{+kl}/n \quad (5.98)$$

und als link Funktion

$$\eta_{jkl} = \log m_{jkl} = \beta + \beta_{A(j)} + \beta_{B(k)} + \beta_{C(l)} + \beta_{BC(kl)}. \quad (5.99)$$

Die Kurzform lautet $(BC|A)$, d.h., die Variable A ist unabhängig von den Variablen B und C, während die Variablen B und C Interaktionen aufweisen können. Die Zeilen sind unabhängig von den Spalten und Schichten; es wird praktisch keine Unterscheidung zwischen Spalten und Schichten gemacht. Deshalb ist das Modell äquivalent zum Unabhängigkeitsmodell in einer $J \times (KL)$ -zweidimensionalen Kontingenztabelle, in der die

Spalten alle Kombinationen der Spalten und Schichten der dreidimensionalen Tabelle aufnehmen. Die Kurzform zeigt gleichzeitig an, dass $\hat{m}_{j++} = y_{j++}$ und $\hat{m}_{+kl} = y_{+kl}$ für alle j, k und l als Randsummenbedingungen zu erfüllen sind.

Das Modell (BC|A) weist J + KL - 1 unabhängige, zu schätzende Parameter auf; die Anzahl der Freiheitsgrade beträgt f = (KL - 1)(J - 1).

Als ML-Schätzungen der π_{jkl} für diese Modell erhält man:

$$\hat{\pi}_{jkl}(BC|A) = \hat{\pi}_{j++}\hat{\pi}_{+kl} = (y_{j++}/n)(y_{+kl}/n)$$

und als ML-Schätzungen für m_{jkl}

$$\hat{m}_{jkl}(BC|A) = n\hat{\pi}_{jkl}(BC|A) = y_{j++}y_{+kl}/n. \quad (5.100)$$

Modell (AC|B):

Die Nullhypothese lautet:

$$\pi_{jkl} = P(B_k)P(A_j \cap C_l) = \pi_{j+l}\pi_{+k+}. \quad (5.101)$$

Damit resultiert als Erwartungswert

$$E(Y_{jkl}) = n \cdot \pi_{j+l}\pi_{+k+} = m_{j+l}m_{+k+}/n \quad (5.102)$$

und als link Funktion

$$\eta_{jkl} = \log m_{jkl} = \beta + \beta_{A(j)} + \beta_{B(k)} + \beta_{C(l)} + \beta_{AC(jl)}. \quad (5.103)$$

Die Kurzform lautet (AC|B), d.h., die Variable B ist unabhängig von den Variablen A und C, während die Variablen A und C Interaktionen aufweisen können. Die Spalten sind unabhängig von den Zeilen und Schichten. Die Kurzform impliziert ausserdem, dass $\hat{m}_{j+l} = y_{j+l}$ und $\hat{m}_{+k+} = y_{+k+}$ für alle j, k und l als Randsummenbedingungen erfüllt sein müssen.

Das Modell (AC|B) weist K + JL - 1 unabhängige, zu schätzende Parameter auf; die Anzahl der Freiheitsgrade beträgt f = (JL - 1)(K - 1).

Als ML-Schätzungen der π_{jkl} für diese Modell erhält man:

$$\hat{\pi}_{jkl}(AC|B) = \hat{\pi}_{j+l}\hat{\pi}_{+k+} = (y_{j+l}/n)(y_{+k+}/n)$$

und als ML-Schätzungen für m_{jkl}

$$\hat{m}_{jkl}(AC|B) = n\hat{\pi}_{jkl}(AC|B) = y_{j+l}y_{+k+}/n. \quad (5.104)$$

Modell (AB|C):

Die Nullhypothese lautet:

$$\pi_{jkl} = P(C_l)P(A_j \cap B_k) = \pi_{jk+}\pi_{++l}. \quad (5.105)$$

Damit resultiert als Erwartungswert

$$E(Y_{jkl}) = n \cdot \pi_{jk+}\pi_{++l} = m_{jk+}m_{++l}/n \quad (5.106)$$

und als link Funktion

$$\eta_{jkl} = \log m_{jkl} = \beta + \beta_{A(j)} + \beta_{B(k)} + \beta_{C(l)} + \beta_{AB(jk)}. \quad (5.107)$$

Die Kurzform lautet (AB|C), d.h., die Variable C ist unabhängig von den Variablen A und B, während die Variablen A und B Interaktionen aufweisen können. Die Schichten sind unabhängig von den Zeilen und Spalten. Die Kurzform impliziert ausserdem, dass $\hat{m}_{jk+} = y_{jk+}$ und $\hat{m}_{++l} = y_{++l}$ für alle j, k und l als Randsummenbedingungen erfüllt sein müssen.

Das Modell (AB|C) weist $L + JK - 1$ unabhängige, zu schätzende Parameter auf; die Anzahl der Freiheitsgrade beträgt $f = (JK - 1)(L - 1)$.

Als ML-Schätzungen der π_{jkl} für diese Modell erhält man:

$$\hat{\pi}_{jkl}(AB|C) = \hat{\pi}_{jk+}\hat{\pi}_{++l} = (y_{jk+}/n)(y_{++l}/n)$$

und als ML-Schätzungen für m_{jkl}

$$\hat{m}_{jkl}(AB|C) = n\hat{\pi}_{jkl}(AB|C) = y_{jk+}y_{++l}/n. \quad (5.108)$$

Modell der bedingten Unabhängigkeit:

In dem Modell der bedingten Unabhängigkeit sind zwei Variablen unabhängig voneinander, wenn für die dritte Variable kontrolliert wird. Hierfür gibt es ebenfalls drei verschiedene Modellspezifikationen.

Modell (AB|AC):

Die Kurzform lautet (AB|AC), d.h., in diesem Modell sind die Variablen B und C bedingt unabhängig, gegeben A_j :

$$P(A_j \cap B_k \cap C_l) / P(A_j) = P(A_j \cap B_k) / P(A_j) \cdot P(A_j \cap C_l) / P(A_j)$$

$$\pi_{jkl} / \pi_{j++} = \pi_{jk+} / \pi_{j++} \cdot \pi_{j+l} / \pi_{j++}.$$

Die Nullhypothese lautet damit:

$$\pi_{jkl} = \pi_{jk+} \pi_{j+l} / \pi_{j++}. \quad (5.109)$$

Damit resultiert als Erwartungswert

$$E(Y_{jkl}) = n \cdot \pi_{jk+} \pi_{j+l} / \pi_{j++} = m_{jk+} m_{j+l} / m_{j++} \quad (5.110)$$

und als link Funktion

$$\eta_{jkl} = \log m_{jkl} = \beta + \beta_{A(j)} + \beta_{B(k)} + \beta_{C(l)} + \beta_{AB(jk)} + \beta_{AC(jl)}. \quad (5.111)$$

Die Kurzform zeigt gleichzeitig an, dass $\hat{m}_{jk+} = y_{jk+}$ und $\hat{m}_{j+l} = y_{j+l}$ und $\hat{m}_{j++} = y_{j++}$ für alle j, k und l als Randsummenbedingungen zu erfüllen sind.

Das Modell (AB|AC) weist $J(K + L - 1)$ unabhängige, zu schätzende Parameter auf und hat somit $f = J(K - 1)(L - 1)$ Freiheitsgrade.

Als ML-Schätzungen der π_{jkl} für diese Modell erhält man:

$$\hat{\pi}_{jkl}(AB|AC) = \hat{\pi}_{jk+} \hat{\pi}_{j+l} / \hat{\pi}_{j++} = (y_{jk+} / n)(y_{j+l} / n)(y_{j++} / n) = y_{jk+} y_{j+l} / y_{j++} n$$

und als ML-Schätzungen für m_{jkl}

$$\hat{m}_{jkl}(AB|AC) = n \hat{\pi}_{jkl}(AB|AC) = y_{jk+} y_{j+l} / y_{j++}. \quad (5.112)$$

Modell (AC|BC):

Die Kurzform lautet (AC|BC), d.h., die Variablen A und B sind bedingt unabhängig, gegeben C_l :

$$P(A_j \cap B_k \cap C_l) / P(C_l) = P(A_j \cap C_l) / P(C_l) \cdot P(B_k \cap C_l) / P(C_l)$$

$$\pi_{jkl} / \pi_{++l} = \pi_{j+l} / \pi_{++l} \cdot \pi_{+kl} / \pi_{++l}.$$

Die Nullhypothese lautet:

$$\pi_{jkl} = \pi_{j+l} \pi_{+kl} / \pi_{++l}. \quad (5.113)$$

Damit resultiert als Erwartungswert

$$E(Y_{jkl}) = n \cdot \pi_{j+l}\pi_{+kl}/\pi_{++l} = m_{j+l}m_{+kl}/m_{++l} \quad (5.114)$$

und als link Funktion

$$\eta_{jkl} = \log m_{jkl} = \beta + \beta_{A(j)} + \beta_{B(k)} + \beta_{C(l)} + \beta_{AC(jl)} + \beta_{BC(kl)}. \quad (5.115)$$

Die Kurzform zeigt gleichzeitig an, dass $\hat{m}_{j+l} = y_{j+l}$ und $\hat{m}_{+kl} = y_{+kl}$ und $\hat{m}_{++l} = y_{++l}$ für alle j, k und l als Randsummenbedingungen zu erfüllen sind.

Das Modell (AC|BC) weist $L(J + K - 1)$ unabhängige, zu schätzende Parameter auf und hat somit $f = L(J - 1)(K - 1)$ Freiheitsgrade.

Als ML-Schätzungen der π_{jkl} für diese Modell erhält man:

$$\hat{\pi}_{jkl}(AC|BC) = \hat{\pi}_{j+l}\hat{\pi}_{+kl}/\hat{\pi}_{++l} = (y_{j+l}/n)(y_{+kl}/n)(y_{++l}/n) = y_{j+l}y_{+kl}/y_{++l}n$$

und als ML-Schätzungen für m_{jkl}

$$\hat{m}_{jkl}(AC|BC) = n\hat{\pi}_{jkl}(AC|BC) = y_{j+l}y_{+kl}/y_{++l}. \quad (5.116)$$

Modell (AB|BC):

Die Kurzform lautet (AB|BC), d.h., die Variablen A und C sind bedingt unabhängig, gegeben B_k :

$$P(A_j \cap B_k \cap C_l)/P(B_k) = P(A_j \cap B_k)/P(B_k) \cdot P(B_k \cap C_l)/P(B_k)$$

$$\pi_{jkl}/\pi_{+k+} = \pi_{jk+}/\pi_{+k+} \cdot \pi_{+kl}/\pi_{+k+}.$$

Die Nullhypothese lautet:

$$\pi_{jkl} = \pi_{jk+}\pi_{+kl}/\pi_{+k+}. \quad (5.117)$$

Damit resultiert als Erwartungswert

$$E(Y_{jkl}) = n \cdot \pi_{jk+}\pi_{+kl}/\pi_{+k+} = m_{jk+}m_{+kl}/m_{+k+} \quad (5.118)$$

und als link Funktion

$$\eta_{jkl} = \log m_{jkl} = \beta + \beta_{A(j)} + \beta_{B(k)} + \beta_{C(l)} + \beta_{AB(jk)} + \beta_{BC(kl)}. \quad (5.119)$$

Die Kurzform zeigt gleichzeitig an, dass $\hat{m}_{jk+} = y_{jk+}$ und $\hat{m}_{+kl} = y_{+kl}$ und $\hat{m}_{+k+} = y_{+k+}$ für alle j, k und l als Randsummenbedingungen zu erfüllen sind.

Das Modell (AB|BC) weist $K(J + L - 1)$ unabhängige, zu schätzende Parameter auf und hat somit $f = K(J - 1)(L - 1)$ Freiheitsgrade.

Als ML-Schätzungen der π_{jkl} für diese Modell erhält man:

$$\hat{\pi}_{jkl}(AB|BC) = \hat{\pi}_{jk+}\hat{\pi}_{+kl}/\hat{\pi}_{+k+} = (y_{jk+}/n)(y_{+kl}/n)/(y_{+k+}/n) = y_{jk+}y_{+kl}/y_{+k+}n$$

und als ML-Schätzungen für m_{jkl}

$$\hat{m}_{jkl}(AB|BC) = n\hat{\pi}_{jkl}(AB|BC) = y_{jk+}y_{+kl}/y_{+k+}. \quad (5.120)$$

Modell ohne Drei-Faktor-Interaktion (AB|AC|BC):

Für dieses Modell ist die link Funktion gegeben mit

$$\eta_{jkl} = \log m_{jkl} = \beta + \beta_{A(j)} + \beta_{B(k)} + \beta_{C(l)} + \beta_{AB(jk)} + \beta_{AC(jl)} + \beta_{BC(kl)}. \quad (5.121)$$

Die Kurzform lautet (AB|AC|BC), d.h., es werden alle paarweisen Assoziationen zwischen den Variablen (Zwei-Variablen-Interaktionen) in das Modell eingeschlossen, jedoch nicht die Drei-Variablen-Interaktion, wodurch die bedingte Assoziation zwischen zwei Variablen (z.B. A und B) nicht mehr von der dritten Variablen (C) abhängt. Das bedeutet jedoch nicht, dass A bzw. B von C unabhängig sind, sondern der Zusammenhang zwischen A und B wird nicht von C beeinflusst. Die Bedingungen in den erwarteten Zellhäufigkeiten m_{jkl} können nur mittels der odds ratios formuliert werden. Dazu wird eine Variable fixiert und die odds ratios der verbleibenden zwei Variablen betrachtet. Wenn z.B. die Variable C auf die Ausprägung l (d.h. die Schicht l in der dreidimensionalen Tabelle) festgelegt wird, so sind die odds ratios $(m_{11l}/m_{1kl})/(m_{j1l}/m_{jkl}) = m_{11l}m_{jkl}/m_{1kl}m_{j1l}$. Diese odds ratios müssen in allen Schichten ($l = 1, \dots, L$) gleich sein, d.h., das Modell kann äquivalent formuliert werden als

$$\frac{m_{11l}m_{jk1}}{m_{1k1}m_{j1l}} = \frac{m_{11l}m_{jkl}}{m_{1kl}m_{j1l}} \quad (5.122)$$

für alle $j = 2, \dots, J$; $k = 2, \dots, K$ und $l = 2, \dots, L$. Das Modell bleibt der Form nach unverändert, wenn man die Zeilen fixiert oder die Spalten fixiert.

