Logit Modelle mit ordinaler Response Variable

Diplomarbeit

zur Erlangung des Grades
eines Diplom-Volkswirtes

an der Wirtschaftswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

vorgelegt von

André Kühnlenz

(Matrikel-Nr. 132982)

Prüfer: Prof. Dr. Bernd Rönz

Berlin, 6. Februar 2002

Inhaltsverzeichnis

1 Einleitung **4**

1.1 Das einfache loglineare Modell für kategoriale Variablen 6

2 Die Logit-Transformation **9**

2.1 Binäre Response-Variablen 10

2.2 Kategoriale Response-Variablen 13

2.3 Ordinale Response - Variablen 15

3 Die Logistische Regression **17**

3.1 Das binäre Logit-Modell 17

3.2 Das multinomiale Logit-Modell 18

3.3 Zur Korrespondenz zwischen dem Logit-Modell und dem loglinearen Modell 19

3.4 Logit-Modelle für ordinale Response-Variablen 21

3.4.1 Nachbarschafts-Modelle (Adjacent-Categories-Models) . . . 21

3.4.2 Das kumulative Logit-Modell 26

3.4.3 Sequentielle Modelle oder *Continuation Ratio Models* 30

3.4.4 Zwei-Schritt (Two-Step) -Modelle 32

4 Das Schätzen und Testen von kumulativen Logit-Modellen **34**

4.1 Die Schätzmethoden 34

4.2 Gewichtete Kleinste Quadrate - Schätzung (Weighted Least Squares - WLS) 34

4.3 Das verallgemeinerte lineare Modell 38

4.4 Die ML-Schätzung für kumulative Logit-Modelle 39

4.5 Hypothesenüberprüfung in kumulativen Logit-Modellen 41

5 Anwendung auf die Perzeption der Wirtschaftslage **41**

5.1 Resultate der bivariaten kumulativen Logit-Modelle 43

6 Anhang **54**

6.1 Stichprobenmodelle und -verteilungen 54

Abbildungsverzeichnis

1 Konstante *odds-ratios* im *uniform association model* 25

2 Einschätzung der Wirtschaftslage in der BRD 45

Tabellenverzeichnis

1 Kontigenztabelle mit den gemeinsamen (bedingten) Zellwahrscheinlichkeiten für (X, Y) 7

2 Häufigkeitsverteilungen der Response-Variablen; relative Häufigkeiten in %. 44

4 Schätzwerte für X1 als Erklärende 48

5 Schätzwerte für X2 als Erklärende 49

6 Schätzwerte für X3 als Erklärende 50

7 Schätzwerte für X4 als Erklärende 51

8 Schätzwerte für X5 als Erklärende 52

9 Schätzwerte für X6 als Erklärende 53

1 Einleitung

In vielen statistischen Untersuchungen spielen ordinale Variablen eine wichtige Rolle. Zum Beispiel können Entscheidungen und Urteile von Menschen der Gegenstand von Befragungen sein, die dann statistisch ausgewertet werden sollen. Sehr oft besitzen Variablen, die die Entscheidungen von Personen reflektieren ordinales Skalenniveau. Ordinale Variablen gehöhren zu den kategorialen Variablen. Bei kategorialen Variablen kommt einzig ihre Verschiedenheit in den Ausprägungen zum Ausdruck. Die Unterschiede zwischen den Ausprägungen sind dabei aber nicht messbar. Ordinal skaliert sind die Daten, wenn sich die Ausprägungen zusätzlich in eine natürliche Reihenfolge bringen lassen. Die absolute Distanz zwischen den Ausprägungen ist jedoch unbekannt und daher nicht quantifizierbar. Man spricht auch von qualitativen Variablen in dem Sinne, dass die verschiedenen Ausprägungen in eine Beziehung größer oder kleiner gebracht werden können. [1]

Generell lassen sich für kategoriale Variablen alle Methoden für nominalskalierte Daten anwenden (z.B. der Pearson Chi-Quadrat Test oder loglinare Modelle). Für ordinale Variablen eignen sich jedoch speziellere Methoden besser, die Natur der Daten wiederzuspiegeln. Bei den in dieser Arbeit vorgestellten Logit-Modellen soll besonderer Wert darauf gelegt werden, dem ordinalen Charakter der zu untersuchenden Variablen gerecht zu werden. Tutz gibt folgende Gründe an, die es sinnvoll machen, der Skala der zu untersuchenden Variablen besondere Beachtung zu schenken:

> „Die Anwendung von Modellen, die ein höheres Skalenniveau benutzen, führt i.a. zu Artefakten, da Information vorausgesetzt wird, die nicht in den Skalen enthalten ist. Ist das Modell hingegen auf zu niedriges Skalenniveau zugeschnitten, wird die in den Skalen enthaltene Information nicht erschöpfend genutzt."
>
> „Obwohl natürlich kategorial-nominale Modelle auf ordinales Skalenniveau anwendbar sind, wird bei dieser Vorgehensweise auf einen Teil der in den Daten entaltenen Informationen verzichtet. Die Folge ist eine nicht notwendige Überparametrisierung..."[2]

[1] siehe auch Rönz & Strohe (1994)
[2] Tutz (1990), S.V.