Das Modell (AB|AC|BC) weist $JKL - (J-1)(K-1)(L-1) = JK+JL+KL-J-K-L+1$ unabhängige Parameter auf. Die Anzahl der Freiheitsgrade ist dementsprechend $f = (J-1)(K-1)(L-1)$. Es gibt keine einfache Darstellung der ML-Schätzungen für π_{jkl} und m_{jkl} . Ein iteratives Schätzverfahren muss verwendet werden, um die ML-Schätzungen zu erhalten. Sie müssen jedoch die Randsummenbedingungen $\hat{m}_{jk+} = y_{jk+}$, $\hat{m}_{j+l} = y_{j+l}$ und $\hat{m}_{+kl} = y_{+kl}$ für alle j , k und l erfüllen.

Für die behandelten Modelle können die Teststatistiken in Analogie zu (5.83) bzw. (5.85) notiert werden. Die Anzahl der Freiheitsgrade beträgt: $f = JKL - \text{Anzahl der unabhängigen (zu schätzenden) Parameter des interessierenden Modells}$. Unter Berücksichtigung der entsprechenden Freiheitsgrade gelten die oben getroffenen Aussagen zur Testentscheidung.

Bestimmte log-lineare Modelle können auch gegeneinander getestet werden. Es handelt sich dabei um sogenannte “nested“ Modelle, bei denen ein Modell (es sei mit M_1 bezeichnet) vollständig in einem anderen Modell (es sei mit M_2 bezeichnet) enthalten ist; man sagt auch: M_2 ist strikt größer als M_1 . Dann ist die Parametermenge von M_1 vollständig in der Parametermenge von M_2 enthalten. Mit einem Test des Modells M_1 gegen das Modell M_2 soll festgelegt werden, ob die in M_2 zusätzlich enthaltenen Parameter wesentlich zur Anpassung der Zellhäufigkeiten sind.

Die Deviance ist stets die Prüfung eines reduzierten Modells gegen das saturierte Modell, das strikt größer ist als alle anderen oben behandelten Modelle. Somit enthält $D(M_1)$ den Test des Modells M_1 gegen das saturierte Modell und $D(M_2)$ den Test des Modells M_2 gegen das saturierte Modell. Da in M_2 mehr Parameter enthalten sind, die möglicherweise zur Verbesserung der Anpassung beitragen, gilt stets $D(M_1) \geq D(M_2)$.

Differenzenbildung führt zu

$$\Delta D = D(M_1) - D(M_2) = 2 \sum_{j=1}^J \sum_{k=1}^K \sum_{l=1}^L y_{jkl} \log \left(\frac{\hat{m}_{jkl}^{(2)}}{\hat{m}_{jkl}^{(1)}} \right). \quad (5.123)$$

Im allgemeinen wird man $\Delta D = D(M_1) - D(M_2)$ verwenden, da $D(M_1)$ und $D(M_2)$ durch vorherige Modellschätzungen schon vorliegen. Wenn ΔD groß ist, tragen die zusätzlichen Parameter in M_2 wesentlich zur besseren Anpassung der Zellhäufigkeiten

bei, andernfalls nicht. Unter der Nullhypothese, dass das Modell M_2 wahr ist, weist ΔD eine asymptotische Chi-Quadrat-Verteilung auf. Die Anzahl der Freiheitsgrade ergibt sich zu $f(\Delta D) = f(M_1) - f(M_2)$.

5.8.2 Beispiel

In Fortführung des Beispiels aus Abschnitt 5.7 wird eine dritte Variable Wohngebiet (C) mit den Kategorien alte Bundesländer (West) = 1 und neue Bundesländer (Ost) = 2 in die Analyse aufgenommen, wobei die zweite Kategorie als Referenzkategorie gewählt wird. Die dreidimensionale Kontingenztabelle mit den beobachteten Zellhäufigkeiten enthält Tabelle 5.10.

Tabelle 5.10: Dreidimensionale Kontingenztabelle für Wahlbeteiligung (A), politisches Interesse (B) und Wohngebiet (C)

Wohngebiet			politisches Interesse			Gesamt
			stark	mittel	wenig	
West	Wahlbeteiligung	ja	573	872	475	1920
		nein	78	162	238	478
	Gesamt		651	1034	713	2398
Ost	Wahlbeteiligung	ja	249	392	262	903
		nein	25	78	109	212
	Gesamt		274	470	371	1115

Unter Verwendung ausschließlich der corner-point-Restriktion und damit der Dummy-Kodierung für die Spezifikation der Kontrast-Variablen werden nachfolgend verschiedene log-lineare Modelle geschätzt. Bezüglich der Interpretation der Schätzergebnisse ist damit stets ein Vergleich zu den Referenzkategorien der Variablen zu führen.

saturiertes Modell

Die link Funktion ist mit (5.89) gegeben. Die geschätzten Zellhäufigkeiten \hat{m}_{jkl} entsprechen den beobachteten Zellhäufigkeiten y_{jkl} , womit alle Residuen gleich Null sind. Anpassungstests sind damit nicht relevant.

Die Anzahl der zu schätzenden Parameter beträgt $J \cdot K \cdot L = 2 \cdot 3 \cdot 2 = 12$. Die geschätzten Parameter des saturierten Modells ergeben sich gemäß (5.92) zu (gerundet auf 4 Dezimalstellen):

$$b = \log y_{232} = \log(109) = 4,6913$$

5.8 Log-lineare Modelle für dreidimensionale Kontingenztafeln

$$\begin{aligned}
b_{A(1)} &= \log y_{132} - \log y_{232} = \log(262) - \log(109) = 0,8770 \\
b_{A(2)} &= 0 \\
b_{B(1)} &= \log y_{212} - \log y_{232} = \log(25) - \log(109) = -1,4725 \\
b_{B(2)} &= \log y_{222} - \log y_{232} = \log(78) - \log(109) = -0,3346 \\
b_{B(3)} &= 0 \\
b_{C(1)} &= \log y_{231} - \log y_{232} = \log(238) - \log(109) = 0,7809 \\
b_{C(2)} &= 0 \\
b_{AB(11)} &= \log y_{112} - \log y_{132} - b_{B(1)} = \log(249) - \log(262) - (-1,4725) = 1,4216 \\
b_{AB(12)} &= \log y_{122} - \log y_{132} - b_{B(2)} = \log(392) - \log(262) - (-0,3346) = 0,7376 \\
b_{AB(13)} &= 0; \quad b_{AB(21)} = 0; \quad b_{AB(22)} = 0; \quad b_{AB(23)} = 0 \\
b_{AC(11)} &= \log y_{131} - \log y_{132} - b_{C(1)} = \log(475) - \log(262) - 0,7809 = -0,1860 \\
b_{AC(12)} &= 0; \quad b_{AC(21)} = 0; \quad b_{AC(22)} = 0 \\
b_{BC(11)} &= \log y_{211} - \log y_{212} - b_{C(1)} = \log(78) - \log(25) - 0,7809 = 0,3569 \\
b_{BC(21)} &= \log y_{221} - \log y_{222} - b_{C(1)} = \log(162) - \log(78) - 0,7809 = -0,0500 \\
b_{BC(12)} &= 0; \quad b_{BC(22)} = 0; \quad b_{BC(31)} = 0; \quad b_{BC(32)} = 0 \\
b_{ABC(111)} &= \log y_{111} - \log y_{232} - b_{A(1)} - b_{B(1)} - b_{C(1)} - b_{AB(11)} - b_{AC(11)} - b_{BC(11)} \\
&= \log(573) - \log(109) - 0,8770 - (-1,4725) - 0,7809 - 1,4216 - (-0,1860) - 0,3569 \\
&= -0,1184 \\
b_{ABC(121)} &= \log y_{121} - \log y_{232} - b_{A(1)} - b_{B(2)} - b_{C(1)} - b_{AB(12)} - b_{AC(11)} - b_{BC(21)} \\
&= \log(872) - \log(109) - 0,8770 - (-0,3346) - 0,7809 - 0,7376 - (-0,1860) - (-0,0500) \\
&= 0,2546 \\
b_{ABC(112)} &= 0; \quad b_{ABC(122)} = 0; \quad b_{ABC(131)} = 0; \quad b_{ABC(132)} = 0; \quad b_{ABC(211)} = 0; \quad b_{ABC(212)} = 0; \\
b_{ABC(221)} &= 0; \quad b_{ABC(222)} = 0; \quad b_{ABC(231)} = 0; \quad b_{ABC(232)} = 0
\end{aligned}$$

Die relevanten Teile des Outputs sind:

```

Data Information
  3513 cases are accepted.
  12 cells are defined.
Variable Information
Factor      Levels  Value
BET         2       Wahlbeteiligung
              1 JA
              2 NEIN
POL_INT     3       Politisches Interesse
              1 stark
              2 mittel
              3 wenig

```

5 Modellierung multinomialer Daten (log-lineare Modelle)

```
WOHNGBE      2      Wohngebiet
                1 Alte Bundesl\{"a}nder
                2 Neue Bundesl\{"a}nder
```

Model and Design Information

Model: Multinomial

Design: Constant + BET + POL_INT + WOHNGBE + BET*POL_INT + BET*WOHNGBE +
POL_INT
*WOHNGBE + BET*POL_INT*WOHNGBE

Correspondence Between Parameters and Terms of the Design

Parameter	Aliased	Term
1		Constant
2		[BET = 1]
3	x	[BET = 2]
4		[POL_INT = 1]
5		[POL_INT = 2]
6	x	[POL_INT = 3]
7		[WOHNGBE = 1]
8	x	[WOHNGBE = 2]
9		[BET = 1]*[POL_INT = 1]
10		[BET = 1]*[POL_INT = 2]
11	x	[BET = 1]*[POL_INT = 3]
12	x	[BET = 2]*[POL_INT = 1]
13	x	[BET = 2]*[POL_INT = 2]
14	x	[BET = 2]*[POL_INT = 3]
15		[BET = 1]*[WOHNGBE = 1]
16	x	[BET = 1]*[WOHNGBE = 2]
17	x	[BET = 2]*[WOHNGBE = 1]
18	x	[BET = 2]*[WOHNGBE = 2]
19		[POL_INT = 1]*[WOHNGBE = 1]
20	x	[POL_INT = 1]*[WOHNGBE = 2]
21		[POL_INT = 2]*[WOHNGBE = 1]
22	x	[POL_INT = 2]*[WOHNGBE = 2]
23	x	[POL_INT = 3]*[WOHNGBE = 1]
24	x	[POL_INT = 3]*[WOHNGBE = 2]
25		[BET = 1]*[POL_INT = 1]*[WOHNGBE = 1]
26	x	[BET = 1]*[POL_INT = 1]*[WOHNGBE = 2]
27		[BET = 1]*[POL_INT = 2]*[WOHNGBE = 1]
28	x	[BET = 1]*[POL_INT = 2]*[WOHNGBE = 2]
29	x	[BET = 1]*[POL_INT = 3]*[WOHNGBE = 1]
30	x	[BET = 1]*[POL_INT = 3]*[WOHNGBE = 2]
31	x	[BET = 2]*[POL_INT = 1]*[WOHNGBE = 1]
32	x	[BET = 2]*[POL_INT = 1]*[WOHNGBE = 2]
33	x	[BET = 2]*[POL_INT = 2]*[WOHNGBE = 1]
34	x	[BET = 2]*[POL_INT = 2]*[WOHNGBE = 2]
35	x	[BET = 2]*[POL_INT = 3]*[WOHNGBE = 1]
36	x	[BET = 2]*[POL_INT = 3]*[WOHNGBE = 2]

Note: 'x' indicates an aliased (or a redundant) parameter.
These parameters are set to zero.

Parameter Estimates

Constant	Estimate
1	4,6913

Note: Constant is not a parameter under multinomial assumption.
Therefore, standard errors are not calculated.

5.8 Log-lineare Modelle für dreidimensionale Kontingenztabellen

Parameter	Estimate	SE	Z-value	Asymptotic 95% CI	
				Lower	Upper
2	,8770	,1240	7,69	,65	1,10
3	,0000	,	,	,	,
4	-1,4725	,2218	-6,64	-1,91	-1,04
5	-,3346	,1483	-2,26	-,63	-,04
6	,0000	,	,	,	,
7	,7809	,1157	6,75	,55	1,01
8	,0000	,	,	,	,
9	1,4216	,2388	5,95	,95	1,89
10	,7376	,1684	4,38	,41	1,07
11	,0000	,	,	,	,
12	,0000	,	,	,	,
13	,0000	,	,	,	,
14	,0000	,	,	,	,
15	-,1860	,1389	-1,34	-,46	,09
16	,0000	,	,	,	,
17	,0000	,	,	,	,
18	,0000	,	,	,	,
19	,3569	,2573	1,39	-,15	,86
20	,0000	,	,	,	,
21	-,0500	,1799	-,28	-,40	,30
22	,0000	,	,	,	,
23	,0000	,	,	,	,
24	,0000	,	,	,	,
25	-,1184	,2791	-,42	-,67	,43
26	,0000	,	,	,	,
27	,2546	,2049	1,24	-,15	,66
28	,0000	,	,	,	,
29	,0000	,	,	,	,
30	,0000	,	,	,	,
31	,0000	,	,	,	,
32	,0000	,	,	,	,
33	,0000	,	,	,	,
34	,0000	,	,	,	,
35	,0000	,	,	,	,
36	,0000	,	,	,	,

Die Interaktionsparameter AC, BC und ABC sind zum 5% Niveau nicht signifikant, da ihre 95% Konfidenzintervalle die Null einschließen. Es kann offensichtlich ein (parameter-)sparsames Modell geschätzt werden.

Modell (A|B|C)

Die geschätzte link Funktion ist gemäß (5.95) gegeben mit

$$\hat{\eta}_{jkl} = b + b_{A(1)} + b_{B(1)} + b_{B(2)} + b_{C(1)}.$$

Die Anzahl der zu schätzenden Parameter beträgt $J + K + L - 2 = 2 + 3 + 2 - 2 = 5$ und damit die Anzahl der Freiheitsgrade $f = JKL - (J + K + L - 2) = 12 - 5 = 7$.

Wichtige Informationen aus dem SPSS-Output:

Model and Design Information

Model: Multinomial

Design: Constant + BET + POL_INT + WOHNGBE

Correspondence Between Parameters and Terms of the Design

Parameter	Aliased	Term
1		Constant
2		[BET = 1]
3	x	[BET = 2]
4		[POL_INT = 1]
5		[POL_INT = 2]
6	x	[POL_INT = 3]
7		[WOHNGBE = 1]
8	x	[WOHNGBE = 2]

Note: 'x' indicates an aliased (or a redundant) parameter.
These parameters are set to zero.

Goodness-of-fit Statistics

	Chi-Square	DF	Sig.
Likelihood Ratio	163,5261	7	6,E-32
Pearson	167,3847	7	9,E-33

Parameter Estimates

Constant	Estimate
1	4,2133

Note: Constant is not a parameter under multinomial assumption.
Therefore, standard errors are not calculated.

Parameter	Estimate	SE	Z-value	Asymptotic 95% CI	
				Lower	Upper
2	1,4089	,0425	33,17	1,33	1,49
3	,0000	,	,	,	,
4	-,1586	,0448	-3,54	-,25	-,07
5	-,3275	,0398	8,22	,25	,41
6	,0000	,	,	,	,
7	,7658	,0362	21,13	,69	,84
8	,0000	,	,	,	,

Pearson's Chi-Quadrat-Teststatistik und die log-likelihood ratio Statistik zeigen an, dass dieses Modell keine ausreichende Anpassung an die Daten liefert. Jedoch sind die Parameter der Haupteffekte zum 5% Niveau signifikant.

Modell (BC|A)

Die geschätzte link Funktion ist gemäß (5.99) gegeben mit

$$\hat{\eta}_{jkl} = b + b_{A(1)} + b_{B(1)} + b_{B(2)} + b_{C(1)} + b_{BC(11)} + b_{AC(21)}.$$

Die Anzahl der zu schätzenden Parameter beträgt $J + KL - 1 = 2 + 6 - 1 = 7$ und die Anzahl der Freiheitsgrade $f = (KL - 1)(J - 1) = (6-1)(2-1) = 5$.

Wichtige Informationen aus dem SPSS-Output:

Model and Design Information

Model: Multinomial

Design: Constant + BET + POL_INT + WOHNGBE + POL_INT*WOHNGBE

Correspondence Between Parameters and Terms of the Design

Parameter	Aliased	Term
1		Constant
2		[BET = 1]
3	x	[BET = 2]
4		[POL_INT = 1]
5		[POL_INT = 2]
6	x	[POL_INT = 3]
7		[WOHNGBE = 1]
8	x	[WOHNGBE = 2]
9		[POL_INT = 1]*[WOHNGBE = 1]
10	x	[POL_INT = 1]*[WOHNGBE = 2]
11		[POL_INT = 2]*[WOHNGBE = 1]
12	x	[POL_INT = 2]*[WOHNGBE = 2]
13	x	[POL_INT = 3]*[WOHNGBE = 1]
14	x	[POL_INT = 3]*[WOHNGBE = 2]

Note: 'x' indicates an aliased (or a redundant) parameter.
These parameters are set to zero.

Goodness-of-fit Statistics

	Chi-Square	DF	Sig.
Likelihood Ratio	158,3702	5	2,E-32
Pearson	164,0702	5	2,E-33

Parameter Estimates

Constant	Estimate
1	4,2877

Note: Constant is not a parameter under multinomial assumption.
Therefore, standard errors are not calculated.

Parameter	Estimate	SE	Z-value	Asymptotic 95% CI	
				Lower	Upper
2	1,4089	,0425	33,17	1,33	1,49
3	,0000	,	,	,	,
4	-,3031	,0797	-3,80	-,46	-,15
5	,2365	,0694	3,41	,10	,37
6	,0000	,	,	,	,
7	,6533	,0640	10,21	,53	,78
8	,0000	,	,	,	,
9	,2121	,0964	2,20	,02	,40
10	,0000	,	,	,	,
11	,1352	,0848	1,59	-,03	,30
12	,0000	,	,	,	,
13	,0000	,	,	,	,
14	,0000	,	,	,	,

Die Gültigkeit dieses Modells bedeutet, dass eine Assoziation zwischen der Variablen politisches Interesse und der Variablen Wohngebiet bestehen kann, während die Variable Wahlbeteiligung von dem politischen Interesse und dem Wohngebiet unabhängig

ist. Diese Hypothese muss jedoch auf dem 5% Niveau abgelehnt werden, denn Pearson's Chi-Quadrat-Teststatistik und die log-likelihood ratio Statistik zeigen keine ausreichende Anpassung an die Daten der Kontingenztabelle an. Ausserdem ist der Interaktionsparameter $b_{BC(21)}$ nicht signifikant zum 5% Niveau.

Modell (AC|B)

Die geschätzte link Funktion ist gemäß (5.103) gegeben mit

$$\hat{\eta}_{jkl} = b + b_{A(1)} + b_{B(1)} + b_{B(2)} + b_{C(1)} + b_{AC(11)}.$$

Die Anzahl der zu schätzenden Parameter beträgt $J + KL - 1 = 3 + 4 - 1 = 6$ und die Anzahl der Freiheitsgrade $f = (JL - 1)(K - 1) = (4-1)(3-1) = 6$.

Wichtige Informationen aus dem SPSS-Output:

Model and Design Information

Model: Multinomial

Design: Constant + BET + POL_INT + WOHNGBE + BET*POL_INT

Correspondence Between Parameters and Terms of the Design

Parameter	Aliased	Term
1		Constant
2		[BET = 1]
3	x	[BET = 2]
4		[POL_INT = 1]
5		[POL_INT = 2]
6	x	[POL_INT = 3]
7		[WOHNGBE = 1]
8	x	[WOHNGBE = 2]
9		[BET = 1]*[WOHNGBE = 1]
10	x	[BET = 1]*[WOHNGBE = 2]
11	x	[BET = 2]*[WOHNGBE = 1]
12	x	[BET = 2]*[WOHNGBE = 2]

Note: 'x' indicates an aliased (or a redundant) parameter.
These parameters are set to zero.

Goodness-of-fit Statistics

	Chi-Square	DF	Sig.
Likelihood Ratio	163,1163	6	2,E-32
Pearson	167,3941	6	2,E-33

Parameter Estimates

Constant	Estimate
1	4,1808

Note: Constant is not a parameter under multinomial assumption.
Therefore, standard errors are not calculated.

Parameter	Estimate	SE	Z-value	Asymptotic 95% CI Lower	Upper
2	1,4491	,0763	18,99	1,30	1,60
3	,0000	,	,	,	,

4	-,1586	,0448	-3,54	-,25	-,07
5	,3275	,0398	8,22	,25	,41
6	,0000	,	,	,	,
7	,8130	,0825	9,85	,65	,97
8	,0000	,	,	,	,
9	-,0587	,0919	-,64	-,24	,12
10	,0000	,	,	,	,
11	,0000	,	,	,	,
12	,0000	,	,	,	,

Die Gültigkeit dieses Modells bedeutet, dass eine Assoziation zwischen der Variablen Wahlbeteiligung und der Variablen Wohngebiet bestehen kann, während die Variable politisches Interesse von der Wahlbeteiligung und dem Wohngebiet unabhängig ist. Diese Hypothese muss jedoch auf dem 5% Niveau abgelehnt werden, denn Pearson's Chi-Quadrat-Teststatistik und die log-likelihood ratio Statistik zeigen keine ausreichende Anpassung an die Daten der Kontingenztabelle an. Ausserdem ist der Interaktionsparameter $b_{AC(11)}$ nicht signifikant zum 5% Niveau.

Modell (AB|C)

Die geschätzte link Funktion ist gemäß (5.107) gegeben mit

$$\hat{\eta}_{jkl} = b + b_{A(1)} + b_{B(1)} + b_{B(2)} + b_{C(1)} + b_{AB(11)} + b_{AB(12)}.$$

Die Anzahl der zu schätzenden Parameter beträgt $L + JK - 1 = 2 + 6 - 1 = 7$ und die Anzahl der Freiheitsgrade $f = (JK - 1)(L - 1) = (6-1)(2-1) = 5$.

Wichtige Informationen aus dem SPSS-Output:

```
Model and Design Information
  Model: Multinomial
Design: Constant + BET + POL_INT + WOHNGBE + BET*POL_INT
Correspondence Between Parameters and Terms of the Design
Parameter    Aliased    Term
1             Constant
2             [BET = 1]
3             x      [BET = 2]
4             [POL_INT = 1]
5             [POL_INT = 2]
6             x      [POL_INT = 3]
7             [WOHNGBE = 1]
8             x      [WOHNGBE = 2]
9             [BET = 1]*[POL_INT = 1]
10            [BET = 1]*[POL_INT = 2]
11            x      [BET = 1]*[POL_INT = 3]
12            x      [BET = 2]*[POL_INT = 1]
13            x      [BET = 2]*[POL_INT = 2]
```

5 Modellierung multinomialer Daten (log-lineare Modelle)

```

14      x      [BET = 2]*[POL_INT = 3]
Note: 'x' indicates an aliased (or a redundant) parameter.
      These parameters are set to zero.

```

Goodness-of-fit Statistics

	Chi-Square	DF	Sig.
Likelihood Ratio	8,8163	5	,1166
Pearson	8,7737	5	,1184

Parameter Estimates

Constant	Estimate
1	4,7017

Note: Constant is not a parameter under multinomial assumption.
Therefore, standard errors are not calculated.

Parameter	Estimate	SE	Z-value	Asymptotic 95% CI	
				Lower	Upper
2	,7533	,0651	11,57	,63	,88
3	,0000	,	,	,	,
4	-1,2146	,1122	-10,82	-1,43	-,99
5	-,3687	,0840	-4,39	-,53	-,20
6	,0000	,	,	,	,
7	,7658	,0362	21,13	,69	,84
8	,0000	,	,	,	,
9	1,3237	,1231	10,75	1,08	1,57
10	,9081	,0959	9,47	,72	1,10
11	,0000	,	,	,	,
12	,0000	,	,	,	,
13	,0000	,	,	,	,
14	,0000	,	,	,	,

Die Gültigkeit dieses Modells bedeutet, dass eine Assoziation zwischen der Variablen Wahlbeteiligung und der Variablen politisches Interesse bestehen kann, während die Variable Wohngebiet von der Wahlbeteiligung und dem politischen Interesse unabhängig ist. Diese Hypothese kann auf dem 5% Niveau nicht abgelehnt werden, wie Pearson's Chi-Quadrat-Teststatistik und die log-likelihood ratio Statistik anzeigen. Alle Parameter sind signifikant zum 5% Niveau.

Modell (AB|AC)

Die geschätzte link Funktion ist gemäß (5.111) gegeben mit

$$\hat{\eta}_{jkl} = b + b_{A(1)} + b_{B(1)} + b_{B(2)} + b_{C(1)} + b_{AB(11)} + b_{AB(12)} + b_{AC(11)}.$$

Die Anzahl der zu schätzenden Parameter beträgt $J(K + L - 1) = 2(3 + 2 - 1) = 8$ und die Anzahl der Freiheitsgrade $f = J(K - 1)(L - 1) = 2(3-1)(2-1) = 4$.

5.8 Log-lineare Modelle für dreidimensionale Kontingenztabellen

Wichtige Informationen aus dem SPSS-Output:

Model and Design Information

Model: Multinomial

Design: Constant + BET + POL_INT + WOHN GEB + POL_INT*WOHN GEB + BET*WOHN GEB

Correspondence Between Parameters and Terms of the Design

Parameter	Aliased	Term
1		Constant
2		[BET = 1]
3	x	[BET = 2]
4		[POL_INT = 1]
5		[POL_INT = 2]
6	x	[POL_INT = 3]
7		[WOHN GEB = 1]
8	x	[WOHN GEB = 2]
9		[BET = 1]*[POL_INT = 1]
10		[BET = 1]*[POL_INT = 2]
11	x	[BET = 1]*[POL_INT = 3]
12	x	[BET = 2]*[POL_INT = 1]
13	x	[BET = 2]*[POL_INT = 2]
14	x	[BET = 2]*[POL_INT = 3]
15		[BET = 1]*[WOHN GEB = 1]
16	x	[BET = 1]*[WOHN GEB = 2]
17	x	[BET = 2]*[WOHN GEB = 1]
18	x	[BET = 2]*[WOHN GEB = 2]

Note: 'x' indicates an aliased (or a redundant) parameter.
These parameters are set to zero.

Goodness-of-fit Statistics

	Chi-Square	DF	Sig.
Likelihood Ratio	8,4066	4	,7078
Pearson	8,3840	4	,0785

Parameter Estimates

Constant	Estimate
1	4,6692

Note: Constant is not a parameter under multinomial assumption.
Therefore, standard errors are not calculated.

Parameter	Estimate	SE	Z-value	Asymptotic 95% CI	
				Lower	Upper
2	,7935	,0909	8,73	,62	,97
3	,0000	,	,	,	,
4	-1,2146	,1122	-10,82	-1,43	-,99
5	-,3687	,0840	-4,39	-,53	-,20
6	,0000	,	,	,	,
7	,8130	,0825	9,85	,65	,97
8	,0000	,	,	,	,
9	1,3237	,1231	10,75	1,08	1,57
10	,9081	,0959	9,47	,72	1,10
11	,0000	,	,	,	,
12	,0000	,	,	,	,
13	,0000	,	,	,	,
14	,0000	,	,	,	,
15	-,0587	,0919	-,64	-,24	,12
16	,0000	,	,	,	,
17	,0000	,	,	,	,
18	,0000	,	,	,	,

Die Gültigkeit dieses Modells bedeutet, dass die Variable politisches Interesse unabhängig ist von der Variablen Wohngebiet für jede Kategorie der Variablen Wahlbeteiligung. Diese Hypothese kann auf dem 5% Niveau nicht abgelehnt werden, wie Pearson's Chi-Quadrat-Teststatistik und die log-likelihood ratio Statistik anzeigen. Jedoch ist der Parameter $b_{AC(11)}$ nicht signifikant zum 5% Niveau.

Modell (AC|BC)

Die geschätzte link Funktion ist gemäß (5.115) gegeben mit

$$\hat{\eta}_{jkl} = b + b_{A(1)} + b_{B(1)} + b_{B(2)} + b_{C(1)} + b_{AC(11)} + b_{BC(11)} + b_{BC(21)}.$$

Die Anzahl der zu schätzenden Parameter beträgt $L(J + K - 1) = 2(2 + 3 - 1) = 8$ und die Anzahl der Freiheitsgrade $f = L(J - 1)(K - 1) = 2(2-1)(3-1) = 4$.

Wichtige Informationen aus dem SPSS-Output:

Model and Design Information

Model: Multinomial

Design: Constant + BET + POL_INT + WOHNGBE + BET*WOHNGBE + POL_INT*WOHNGBE

Correspondence Between Parameters and Terms of the Design

Parameter	Aliased	Term
1		Constant
2		[BET = 1]
3	x	[BET = 2]
4		[POL_INT = 1]
5		[POL_INT = 2]
6	x	[POL_INT = 3]
7		[WOHNGBE = 1]
8	x	[WOHNGBE = 2]
9		[BET = 1]*[WOHNGBE = 1]
10	x	[BET = 1]*[WOHNGBE = 2]
11	x	[BET = 2]*[WOHNGBE = 1]
12	x	[BET = 2]*[WOHNGBE = 2]
13		[POL_INT = 1]*[WOHNGBE = 1]
14	x	[POL_INT = 1]*[WOHNGBE = 2]
15		[POL_INT = 2]*[WOHNGBE = 1]
16	x	[POL_INT = 2]*[WOHNGBE = 2]
17	x	[POL_INT = 3]*[WOHNGBE = 1]
18	x	[POL_INT = 3]*[WOHNGBE = 2]

Note: 'x' indicates an aliased (or a redundant) parameter.
These parameters are set to zero.

Goodness-of-fit Statistics

	Chi-Square	DF	Sig.
Likelihood Ratio	157,9604	4	4,E-33
Pearson	163,4394	4	3,E-34

Parameter Estimates

Constant Estimate

5.8 Log-lineare Modelle für dreidimensionale Kontingenztabellen

1 4,2562

Note: Constant is not a parameter under multinomial assumption.

Therefore, standard errors are not calculated.

Parameter	Estimate	SE	Z-value	Asymptotic 95% CI	
				Lower	Upper
2	1,4491	,0763	18,99	1,30	1,60
3	,0000	,	,	,	,
4	-,3031	,0797	-3,80	-,46	-,15
5	,2365	,0694	3,41	,10	,37
6	,0000	,	,	,	,
7	,7005	,0979	7,15	,51	,89
8	,0000	,	,	,	,
9	-,0587	,0919	-,64	-,24	,12
10	,0000	,	,	,	,
11	,0000	,	,	,	,
12	,0000	,	,	,	,
13	,2121	,0964	2,20	,02	,40
14	,0000	,	,	,	,
15	,1352	,0848	1,59	-,03	,30
16	,0000	,	,	,	,
17	,0000	,	,	,	,
18	,0000	,	,	,	,

Die Gültigkeit dieses Modells bedeutet, dass die Variable Wahlbeteiligung unabhängig ist von der Variablen politisches Interesse für jede Kategorie der Variablen Wohngebiet. Diese Hypothese muss auf dem 5% Niveau abgelehnt werden, denn Pearson's Chi-Quadrat-Teststatistik und die log-likelihood ratio Statistik zeigen eine unbefriedigende Anpassung des Modells an die Daten der Kontingenztafel. Ausserdem sind die Parameter $b_{AC(11)}$ und $b_{BC(21)}$ nicht signifikant zum 5% Niveau.

Modell (AB|BC)

Die geschätzte link Funktion ist gemäß (5.119) gegeben mit

$$\hat{\eta}_{jkl} = b + b_{A(1)} + b_{B(1)} + b_{B(2)} + b_{C(1)} + b_{AB(11)} + b_{AB(12)} + b_{BC(11)} + b_{BC(21)}.$$

Die Anzahl der zu schätzenden Parameter beträgt $K(J + L - 1) = 3(2 + 2 - 1) = 9$ und die Anzahl der Freiheitsgrade $f = K(J - 1)(L - 1) = 3(2-1)(2-1) = 3$.

Wichtige Informationen aus dem SPSS-Output:

Model and Design Information

Model: Multinomial

Design: Constant + BET + POL_INT + WOHNGBE + POL_INT*WOHNGBE + POL_INT*WOHNGBE

Correspondence Between Parameters and Terms of the Design

Parameter	Aliased	Term
1		Constant
2		[BET = 1]

5 Modellierung multinomialer Daten (log-lineare Modelle)

3	x	[BET = 2]
4		[POL_INT = 1]
5		[POL_INT = 2]
6	x	[POL_INT = 3]
7		[WOHNGBE = 1]
8	x	[WOHNGBE = 2]
9		[BET = 1]*[POL_INT = 1]
10		[BET = 1]*[POL_INT = 2]
11	x	[BET = 1]*[POL_INT = 3]
12	x	[BET = 2]*[POL_INT = 1]
13	x	[BET = 2]*[POL_INT = 2]
14	x	[BET = 2]*[POL_INT = 3]
15		[POL_INT = 1]*[WOHNGBE = 1]
16	x	[POL_INT = 1]*[WOHNGBE = 2]
17		[POL_INT = 2]*[WOHNGBE = 1]
18	x	[POL_INT = 2]*[WOHNGBE = 2]
19	x	[POL_INT = 3]*[WOHNGBE = 1]
20	x	[POL_INT = 3]*[WOHNGBE = 2]

Note: 'x' indicates an aliased (or a redundant) parameter.
These parameters are set to zero.

Goodness-of-fit Statistics

	Chi-Square	DF	Sig.
Likelihood Ratio	3,6604	3	,3006
Pearson	3,5929	3	,3089

Parameter Estimates

Constant	Estimate
1	4,7771

Note: Constant is not a parameter under multinomial assumption.
Therefore, standard errors are not calculated.

Parameter	Estimate	SE	Z-value	Asymptotic 95% CI	
				Lower	Upper
2	,7533	,0651	11,57	,63	,88
3	,0000	,	,	,	,
4	-1,3591	,1301	-10,44	-1,61	-1,10
5	-,5496	,1014	-4,53	-,66	-,26
6	,0000	,	,	,	,
7	,6533	,0640	10,21	,53	,78
8	,0000	,	,	,	,
9	1,3237	,1231	10,21	1,08	1,57
10	,9081	,0959	9,47	,72	1,10
11	,0000	,	,	,	,
12	,0000	,	,	,	,
13	,0000	,	,	,	,
14	,0000	,	,	,	,
15	,2121	,0964	2,20	,02	,40
16	,0000	,	,	,	,
17	,1352	,0848	1,59	-,03	,30
18	,0000	,	,	,	,
19	,0000	,	,	,	,
20	,0000	,	,	,	,

Die Gültigkeit dieses Modells bedeutet, dass die Variable Wahlbeteiligung unabhängig ist von der Variablen Wohngebiet für jede Kategorie der Variablen politisches Interesse.

Diese Hypothese kann auf dem 5% Niveau nicht abgelehnt werden, wie Pearson's Chi-Quadrat-Teststatistik und die log-likelihood ratio Statistik anzeigen. Allerdings ist der Parameter $b_{BC(21)}$ nicht signifikant zum 5% Niveau.

Modell (AB|AC|BC)

Die geschätzte link Funktion ist gemäß (5.121) gegeben mit

$$\hat{\eta}_{jkl} = b + b_{A(1)} + b_{B(1)} + b_{B(2)} + b_{C(1)} + b_{AB(11)} + b_{AB(12)} + b_{AC(11)} + b_{BC(11)} + b_{BC(21)}.$$

Die Anzahl der zu schätzenden Parameter beträgt $JK + JL + KL - J - K - L + 1 = 2 \cdot 3 + 2 \cdot 2 + 3 \cdot 2 - 2 - 3 - 2 + 1 = 10$

und die Anzahl der Freiheitsgrade $f = (J - 1)(K - 1)(L - 1) = (2-1)(3-1)(2-1) = 2$.

Wichtige Informationen aus dem SPSS-Output:

Model and Design Information

Model: Multinomial

Design: Constant + BET + POL_INT + WOHN GEB + BET*POL_INT + BET*WOHN GEB + POL_INT

*WOHN GEB

Correspondence Between Parameters and Terms of the Design

Parameter	Aliased	Term
1		Constant
2		[BET = 1]
3	x	[BET = 2]
4		[POL_INT = 1]
5		[POL_INT = 2]
6	x	[POL_INT = 3]
7		[WOHN GEB = 1]
8	x	[WOHN GEB = 2]
9		[BET = 1]*[POL_INT = 1]
10		[BET = 1]*[POL_INT = 2]
11	x	[BET = 1]*[POL_INT = 3]
12	x	[BET = 2]*[POL_INT = 1]
13	x	[BET = 2]*[POL_INT = 2]
14	x	[BET = 2]*[POL_INT = 3]
15		[BET = 1]*[WOHN GEB = 1]
16	x	[BET = 1]*[WOHN GEB = 2]
17	x	[BET = 2]*[WOHN GEB = 1]
18	x	[BET = 2]*[WOHN GEB = 2]
19		[POL_INT = 1]*[WOHN GEB = 1]
20	x	[POL_INT = 1]*[WOHN GEB = 2]
21		[POL_INT = 2]*[WOHN GEB = 1]
22	x	[POL_INT = 2]*[WOHN GEB = 2]
23	x	[POL_INT = 3]*[WOHN GEB = 1]
24	x	[POL_INT = 3]*[WOHN GEB = 2]

Note: 'x' indicates an aliased (or a redundant) parameter.

These parameters are set to zero.

Goodness-of-fit Statistics

Chi-Square	DF	Sig.
------------	----	------

5 Modellierung multinomialer Daten (log-lineare Modelle)

Likelihood Ratio	2,3472	2	,3092
Pearson	2,3520	2	,3085

Parameter Estimates	
Constant	Estimate
1	4,7282

Note: Constant is not a parameter under multinomial assumption.
Therefore, standard errors are not calculated.

Parameter	Estimate	SE	Z-value	Asymptotic 95% CI	
				Lower	Upper
2	,8244	,0904	9,12	,65	1,00
3	,0000	,	,	,	,
4	-1,3784	,1321	-10,43	-1,64	-1,12
5	-,4739	,0910	-4,60	-,68	-,27
6	,0000	,	,	,	,
7	,7267	,0910	7,98	,55	,91
8	,0000	,	,	,	,
9	1,3291	,1233	10,78	1,09	1,57
10	,9117	,0960	9,50	,72	1,10
11	,0000	,	,	,	,
12	,0000	,	,	,	,
13	,0000	,	,	,	,
14	,0000	,	,	,	,
15	-,1074	,0941	-1,14	-,29	,08
16	,0000	,	,	,	,
17	,0000	,	,	,	,
18	,0000	,	,	,	,
19	,2344	,0983	2,38	,04	,43
20	,0000	,	,	,	,
21	,1523	,0861	1,77	-,02	,32
22	,0000	,	,	,	,
23	,0000	,	,	,	,
24	,0000	,	,	,	,

Die Gültigkeit dieses Modells bedeutet, dass die Assoziation zwischen den Variablen Wahlbeteiligung und politisches Interesse unabhängig ist von der Variablen Wohngebiet, die Assoziation zwischen den Variablen Wahlbeteiligung und Wohngebiet unabhängig ist von der Variablen politisches Interesse sowie die Assoziation zwischen den Variablen politisches Interesse und Wohngebiet unabhängig ist von der Variablen Wahlbeteiligung. Diese Hypothese kann auf dem 5% Niveau nicht abgelehnt werden, wie Pearson's Chi-Quadrat-Teststatistik und die log-likelihood ratio Statistik anzeigen. Allerdings sind die Parameter $b_{AC(11)}$ und $b_{BC(21)}$ nicht signifikant zum 5% Niveau.

Zusammenfassende Beurteilung:

Für die Modelle AB|C, AB|AC, AB|BC und AB|AC|BC wurde die Nullhypothese einer ausreichenden Anpassung des Modells an die Daten der Kontingenztafel auf dem 5% Niveau nicht abgelehnt. Vergleicht man diese Modelle unter den Aspekten der Signifikanz

der Parameter, der Anpassungsgüte (Pearson's Chi-Quadrat Teststatistik) sowie der Parametersparsamkeit, so zeigt sich:

- Der einzige Parameter für die Interaktion zwischen Wahlbeteiligung und Wohngebiet $b_{AC(11)}$ in den Modellen $AB|AC$ und $AB|AC|BC$ ist nicht signifikant. Da das Modell $AB|AC|BC$ jedoch gegenüber dem Modell $AB|AC$ einen weiteren signifikanten Parameter für die Interaktion zwischen politischem Interesse und Wohngebiet enthält, kann das Modell $AB|AC$ aus der weiteren Betrachtung herausgenommen werden.
- Die Modelle $AB|BC$ und $AB|AC|BC$ unterscheiden sich hinsichtlich der Anpassungsgüte nur geringfügig. Da es sich bei diesen beiden Modellen um nested Modelle handelt, denn das Modell $AB|BC$ ist vollständig in dem Modell $AB|AC|BC$ enthalten, kann ΔD nach (5.123) berechnet werden:

$$\Delta D = D(AB|BC) - D(AB|AC|BC) = 3,6604 - 2,3472 = 1,3132$$

mit $f(\Delta D) = f(AB|BC) - f(AB|AC|BC) = 3 - 2 = 1$. Aus der Tabelle der Verteilungsfunktion der Chi-Quadrat-Verteilung findet man $\chi_{0,95;1}^2 = 3,84$, so dass kein signifikanter Unterschied zum 5% Niveau vorliegt. Ausserdem ist der zusätzliche Parameter $b_{AC(11)}$ im Modell $AB|AC|BC$ nicht signifikant. Das Modell $AB|AC|BC$ kann aus der weiteren Betrachtung herausgenommen werden.

- Das Modell $AB|BC$ enthält gegenüber dem Modell $AB|C$ zusätzlich die Interaktion zwischen politischem Interesse und Wohngebiet, wobei der Parameter $b_{BC(11)}$ signifikant, der Parameter $b_{BC(21)}$ jedoch nicht signifikant ist. Da es sich bei diesen beiden Modellen ebenfalls um nested Modelle handelt, denn das Modell $AB|C$ ist vollständig in dem Modell $AB|BC$ enthalten, ergibt sich nach (5.123) eine Verbesserung der Anpassungsgüte um:

$$\Delta D = D(AB|C) - D(AB|BC) = 8,8163 - 3,6604 = 5,1559$$

mit $f(\Delta D) = f(AB|C) - f(AB|BC) = 5 - 3 = 2$ Freiheitsgraden. Dieser Unterschied ist aber zum 5% Niveau nicht signifikant, denn aus der Tabelle der Verteilungsfunktion der Chi-Quadrat-Verteilung findet man $\chi_{0,95;2}^2 = 5,99$. Das Modell $AB|BC$ kann aus der weiteren Betrachtung herausgenommen werden.

Es verbleibt das Modell $AB|C$. Die inhaltliche Schlussfolgerung ist, dass die Variable Wohngebiet unabhängig von den Variablen Wahlbeteiligung und politisches Interesse ist, während die Variablen Wahlbeteiligung und politisches Interesse Interaktionen aufweisen.

Diese Modellauswahl kann auch mittels der Prozedur “Loglinear Model Selection“ unter SPSS Windows Release 10.0.7⁴⁴ unterstützt werden. Diese Prozedur dient der Auswahl derjenigen Terme hierarchischer log-linearer Modelle, die für eine zufriedenstellende Anpassung der Daten unbedingt erforderlich sind. Die relevanten Teile des SPSS-Outputs sind nachfolgend angegeben.

DESIGN 1 has generating class

```
BET*POL_INT*WOHNGBE
Goodness-of-fit test statistics
Likelihood ratio chi-square =      ,00000    DF = 0    P =      1,000
Pearson chi-square =      ,00000    DF = 0    P =      1,000
```

Tests that K-way and higher order effects are zero.

K	DF	L.R. Chisq	Prob	Pearson Chisq	Prob	Iteration
3	2	2,347	,3093	2,352	,0385	3
2	7	163,526	,0000	167,385	,0000	2
1	11	2181,732	,0000	2315,212	,0000	0

Tests that K-way effects are zero.

K	DF	L.R. Chisq	Prob	Pearson Chisq	Prob	Iteration
3	4	2018,206	,0000	2147,827	,0000	0
2	5	161,179	,0000	165,033	,0000	0
1	2	2,347	,3093	2,352	,3085	0

Tests of PARTIAL associations.

Effect Name	DF	Partial Chisq	Prob	Iter
BET*POL_INT	2	155,613	,0000	2
BET*WOHNGBE	1	1,313	,2518	2
POL_INT*WOHNGBE	2	6,060	,0483	2
BET	1	1389,443	,0000	2
POL_INT	2	149,176	,0000	2
WOHNGBE	1	479,586	,0000	2

Backward Elimination (p = ,050) for DESIGN 1 with generating class

⁴⁴Siehe SPSS Advanced Models 9.0, S. 41 ff. und S. 139 ff.

5.8 Log-lineare Modelle für dreidimensionale Kontingenztafeln

```

BET*POL_INT*WOHNGB
Likelihood ratio chi square =      ,00000    DF = 0    P =      1,000
If Deleted Simple Effect is      DF    L.R. Chisq Change      Prob    Iter
BET*POL_INT*WOHNGB                2                2,347      ,3093      3

```

Step 1

The best model has generating class

```

BET*POL_INT
BET*WOHNGB
POL_INT*WOHNGB

```

```

Likelihood ratio chi square =      3,34713    DF = 2    P =      ,309
If Deleted Simple Effect is      DF    L.R. Chisq Change      Prob    Iter
BET*POL_INT                      2                155,613      ,0000      2
BET*WOHNGB                      1                1,313      ,2518      2
POL_INT*WOHNGB                  2                6,060      ,0483      2

```

Step 2

The best model has generating class

```

BET*POL_INT
POL_INT*WOHNGB

```

```

Likelihood ratio chi square =      3,66035    DF = 3    P =      ,301
If Deleted Simple Effect is      DF    L.R. Chisq Change      Prob    Iter
BET*POL_INT                      2                154,710      ,0000      2
POL_INT*WOHNGB                  2                5,156      ,0759      2

```

Step 3

The best model has generating class

```

BET*POL_INT
WOHNGB

```

```

Likelihood ratio chi square =      8,81626    DF = 5    P =      ,117
If Deleted Simple Effect is      DF    L.R. Chisq Change      Prob    Iter
BET*POL_INT                      2                154,710      ,0000      2
WOHNGB                          1                479,586      ,0000      2

```

Step 4

The best model has generating class

```

BET*POL_INT
WOHNGB

```

```

Likelihood ratio chi square =      8,81626    DF = 5    P =      ,117

```

The final model has generating class

```

BET*POL_INT
WOHNGB

```

Goodness-of-fit test statistics

```

Likelihood ratio chi-square =      8,81626    DF = 5    P =      ,117
Pearson chi-square =          8,77366    DF = 5    P =      ,118

```

Diese Modellselektionsprozedur führt letztendlich zu dem auch oben ausgewählten Modell AB|C.

5.9 Weitere Probleme

Beziehung zwischen log-linearen Modellen und Logit Modellen:

Zur Verdeutlichung dieser Beziehung soll auf das Beispiel 4.1 zurückgegriffen werden.

Es ist für das Jahr 1991:

- A = Y1 - Einschätzung der Wirtschaftslage, $j = 1, \dots, 5$ mit 5. Kategorie als Referenzkategorie,
- B = X1 - Erhebungsgebiet mit 1 - West und 2 - Ost (Referenzkategorie), $k = 1, 2$,
- C = X2 Geschlecht mit 1 = Mann und 2 = Frau (Referenzkategorie), $l = 1, 2$.

y_{jkl} ist die beobachtete Häufigkeit der Kategorienkombination $A_j \cap B_k \cap C_l$. Geeignet darstellen lassen sich diese Häufigkeiten in einer dreidimensionalen $(5 \times 2 \times 2)$ -Kontingenztafel. Die interessierende Frage ist nun die Zusammenhangsstruktur der drei Variablen, aber offensichtlich im Hinblick auf die Variable Einschätzung der Wirtschaftslage (A) als response Variable, da man wissen will, wie die Variablen Erhebungsgebiet (B) und Geschlecht (C) als erklärende Variable die Wahrscheinlichkeit des Auftretens der Kategorien der Einschätzung der Wirtschaftslage beeinflussen. Somit liegt eine asymmetrische Fragestellung vor. Diese lässt sich jedoch auch im Rahmen log-linearer Modelle durch die Analyse der Interaktionsterme, die die Variable A beinhalten (d.h. $b_{AB(jk)}$ und $b_{AC(jl)}$), realisieren. Es kann z.B. das Modell (AB|AC|BC) spezifiziert werden:

$$\log \hat{n}_{jkl} = \beta + \beta_{A(j)} + \beta_{B(k)} + \beta_{C(l)} + \beta_{AB(jk)} + \beta_{AC(jl)} + \beta_{BC(kl)}, \quad (5.121)$$

in dem die bedingte Assoziation zwischen zwei Variablen nicht mehr von der dritten Variablen abhängt. Mit Blickrichtung auf die zu erklärende Variable A bildet man die Differenz

$$\begin{aligned} \log(\hat{m}_{jkl}/\hat{m}_{Jkl}) &= \log m_{jkl} - \log m_{Jkl} \\ &= [b + b_{A(j)} + b_{B(k)} + b_{C(l)} + b_{AB(jk)} + b_{AC(jl)} + b_{BC(kl)}] \\ &\quad - [b + b_{A(J)} + b_{B(k)} + b_{C(l)} + b_{AB(Jk)} + b_{AC(Jl)} + b_{BC(kl)}] \\ &= b_{A(j)} + b_{AB(jk)} + b_{AC(jl)}, \end{aligned}$$

Dieses Ergebnis erhält man wegen der Restriktion $b_{A(J)} = 0$, $b_{AB(Jk)} = 0$ und $b_{AC(Jl)} = 0$ bei Unterstellung der corner-point-Restriktion und damit der Dummy-Kodierung. Entsprechende Ergebnisse resultieren auch bei der Verwendung der sum-to-zero Restriktion und der dazugehörigen Effekt-Kodierung.

Es sind dabei alle Parameter entfallen, die sich ausschließlich auf die erklärenden Variablen beziehen: $b_{B(k)}$, $b_{C(l)}$ und $b_{BC(kl)}$.

Mit $\hat{m}_{jkl}/n = \hat{\pi}_{jkl}$ folgt:

$$\log(\hat{m}_{jkl}/\hat{m}_{Jkl}) = \log(\hat{\pi}_{jkl}/\hat{\pi}_{Jkl}) = b_{A(j)} + b_{AB(jk)} + b_{AC(jl)}.$$

Dies ist das Logit Modell entsprechend (4.4), in dem die logarithmierten odds als Summe von Effekten der erklärenden Variablen dargestellt werden. Dies wird auch deutlich, wenn man berücksichtigt, dass

$$\log \frac{\pi_{jkl}}{\pi_{Jkl}} = \log \frac{P(A_j \cap B_k \cap C_l)}{P(A_J \cap B_k \cap C_l)} = \log \frac{\frac{P(A_j \cap B_k \cap C_l)}{P(B_k \cap C_l)}}{\frac{P(A_J \cap B_k \cap C_l)}{P(B_k \cap C_l)}} = \log \frac{P(A_j|B_k \cap C_l)}{P(A_J|B_k \cap C_l)}$$

gilt. Das Interesse ist also auf die Variable A = Einschätzung der Wirtschaftslage gerichtet, wobei für die Variable B = Erhebungsgebiet und die Variable C = geschlecht eine bestimmte Kategorie vorliegt und die Assoziation zwischen B und C nicht betrachtet wird. Bei den Logit Modellen ist eine Variable abhängig von den anderen Variablen, bei gegebenen Kategorien der erklärenden Variablen.

Die Konstante des Logit Modells ist identisch mit dem Parameter $b_{A(j)}$ im log-linearen Modell, d.h. dem Parameter des Haupteffekts derjenigen Variablen, die im Logit Modell die zu erklärende Variable ist.

Dem Parameter des Einflusses der erklärenden X-Variablen im Logit Modell entspricht der Interaktionsparameter $b_{AB(jk)}$ bzw. $b_{AC(jl)}$ im log-linearen Modell, d.h. der Interaktionsparameter zwischen der zu erklärenden Variablen und der jeweiligen erklärenden X-Variablen des Logit Modells.

Relevante Teile des SPSS-Outputs:

```
Variable Information
Factor      Levels Value
```

5 Modellierung multinomialer Daten (log-lineare Modelle)

```

Y1_91          5          WIRTSCHAFTSLAGE IN DER BRD HEUTE
                    1 sehr gut
                    2 gut
                    3 teils teils
                    4 schlecht
                    5 sehr schlecht
X1_91           2          ERHEBUNGSGEBIET: WEST - OST
                    1 ALTE BUNDESLAENDER
                    2 NEUE BUNDESLAENDER
                    GESCHLECHT, BEFRAGTE
                    1 MANN
                    2 FRAU

```

Model and Design Information

Model: Multinomial

Design: Constant + X1_91 + X2_91 + Y1_91 + X1_91*X2_91 + Y1_91*X1_91 + Y1_91
*X2_91

Correspondence Between Parameters and Terms of the Design

Parameter	Aliased	Term
1		Constant
2		[X1_91 = 1]
3	x	[X1_91 = 2]
4		[X2_91 = 1]
5	x	[X2_91 = 2]
6		[Y1_91 = 1]
7		[Y1_91 = 2]
8		[Y1_91 = 3]
9		[Y1_91 = 4]
10	x	[Y1_91 = 5]
11		[X1_91 = 1]*[X2_91 = 1]
12	x	[X1_91 = 1]*[X2_91 = 2]
13	x	[X1_91 = 2]*[X2_91 = 1]
14	x	[X1_91 = 1]*[X2_91 = 2]
15		[Y1_91 = 1]*[X1_91 = 1]
16	x	[Y1_91 = 1]*[X1_91 = 2]
17		[Y1_91 = 2]*[X1_91 = 1]
18	x	[Y1_91 = 2]*[X1_91 = 2]
19		[Y1_91 = 3]*[X1_91 = 1]
20	x	[Y1_91 = 3]*[X1_91 = 2]
21		[Y1_91 = 4]*[X1_91 = 1]
22	x	[Y1_91 = 4]*[X1_91 = 2]
23	x	[Y1_91 = 5]*[X1_91 = 1]
24	x	[Y1_91 = 5]*[X1_91 = 2]
25		[Y1_91 = 1]*[X2_91 = 1]
26	x	[Y1_91 = 1]*[X2_91 = 2]
27		[Y1_91 = 2]*[X2_91 = 1]
28	x	[Y1_91 = 2]*[X2_91 = 2]
29		[Y1_91 = 3]*[X2_91 = 1]
30	x	[Y1_91 = 3]*[X2_91 = 2]
31		[Y1_91 = 4]*[X2_91 = 1]
32	x	[Y1_91 = 4]*[X2_91 = 2]
33	x	[Y1_91 = 5]*[X2_91 = 1]
34	x	[Y1_91 = 5]*[X2_91 = 2]

Note: 'x' indicates an aliased (or a redundant) parameter.
These parameters are set to zero.

Goodness-of-fit Statistics

Chi-Square	DF	Sig.
------------	----	------

Likelihood Ratio	4,2589	4	,3721
Pearson	4,2371	4	,3749

Parameter Estimates

Constant	Estimate
1	2,4357

Note: Constant is not a parameter under multinomial assumption.
Therefore, standard errors are not calculated.

Parameter	Estimate	SE	Z-value	Asymptotic 95% CI	
				Lower	Upper
2	-1,1613	,5128	-2,26	-2,17	-,16
3	,0000	,	,	,	,
4	-,9148	,4833	-1,89	-1,86	,03
5	,0000	,	,	,	,
6	1,9164	,3015	6,36	1,33	2,51
7	3,2282	,2904	11,12	2,66	3,80
8	3,5432	,2894	12,24	2,98	4,11
9	,7701	,3384	2,28	,11	1,43
10	,0000	,	,	,	,
11	-,0064	,0750	-,09	-,15	,14
12	,0000	,	,	,	,
13	,0000	,	,	,	,
14	,0000	,	,	,	,
15	1,4080	,5230	2,69	,38	2,43
16	,0000	,	,	,	,
17	1,3911	,5155	2,70	,38	2,40
18	,0000	,	,	,	,
19	,7622	,5160	1,48	-,25	1,77
20	,0000	,	,	,	,
21	1,0612	,5603	1,89	-,04	2,16
22	,0000	,	,	,	,
23	,0000	,	,	,	,
24	,0000	,	,	,	,
25	1,0554	,4944	2,13	,09	2,02
26	,0000	,	,	,	,
27	1,0121	,4866	2,08	,06	1,97
28	,0000	,	,	,	,
29	,5563	,4870	1,14	-,40	1,51
30	,0000	,	,	,	,
31	,5016	,5359	,94	-,55	1,55
32	,0000	,	,	,	,
33	,0000	,	,	,	,
34	,0000	,	,	,	,

Aus diesem Output lassen sich die link Funktionen des log-linearen Modells und die oben angegebenen Differenzen ermitteln, z.B. für Zelle (111) im Vergleich zur Zelle(511):

$$\begin{aligned}
 \log(\hat{m}_{111}/\hat{m}_{511}) &= \log \hat{m}_{111} - \log \hat{m}_{511} \\
 &= [b + b_{A(1)} + b_{B(1)} + b_{C(1)} + b_{AB(11)} + b_{AC(11)} + b_{BC(11)}] \\
 &\quad - [b + b_{A(5)} + b_{B(1)} + b_{C(1)} + b_{AB(51)} + b_{AC(51)} + b_{BC(11)}] \\
 &= b_{A(1)} + b_{AB(11)} + b_{AC(11)}
 \end{aligned}$$

5 Modellierung multinomialer Daten (log-lineare Modelle)

$$\begin{aligned}
&= [2,4357 + 1,9164 + (-1,1613) + (-0,9148) + 1,4080 + 1,0551 + (-0,0064)] \\
&\quad - [2,4357 + 0 + (-1,1613) + (-0,9148) + 0 + 0 + (-0,0064)] \\
&= 1,9164 + 1,4080 + 1,0554
\end{aligned}$$

Analog erhält man:

$$\begin{aligned}
\log(\hat{m}_{211}/\hat{m}_{511}) &= \log \hat{m}_{211} - \log \hat{m}_{511} = b_{A(2)} + b_{AB(21)} + b_{AC(21)} \\
&= 3,2282 + 1,3911 + 1,0121 \\
\log(\hat{m}_{311}/\hat{m}_{511}) &= \log \hat{m}_{311} - \log \hat{m}_{511} = b_{A(3)} + b_{AB(31)} + b_{AC(31)} \\
&= 3,5432 + 0,7622 + 0,5563 \\
\log(\hat{m}_{411}/\hat{m}_{511}) &= \log \hat{m}_{411} - \log \hat{m}_{511} = b_{A(4)} + b_{AB(41)} + b_{AC(41)} \\
&= 0,7701 + 1,0612 + 0,5016
\end{aligned}$$

Diese Parameter entsprechen denjenigen aus der Tabelle 4.6 für Y1 im Jahre 1991.

Wie sich leicht nachprüfen lässt, führt auch das log-lineare Modell (AB|AC)

$$\log \hat{m}_{jkl} = b + b_{A(j)} + b_{B(k)} + b_{C(l)} + b_{AB(jk)} + b_{AC(jl)}, \quad (5.111)$$

zu dem Logit Modell $\log(\hat{m}_{jkl}/\hat{m}_{Jkl}) = \log \hat{m}_{jkl} - \log \hat{m}_{Jkl} = b_{A(j)} + b_{AB(jk)} + b_{AC(jl)}$, d.h., der Übergang vom log-linearen Modell zum logit Modell ist nicht eindeutig.

Relevante Teile des SPSS-Outputs sind dazu:

```

Model and Design Information
  Model: Multinomial
Design: Constant + X1_91 + X2_91 + Y1_91 + Y1_91*X1_91 + Y1_91*X2_91
Correspondence Between Parameters and Terms of the Design
Parameter   Aliased   Term
      1                Constant
      2                [X1_91 = 1]
      3                x    [X1_91 = 2]
      4                [X2_91 = 1]
      5                x    [X2_91 = 2]
      6                [Y1_91 = 1]
      7                [Y1_91 = 2]
      8                [Y1_91 = 3]
      9                [Y1_91 = 4]
     10                x    [Y1_91 = 5]
     11                [Y1_91 = 1]*[X1_91 = 1]
     12                x    [Y1_91 = 1]*[X1_91 = 2]
     13                [Y1_91 = 2]*[X1_91 = 1]
     14                x    [Y1_91 = 2]*[X1_91 = 2]

```

15		[Y1_91 = 3]*[X1_91 = 1]
16	x	[Y1_91 = 3]*[X1_91 = 2]
17		[Y1_91 = 4]*[X1_91 = 1]
18	x	[Y1_91 = 4]*[X1_91 = 2]
19	x	[Y1_91 = 5]*[X1_91 = 1]
20	x	[Y1_91 = 5]*[X1_91 = 2]
21		[Y1_91 = 1]*[X2_91 = 1]
22	x	[Y1_91 = 1]*[X2_91 = 2]
23		[Y1_91 = 2]*[X2_91 = 1]
24	x	[Y1_91 = 2]*[X2_91 = 2]
25		[Y1_91 = 3]*[X2_91 = 1]
26	x	[Y1_91 = 3]*[X2_91 = 2]
27		[Y1_91 = 4]*[X2_91 = 1]
28	x	[Y1_91 = 4]*[X2_91 = 2]
29	x	[Y1_91 = 5]*[X2_91 = 1]
30	x	[Y1_91 = 5]*[X2_91 = 2]

Note: 'x' indicates an aliased (or a redundant) parameter.
These parameters are set to zero.

Goodness-of-fit Statistics

	Chi-Square	DF	Sig.
Likelihood Ratio	4,2661	5	,5118
Pearson	4,2437	5	,5149

Parameter Estimates

Constant	Estimate
1	2,4361

Note: Constant is not a parameter under multinomial assumption.
Therefore, standard errors are not calculated.

Parameter	Estimate	SE	Z-value	Asymptotic 95% CI	
				Lower	Upper
2	-1,1632	,5123	-2,27	-2,17	-,16
3	,0000	,	,	,	,
4	-,9163	,4830	-1,90	-1,86	,03
5	,0000	,	,	,	,
6	1,9179	,3009	6,37	1,33	2,51
7	3,2296	,2898	11,14	2,66	3,80
8	3,5438	,2892	12,25	2,98	4,11
9	,7709	,3382	2,28	,11	1,43
10	,0000	,	,	,	,
11	1,4064	,5226	2,69	,38	2,43
12	,0000	,	,	,	,
13	1,3896	,5152	2,70	,38	2,40
14	,0000	,	,	,	,
15	,7614	,5159	1,48	-,25	1,77
16	,0000	,	,	,	,
17	1,0605	,5602	1,89	-,04	2,16
18	,0000	,	,	,	,
19	,0000	,	,	,	,
20	,0000	,	,	,	,
21	1,0533	,4938	2,13	,09	2,02
22	,0000	,	,	,	,
23	1,0101	,4860	2,08	,06	1,96
24	,0000	,	,	,	,
25	,5553	,4868	1,14	-,40	1,51
26	,0000	,	,	,	,
27	,5001	,5356	,93	-,55	1,55

28	,0000	,	,	,	,
29	,0000	,	,	,	,
30	,0000	,	,	,	,

Es ergibt sich:

$$\begin{aligned}
 \log(\hat{m}_{jkl}/\hat{m}_{5kl}) &= \log \hat{m}_{jkl} - \log \hat{m}_{5kl} \\
 &= [b + b_{A(j)} + b_{B(k)} + b_{C(l)} + b_{AB(jk)} + b_{AC(jl)}] \\
 &\quad - [b + b_{A(5)} + b_{B(k)} + b_{C(l)} + b_{AB(5k)} + b_{AC(5l)}] \\
 &= b_{A(j)} + b_{AB(jk)} + b_{AC(jl)},
 \end{aligned}$$

$$\begin{aligned}
 \log(\hat{m}_{111}/\hat{m}_{511}) &= \log m_{111} - \log \hat{m}_{511} = b_{A(1)} + b_{AB(11)} + b_{AC(11)} \\
 &= 1,9179 + 1,4064 + 1,0533
 \end{aligned}$$

$$\begin{aligned}
 \log(\hat{m}_{211}/\hat{m}_{511}) &= \log m_{211} - \log \hat{m}_{511} = b_{A(2)} + b_{AB(21)} + b_{AC(21)} \\
 &= 3,2296 + 1,3896 + 1,0101
 \end{aligned}$$

$$\begin{aligned}
 \log(\hat{m}_{311}/\hat{m}_{511}) &= \log \hat{m}_{311} - \log \hat{m}_{511} = b_{A(3)} + b_{AB(31)} + b_{AC(31)} \\
 &= 3,5438 + 0,7614 + 0,5553
 \end{aligned}$$

$$\begin{aligned}
 \log(\hat{m}_{411}/\hat{m}_{511}) &= \log \hat{m}_{411} - \log \hat{m}_{511} = b_{A(4)} + b_{AB(41)} + b_{AC(41)} \\
 &= 0,7709 + 1,0605 + 0,5001
 \end{aligned}$$

Bei vielen praktischen Untersuchungen soll erst die Zusammenhangsstruktur aufgedeckt werden, d.h., keine konkrete Hypothese und kein spezielles log-lineares Modell sind vorab formuliert. Die Aufgabe besteht somit darin, ein “bestes“ Modell unter allen möglichen log-linearen Modellen für die gegebene Kontingenztabelle zu suchen, wobei dieses Modell die Daten mit möglichst wenigen Parametern möglichst gut beschreiben, alle signifikanten Parameter (Effekte) einschließen und auch noch einfach zu interpretieren sein sollte.

Es wurden hier relativ einfache log-lineare Modelle behandelt. Eine Vielzahl von möglichen Modellspezifikationen und Problemen⁴⁵ blieb dabei unberührt, z.B.

- log-lineare Modelle für höherdimensionale Kontingenztabelle auf Basis des produkt-multinomialen Stichprobenmodells

⁴⁵Es wird diesbezüglich auf die angegebene Literatur verwiesen.

Das produkt-multinomiale Stichprobenmodell bei mehr als zwei Variablen wird vielfältiger, da die Grundgesamtheit entweder nach einer oder mehreren Variablen segmentiert werden kann. Andererseits wird beim produkt-multinomialen Stichprobenmodell die Menge der zulässigen log-linearen Modelle gegenüber dem multinomialen Stichprobenmodell eingeschränkt, da zusätzlich Restriktionen erforderlich sind und bestimmte Randtabellen bzw. -verteilungen als fest vorgegeben angesehen werden.

- Die behandelten Modelle sind sogenannte umfassende Modelle, da sie stets alle Haupteffekte und damit alle Variablen enthalten. Darüber hinaus gibt es die nicht umfassenden Modelle, bei denen auch Haupteffekte aus dem Modell entfernt werden, d.h., sie enthalten nicht alle Variablen.

- Kontingenztabellen mit leeren Zellen

Es wurde bisher immer vorausgesetzt, dass die Häufigkeiten für alle Zellen größer als Null sind. Das ist bei praktischen Untersuchungen jedoch oft unrealistisch. Leere Zellen können sowohl problem- als auch stichprobenbedingt sein. Problembedingt heisst dabei, dass eine bestimmte Kombination von Variablenkategorien gar nicht auftreten kann, d.h., ihre Auftretenswahrscheinlichkeit ist Null. Stichprobenbedingt können leere Zelle vor allem bei höherdimensionalen Kontingenztabellen und kleinem Stichprobenumfang auftreten, d.h., die Kombination von Variablenkategorien ist nicht unmöglich, sie wurde in der Stichprobe jedoch nicht beobachtet.

- Modellspezifikation bei ordinalskalierten Variablen

Wenn alle oder einige der Variablen ordinalskaliert sind, kann die zusätzliche Information, die mit der Ordnung der Kategorien gegeben ist, berücksichtigt werden, indem den Kategorien dieser Variablen Scores (z.B. Rangzahlen) zugeordnet werden. Bei der Spezifikation log-linearer Modelle werden diese Scores als Terme in das Modell aufgenommen. Mögliche Modellspezifikationen sind z.B. das linear-by-linear association model, das row-effects model und das column-effects model.

Literaturverzeichnis

- [1] Agresti, A. (1990), *Categorical Data Analysis*, John Wiley & Sons, New York
- [2] Agresti, A. (1996), *An Introduction to Categorical Data Analysis*, John Wiley & Sons, New York
- [3] Agresti, A. (1984), *Analysis of Ordinal Categorical Data*, John Wiley & Sons, New York et al.
- [4] Agresti, A., Kezouth, A. (1983) Association models for multidimensional crossclassifications of ordinal variables, *Communication in Statistics, Part A - theory and Method*, 12 (1983), 1261 - 1276
- [5] Allgemeine Bevölkerungsumfrage der Sozialwissenschaften, ALLBUS 1980-94, Codedbuch, ZA-Nr. 1795, herausgegeben vom Zentralarchiv für Empirische Sozialforschung Köln und vom Zentrum für Umfragen, Methoden und Analysen Mannheim.
- [6] Andersen, Erling B. (1990), *The Statistical Analysis of Categorical Data*, Springer, Berlin
- [7] Andersen, Erling B. (1980), *Discrete Statistical Models With Social Science Applications*, Amsterdam, North-Holland Publishing Company
- [8] Andreß, H.-J., Hagenaars, J., Kühnel, S.M. (1997) *Analyse von Tabellen und kategorialen Daten*, Springer, Berlin et. al.

- [9] Backhaus, K., Erichson, B., Plinke, W., Weiber, R. (2000), *Multivariate Analysemethoden*, Springer, Berlin, Heidelberg, New York
- [10] Bamberg, G., Baur, F. (1996), *Statistik*, Oldenbourg, München, Wien
- [11] Becker, M.P. (1989), Models for the analysis of association in multivariate contingency tables, *Journal of the American Statistical Association* 84, 1014-1019
- [12] Becker, M.P., Clogg, C.C. (1989), Analysis of sets of two-way contingency tables using association models, *Journal of the American Statistical Association* 84, 142-151
- [13] Becker, M.P. (1992), Exploratory analysis of association models using loglinear models and singular values decompositions, *Computational Statistics & Data Analysis* 13, 253-267
- [14] Ben-Akiva, Moshe, Lerman, Steven R. (1985), *Discrete Choice Analysis: Theory and Applications to Travel Demand*, MIT Press, Cambridge, MA
- [15] Benedetti, J.K., Brown, M.B. (1978), Strategies for the selection of log-linear models, *Biometrics*, 34, 680 - 686
- [16] Berry, D.A., Lindgren, B.W. (1990), *Statistics: Theory and Methods*, Brooks/Cole Publishing Company, Pacific Grove, California
- [17] Bishop, Y.M.M., Fienberg, S.E., Holland, P.W. (1975), *Discrete Multivariate Analysis: Theory and Practice*, MIT Press Cambridge (Mass.), London
- [18] Bock, R.D. (1985), *Multivariate Statistical Methods in Behavioral Research*, Mooresvills, Ind., Scientific Software, Inc.
- [19] Bortz, J. (1993) *Statistik*, Springer, Berlin et al.
- [20] Bosch, K. (1992), *Statistik-Taschenbuch*, Oldenbourg, München, Wien
- [21] Bühl, A., Zöfel, P. (1994), *SPSS für Windows Version 6*, Addison-Wesley, Bonn

- [22] Büning, H., Trenkler, G. (1994), Nichtparametrische statistische Methoden, Walter de Gruyter, Berlin, New York
- [23] Burda, M. (1993), The determinants of east-west german migration, *European Economic Review* 37: 452 - 461
- [24] Choulakian, V. (1988), Exploratory analysis of the contingency tables by log-linear formulation and generalizations of correspondence analysis, *Psychometrika* 53, 235-250
- [25] Christensen, R. (1990), *Log-Linear Models*, Springer, New York et al.
- [26] Clogg, C.C. (1982), Some models for the analysis of association in multiway cross-classifications having ordered categories, *Journal of the American Statistical Association*, 77, 803-815
- [27] Clogg, C.C. (1982), Using association models in sociological research: some examples, *American Journal of Sociology*, 88, 114-134
- [28] Clogg, C.C., Eliason, S.R. (1988), Some common problems in log-linear analysis, in: *Common Problems/Proper Solutions: Avoiding Error in Quantitative Research*, J. Scott Long, ed., Sage Publications, Beverly Hills
- [29] Clogg, C.C., Eliason, S.R., Grego, J. (1990), Models for the analysis of change in discrete variables, in: *Statistical Methods in Longitudinal Research*, Volume II ed. A. von Eye, Academic Press, Inc.
- [30] Coppi, R., Bolasco, S. (1989, Eds.), *Multiway Data Analysis*, North-Holland, Amsterdam
- [31] Collet, D. (1991), *Modelling Binary Data*, Chapman & Hall, London et al.
- [32] Cox, D.R., Hinkley, D.V. (1974), *Theoretical Statistics*, Chapman and Hall, London
- [33] Cox, D.R., Snell, E.J. (1989), *Analysis of Binary Data*, Chapman & Hall, London

- [34] Cramer, J.S. (1991), The Logit Model, Edward Arnold, London et al.
- [35] Dobson, A.J. (1991), An Introduction to Generalized Linear Models, Chapman & Hall, London et al.
- [36] Draper, N.R., Smith, H. (1981), Applied Regression Analysis, Wiley, New York
- [37] Everitt, B.S. (1977), The Analysis of Contingency Tables, London, Chapman and Hall
- [38] Fahrmeir, L., Hamerle, A. (1984), Multivariate statistische Verfahren, Walter de Gruyter, Berlin, New York
- [39] Fahrmeir, L., Kaufman, H. (1985), Consistency and asymptotic normality of the maximum likelihood estimator in generalized Linear Models, Annals of Statistics, 13, S. 342-368
- [40] Fahrmeir, L., Tutz, G. (1994), Multivariate Statistical Modelling Based on Generalized Linear Models, Springer Verlag, New York et al.
- [41] Fienberg, Stephen E. (1980), The Analysis of Cross Classified Categorical Data, Second Edition, MIT Press, Cambridge, MA
- [42] , J.D. (1974), A General Model for Multivariate Analysis, New York, Holt, Rinehart and Winston
- [43] Freeman, D.H. jr. (1987), Applied Categorical Data Analysis, New York, Dekker
- [44] Freund, R.J. (1980), The case of missing cell, The American Statistician, 34, 94-98
- [45] Gilula, Z., Haberman, S.J. (1988), The analysis of multivariate contingency tables by restricted canonical and restricted association models, Journal of the American Statistical Association 83, 760-771
- [46] Gilula, Z., Krieger, A.M., Ritov, Y. (1988), Ordinal association in contingency tables; some interpretative aspects, Journal of the American Statistical Association 83, 540-545

- [47] Goodman, L.A. (1964), Simple methods of analyzing three-factor interaction in contingency tables, *Journal of the American Statistical Association*, 59, 319 - 352
- [48] Goodman, L.A. (1978), *Analyzing qualitative/categorical data*, Cambridge, Mass., Abt Books
- [49] Goodman, L.A. (1978), *Analyzing qualitative/categorical data: Log-linear models and latent structure analysis*. J. Madigson ed. United Press of America, Inc. Boston
- [50] Goodman, L.A. (1979), Simple models for the analysis of association in cross-classifications having ordered categories, *Journal of the American Statistical Association*, 74, 537-552
- [51] Goodman, L.A. (1984), *The analysis of cross-classified data having ordered categories*, Cambridge, Mass., Harvard University Press
- [52] Goodman, L.A. (1985), The analysis of cross-classified data having ordered and/or unordered categories: association models, correlation models, and asymmetry models for contingency tables with or without missing entries, *Annals of Statistics* 13, 1069
- [53] Goodman, L.A. (1986), Some useful extensions of the usual correspondence analysis approach and the usual log-linear models approach in the analysis of contingency tables. *International Statistical Review* 54, 243-309
- [54] Goodman, L.A. (1987), New methods for analyzing the intrinsic character of qualitative variables using cross-classified data, *American Journal of Sociology*, 93, 529-583
- [55] Green, P.J., Silverman, B.W., (1993), *Nonparametric Regression and Generalized Linear Models*, Chapman & Hall, London et al.
- [56] Grizzle, J.E., Starmer, C., Koch, G.G., (1969), *Analysis of Categorical Data by Linear Models*, *Biometrics* 25, 489-504

- [57] Haberman, S.J. (1978), Analysis of Qualitative Data, Vol. I: Introductory Topics, New York, Academic Press
- [58] Haberman, S.J. (1978), Analysis of Qualitative Data, Vol. II: New Developments, New York, Academic Press
- [59] Haberman, S.J. (1982), Analysis of Dispersion of multinomial responses, Journal of the American Statistical Association, 77, 568 - 580
- [60] Hartung, J., Elpelt, B., Klösener, K.-H. (1993), Statistik, 9. Auflage, Oldenbourg, München, Wien
- [61] Hilbe, J.M. (1994) Generalized linear models, The American Statistician
- [62] Hochstädter, D., Kaiser, U. (1988), Varianz- und Kovarianzanalyse, Harri Deutsch, Frankfurt am Main, Thun
- [63] Hosmer, D.W., Lemeshow, S. (1989), Applied Logistic Regression, John Wiley & Sons, New York et al.
- [64] Jobson, J.D. (1992), Applied Multivariate Data Analysis; Volume II: Categorical and Multivariate Methods, Springer Verlag, New York et al.
- [65] Judge, G.G., Hill, R.C., Griffiths, W.E., Lütkepohl, H., Lee, T.-Chao (1988), Introduction to the Theory and Practice of Econometrics, John Wiley & Sons, New York et al.
- [66] Kleinbaum, D.G. (1994), Logistic Regression, Springer, New York et al.
- [67] Kleinbaum, D.G., Kupper, L.L., Muller, K.E., (1988), Applied Regression Analysis and other Multivariate Methods, P.W.S.-Kent, Boston, Mass.
- [68] Koch, G.G., Landis, J.R., Freeman, J.L., Freeman, D.H., Lehnen, R.G. (1977), A General Methodology for the Analysis of Experiments with Repeated Measurements of Categorical Data, Biometrics 33, 133 - 158

- [69] Kühnel, S.M. (1997), Benutzerdefinierte Design-Matrizen in log-linearen Analysen: Realisierungsmöglichkeiten in den SPSS-Prozeduren GENLOG oder LOGLINEAR, ZA-Information 40, Mai 1997, Zentralarchiv für Empirische Sozialforschung an der Universität Köln, S. 60 - 86
- [70] Langenheine, Rolf (1980), Log-lineare Modelle zur multivariaten Analyse qualitativer Daten, Oldenbourg, München
- [71] Lauro, N.C., Siciliano, Exploratory methods and modelling for contingency tables and analysis: an integrated approach, Statistica Applicata, Italian Journal of Applied Statistics, 1, 532
- [72] Law, H.G., Snyder, C.W., Hattie, J.A., McDonalds, R.P. (1984, Eds.), Research Methods for Multimode Data Analysis, Praeger, New York
- [73] Lindsey, J.K. (1997), Applying Generalized Linear Models, Chapman & Hall, London et al.
- [74] Maddala, G.S. (1983), Limited Dependent and Qualitative Variables in Econometrics, Cambridge University Press, Cambridge, MA
- [75] Madigson, J. (1981), Qualitative variance, entropy and correlation ratios for nominal dependent variables, Social Science Research, 10, 177-194
- [76] McCullagh, P., Nelder, J.A., (1991), Generalized Linear Models, 2nd Edition, Chapman & Hall, London et al.
- [77] Menard, S. (1995), Applied Logistic Regression, Laye
- [78] Nelder, J.A., Wedderburn, R.W.M. (1972), Generalized linear models, Journal of the Royal Statistical Society, Series A 135 (3), 370-384
- [79] Reynolds, H.T. (1977), The analysis of Cross-Classifications, New York, The Free Press

- [80] Rönz, B. (1999), Modelling the perception of current and prospective economic situation. Statistical Research Report No. 99.002, Centre for Mathematics and its Applications, School of Mathematical Science, The Australian National University Canberra
- [81] Rönz, B., Förster, E. (1992), Regressions- und Korrelationsanalyse, Gabler-Verlag, Wiesbaden
- [82] Rönz, B., Strohe, H.G. (Hrsg.) (1994), Lexikon Statistik, Gabler-Verlag, Wiesbaden
- [83] Santner, Th.J., Duffy, D.E. (1989), The Statistical Analysis of Discrete Data, Springer, New York et al.
- [84] Schlittgen, R. (1990), Einführung in die Statistik, Oldenbourg, München, Wien
- [85] Schwarze, J. (1990), Grundlagen der Statistik II, Verlag Neue Wirtschafts-Briefe, Herne, Berlin
- [86] Seber, G.A.F. (1977), Linear Regression Analysis, Wiley, New York
- [87] Siciliano, R., Mooijaart, Ab (1997), Three-factor association models for three-way contingency tables, Computational Statistics and Data Analysis, Volume 24 Number 3 (12 May 1997), 337-356
- [88] Siciliano, R., Lauro, N.C., Mooijaart, Ab (1990), Exploratory approach and maximum likelihood estimation of models for non-symmetrical analysis of two-way multiple contingency tables, Compstat '90, 157-162
- [89] Siciliano, R., van der Heijden, P.G.M. (1994), Simultaneous latent budget analysis of a set of two-way tables with constant-row-sum data, Metron, 53, 1-2
- [90] SPSS for Windows: Advanced Statistics Release 6.0 (1993)
- [91] SPSS Regression Models 9.0 (1999)
- [92] SPSS Advanced Models 9.0 (1999)

- [93] Tutz, G. (2000), Die Analyse kategorialer Daten, Oldenbourg Verlag, München
- [94] Upton, G.J.G (1978), The Analysis of Cross-Tabulated Data, New York, John Wiley and Sons
- [95] Vach, W. (1994), Logistic Regression with Missing Values in the Covariates, Springer Verlag, New York et al.
- [96] Winer, B.J. (1971), Statistical Principles in Experimental Design, New York
- [97] Wong, R.R.S.K. (1995), Extensions in the use of log-multiplicative scaled associated models in multiway tables, Sociol. Methods and Research, 23, 507-538
- [98] Xie, Y. (1992), The log-multiplicative layer effect model for comparing mobility tables, American Socio. Review, 57, 380-395

Index

- abhängige Variable, 2
- Abweichungskodierung, 88
- ALLBUS, 121
- Alternativhypothese, 6, 14, 69, 79, 85, 93
- arithmetisches Mittel, 90, 152, 164
- Assoziation, 79, 159, 208, 215, 218, 228, 229
- asymptotisch, 37

- Backward-Selection, 69
- bedingte relative Häufigkeit, 58, 124
- bedingte relative Häufigkeitsverteilungen, 139
- bedingte Unabhängigkeit, 205
- bedingte Wahrscheinlichkeit, 52, 54, 116, 117, 145, 149
- Bernoulli-Verteilung, 50, 53
- Bezugskategorie, 84, 85, 88, 90, 120, 124, 131
- binär, 49, 116, 118–120
- binomialverteilt, 40, 61, 65
- Binomialverteilung, 28, 41, 51–53, 141, 158

- Chi-Quadrat-Statistik, 168
- Chi-Quadrat-Teststatistik, 180, 214, 216–218, 220, 221, 223, 224
- chi-quadrat-verteilt, 44, 69, 70, 169, 171, 181
- Chi-Quadrat-Verteilung, 15, 68, 80, 86, 94, 98, 126, 169, 171, 180, 210
- Chi-Quadrat-Verteilung, 225
- Chi-Quadrat-Verteilung, 226
- Cook's Distanz, 72
- corner-point-Restriktion, 152, 166, 173, 175, 187, 202, 210, 229
- Cox&Snell, 70, 126
- cut point, 72
- cut value, 81, 94, 99

- Delta-Wert, 182
- Design-Matrix, 88, 90, 158, 173, 174, 186
- Deviance, 44, 45, 67, 68, 80, 86, 93, 98, 126, 170, 209
- Deviance-Residuum, 46, 71
- deviation, 88
- DFBETAS, 71

- dichotom, 51, 60, 77, 83, 95, 109, 118, 124, 131
- Dichtefunktion, 8, 22, 27, 61, 64
- diskret, 22
- Drei-Variablen-Interaktion, 200
- dreidimensionale Kontingenztafel, 198
- Dummy-Kodierung, 152, 173, 174, 181, 187, 194, 210, 229
- Dummy-Variable, 156, 166
- Effekt, 8, 89
- Effekt-Kodierung, 152, 173, 174, 177, 184, 187, 190, 196, 229
- einfache lineare Regression, 58
- einfache Zufallsstichprobe, 8
- erklärende Variable, 2, 51, 52, 63
- erwartete absolute Häufigkeit, 140
- Erwartungswert, 23, 31, 33, 36, 37, 39, 40, 42, 60, 65, 119
- eviation, 184
- exponentielle Familie, 22, 54
- exponentielle Familie, 158
- F-Test, 14, 47
- F-Verteilung, 15, 47
- Faktor, 6, 52, 53
- Faktorstufen, 6
- Forward-Selection, 69
- full model, 5
- Güte der Anpassung, 37, 43
- gemeinsame Häufigkeitsverteilung, 138
- gemeinsame log-Likelihood-Funktion, 26
- geometrisches Mittel, 91
- gesättigtes Modell, 5
- Goodness-of-Fit, 68, 105, 113, 179
- Goodness-of-fit, 168
- hat-Matrix, 71
- Haupteffekt, 151, 155, 164, 167, 181, 182, 186, 189, 200, 214, 229
- hierarchische log-lineare Modelle, 200, 226
- hierarchisches Modell, 159
- Homogenität, 155
- Homogenitätshypothese, 149
- Hypothese, 47
- Hypothesenprüfung, 3, 37, 46, 63, 168
- Indikator-Kodierung, 85, 87, 124, 131
- Information, 25
- Informationsmatrix, 26, 27, 34, 35, 38–42, 65, 70, 71, 119
- Interaktion, 101, 135, 148, 203, 225
- Interaktionseffekte, 154, 188, 189, 192
- Interaktionsparameter, 164, 167, 186, 189, 213, 216, 217, 229
- Interaktionsterm, 159, 228
- Iterationsprozess, 37
- iterative gewichtete Kleinst-Quadrate-Schätzung, 36
- kanonische Form, 22, 28–30, 33

- kategoriale Variable, 53, 115, 120, 122
- Klassifikationstabelle, 72
- Kodierung, 157
- Konfidenzintervall, 43, 182, 186, 213
- Kontingenztafel, 52, 54, 58, 83, 90, 92, 96, 125, 138, 159
- Kontrast-Variable, 84, 85, 88, 92, 120, 124, 131, 173, 175, 177, 181, 186, 188
- Konvergenzschranke, 37
- korrigiertes standardisiertes Residuum, 141
- Kovariante, 2, 16
- Kovarianz, 38
- Kovarianzanalyse, 2

- Leverage, 71
- Likelihood-Funktion, 9, 11, 12, 39, 40, 44, 55, 67, 70, 109, 118, 160
- lineare Modelle, 1, 21, 31
- lineares Regressionsmodell, 17, 60
- Linearität, 96, 101
- link Funktion, 31–34, 39, 41, 43, 56, 58, 60, 62, 73, 78, 82, 95, 99, 105, 109, 113, 114, 117, 125, 128, 131, 133, 159, 173, 176, 179, 186, 200, 202, 204–208, 210, 213, 214, 216–218, 220, 221, 223, 231
- log odds, 62, 74, 78, 117, 127, 129
- log odds ratio, 82
- log-likelihood ratio Statistik, 44, 170, 214, 216–218, 220, 221, 223, 224
- log-Likelihood-Funktion, 11, 13, 23, 25, 33, 35, 39, 40, 44, 55, 63, 68, 77, 81, 86, 94, 98, 110, 118, 127, 160, 161, 170
- log-lineares Modell, 151, 158, 173, 199, 228, 229, 234
- logarithmierte Likelihood-Funktion, 10
- logistische Regression, 74, 77, 83, 92
- logistische Verteilung, 61, 64
- logit, 62, 110, 111, 129
- logit Funktion, 62
- Logit Modell, 62, 70, 102, 104, 110, 113, 114, 119, 120, 127, 228, 229
- logit Residuen, 71
- Logit-Transformation, 110
- logits, 90, 117, 127, 128, 135

- maximales Modell, 43, 44, 66, 68
- Maximum-Likelihood-Methode, 8, 33
- Maximum-Likelihood-Schätzer, 63, 144
- Maximum-Likelihood-Schätzung, 9
- McFadden, 126
- mehrkategorial, 52, 83
- Methode der kleinsten Quadrate, 17
- metrisch, 92, 95, 101, 120, 137
- metrisch skaliert, 52, 53
- ML-Schätzer, 41, 44, 119, 161–163
- Modelldiagnose, 101

- Modellprüfung, 70
- Modellspezifikation, 3, 31
- multinomial, 137
- multinomiale logistische Verteilung, 117
- multinomiales Logit Modell, 115, 119, 127
- multinomiales Stichprobenmodell, 141, 145, 148, 153, 159, 161, 172, 181, 235
- Multinomialkoeffizient, 142
- multinomialverteilt, 144, 145, 148, 160, 169, 172
- Multinomialverteilung, 116, 141, 142, 146, 149, 158, 199
- multiple logistische Regression, 95
- multiple Logit Modell, 133
- Multiplikationssatz, 140
- Nagelkerke, 70, 81, 87, 94, 99, 126
- nested Modelle, 209, 225
- Newton-Raphson-Methode, 35
- Newton-Raphson-Prozedur, 119
- nichtlineare logits, 101
- nominalskaliert, 51, 52, 137, 172
- Normalgleichungen, 19
- normalverteilt, 7, 16, 42, 44, 60, 169
- Normalverteilung, 21, 27, 38, 40, 46, 58, 61, 109
- Nullhypothese, 6, 13, 68, 70, 76, 79, 85, 93, 105, 113, 132, 148, 149, 151, 152, 162, 167, 169, 171, 173, 180, 202, 204–206, 210, 224
- odds, 62, 74, 78, 82, 91, 120, 130, 131, 229
- odds ratio, 78, 82, 88, 91, 92, 95, 120, 128, 129, 190, 208
- ordinalskaliert, 137, 235
- Pearson Residuum, 46
- Perzeption, 121
- Poisson-Prozess, 147
- Poisson-Stichprobenmodell, 146
- poisson-verteilt, 147
- Poisson-Verteilung, 30, 147, 158
- Polymialverteilung, 142
- probit, 108, 110, 111
- Probit-Funktion, 108
- Probit-Modell, 108, 109, 112, 113
- Probit-Transformation, 108, 110
- produkt-multinomiales Stichprobenmodell, 145, 149, 153, 159, 162, 234
- Produkt-Multinomialverteilung, 146
- Pseudo R-Square, 127
- Rückwärts-Selektion, 101
- random error term, 7
- Randtabelle, 198
- Randverteilung, 138, 148, 162, 170, 198
- Randwahrscheinlichkeit, 140, 169
- Referenzkategorie, 116, 117, 120, 121, 124, 129, 131, 152, 155, 157, 187, 194, 210
- Regressionsfunktion, 16

Index

- Regressionskoeffizient, 16
- Regressionskonstante, 16
- Regressionsmodell, 2, 55
- Reparametrisierung, 179
- Residuen, 179, 210
- Residuum, 4, 71, 140, 171
- response, 49, 53, 60, 105, 107, 115, 116,
121, 124, 130, 135, 172
- Restriktion, 151, 155, 157, 173, 191, 200
- saturiertes log-lineares Modell, 154
- saturiertes Modell, 5, 66, 156, 163, 166,
170, 173, 186, 187, 190, 194, 196,
199, 209, 210
- Scatterplot, 58, 102, 104
- Schätzung, 3, 33, 53, 55, 63, 85, 89, 131,
133, 149, 150, 160, 172, 181
- Scheinvariable, 85, 166
- Score, 23, 26, 39, 40, 69, 235
- Score-Test, 69
- Scores, 37
- Scoring-Methode, 35
- Signifikanzniveau, 76
- Skalenniveau, 3, 53, 122
- SPSS, 76, 93, 102, 112, 124, 172, 181
- Standardfehler, 43, 70, 141, 167
- standardisiertes Residuum, 46, 71, 141,
169, 171
- Standardnormalverteilung, 41, 43, 108
- Stepwise-Selektion, 101
- stetig, 22
- Stichprobenfunktion, 38–40
- Stichprobenmodell, 141, 145
- Stichprobenvarianz, 45
- Stichprobenverteilung, 37, 41, 44
- sum-to-zero Restriktion, 152, 163, 173,
201, 229
- Syntax, 184
- systematische Komponente, 4, 7, 17
- Teststatistik, 20, 169, 170
- Toleranz-Verteilung, 61, 108
- Trennpunkt, 72
- Unabhängigkeit, 140, 146, 148, 155, 180,
202
- Unabhängigkeitssatz, 148
- unstandardisiertes Residuum, 140
- Varianz, 25, 33, 37, 60, 65
- Varianz-Kovarianz-Matrix, 26, 34, 38, 42,
66, 71, 119, 143
- Varianzanalyse, 2, 8
- verallgemeinerte lineare Modelle, 22
- verallgemeinertes lineares Modell, 2, 158
- Verschiebungssatzes, 14
- Verteilungsfunktion, 58, 60, 64, 108
- Vorhersage, 72
- Vorwärts-Selektion, 101
- Wahrscheinlichkeitsfunktion, 8, 22, 28, 30
- Wahrscheinlichkeitsplots, 46

Wahrscheinlichkeitsverteilung, 50, 137

Wald-Test, 70, 77, 81, 95, 99, 128

Wechselwirkung, 148, 154

Zellhäufigkeit, 138, 143, 149, 172

Zentraler Grenzwertsatz, 38

Zufallsexperiment, 49, 142, 143

Zufallskomponente, 7, 16

Zwei-Variablen-Interaktion, 200, 208