Bei der Untersuchung von mehreren Variablen kann je nach dem Zweck zwischen einer symmetrischen und einer asymmetrischen Fragestellung unterschieden werden. Die symmetrische Sichtweise stellt sich zur Aufgabe, die Interaktion zwischen den Variablen zu untersuchen. Für diese Fragestellung eignen sich besonders loglineare Modelle. In den in dieser Arbeit betrachteten Logit-Modellen geht man dagegen asymmetrisch an die beobachteten Variablen heran. Dabei soll die Art und Stärke des Einflusses von exogenen kategorialen Variablen oder metrischen Kovarialen auf eine abhängige kategoriale Variable untersucht werden. Schwerpunkt dieser Arbeit bildet die Modellierung der Abhängigkeit, wenn die abhängige Variable ein ordinales Skalenniveau besitzt.

Häufig resultieren ordinale Daten auch aus einer diskreten Darstellung von stetigen Variablen wie zum Beispiel dem Alter oder die Bildung. Dabei wird eine stetige Variable durch Intervallbildung kategorisiert. Dies kann nützlich sein im Sinne einer besseren Approximation der zugrunde liegenden Skala. Diese wird in verschiedene Gruppen eingeteilt. Ist die stetige Variable monoton, können die Gruppen in eine Rangordnung gebracht werden.

Oft können ordinale Variablen auch auf die Wirkungsweise latenter stetiger Variablen zurückgeführt werden. Dabei nimmt man an, dass eine bestimmte Größe nicht beobachtbar ist. Nimmt man zum Beispiel die Einschätzung der Wirtschaftslage, können die Befragten diese auf einer objektiv nicht messbaren Skala vornehmen. Diese Skala ist nicht messbar, da jeder Befragte seine Einschätzung der wirtschaftlichen Tatsachen nach subjektiven Gesichtspunkten geben wird. Darüber hinaus hat man bei einer Untersuchung keine Anhaltspunkte über die Art der Skala, die jeder Befragte bei seiner Einschätzung hat. Diese könnte sich theoretisch auf einen unendlichen Bereich erstrecken und wäre somit eine stetige Größe. Realistischer ist es anzunehmen, dass jeder Befragte seine eigenen Kategorien haben wird, um diesen unendlichen Bereich einzugrenzen. Diese verschiedenen subjektiven Einschätzungen versucht man nun, mit Hilfe einer ordinalen Variablen zu vereinheitlichen. Indem man den Befragten Kategorien vorgibt, für die sie sich entscheiden müssen. Bei einer Umfrage nach der Einschätzung der Wirtschaftslage könnte man zum Beispiel die Kategorien „sehr gut”, „gut”, „teils/teils”, „schlecht” oder „sehr schlecht” zu Hilfe nehmen. Fragt man nach der Einschätzung der zukünfigen Wirtschaftslage wären Kategorien wie „wesentlich besser”, „etwas besser”, „gleichbleibend”, „etwas schlechter” und „wesentlich schlechter” gut geeignet, die Beurteilung der Wirtschaftslage zu erfassen.

In diesem Kapitel soll kurz ein Überblick über das loglineare Modell gegeben werden. Da eine Beziehung zwischen dem loglinearen und dem Logit-Modell besteht, wird es zur Einführung hier kurz vorgestellt. In dem folgenden zweiten Kapitel betrachten wir die Notwendigkeit und die verschiedenen Möglichkeiten der Bildung von Logits. Diese selbst werden im dritten Kapitel zum Gegenstand von Modellen der logistischen Regression (Logit-Modellen). Dabei werden hier die Logit-Modelle mit ordinaler Response-Variable im Zentrum der Betrachtung stehen. Vorher sollen aber noch die Logit-Modelle für verschiedene andere kategoriale Variablen vorgestellt werden, da die ordinalen Logit-Modelle zum Teil auf diesen basieren. Abschließend soll ein Logit-Modell auf die Prezeption der Wirtschaftslage empirisch angewandt werden.

1.1 Das einfache loglineare Modell für kategoriale Variablen

Beim loglinearen Modell steht die symmetrische Fragestellung im Mittelpunkt. In erster Linie kann mit diesem Modell die gemeinsame Wahrscheinlichkeitsverteilung von mindestens zwei kategorialen Zufallsvariablen geschätzt werden. Darüber hinaus dient das Modell dazu, Aussagen über die Beziehung zwischen den Variablen zu treffen.

Die gemeinsame Wahrscheinlichkeitsverteilung läßt sich in einer Kontigenztabelle darstellen. Für zwei kategoriale Variablen X und Y mit jeweils I und J verschiedenen Ausprägungen erhält man $N = IJ$ verschiedene Möglichkeiten der gemeinsamen Ausprägung, die in einer $I \times J$-Kontingenztabelle dargestellt werden können. Tabelle 1 zeigt die Wahrscheinlichkeiten $\{\pi_{ij}\}$ dafür, dass (X, Y) in eine Zelle in der Zeile i und Spalte j fällt.

Bei einem Gesamtstichprobenumfang n lassen sich die einzelnen Zellhäufigkeiten n_{ij} (die sich zu n addieren) analog in einer Kontigenztabelle darstellen. Neben der gemeinsamen Verteilung sind in der Kontingenztabelle auch die Randverteilungen angegeben, die sich wie folgt ergeben

$$\pi_{i+} = \sum_j \pi_{ij}, \pi_{+j} = \sum_i \pi_{ij} \text{ und } \pi_{++} = \sum_i \pi_{i+} = \sum_j \pi_{+j} = \sum_i \sum_j \pi_{ij} = 1.$$

Analog zu den Randwahrscheinlichkeiten lassen sich auf diese Weise die Randhäufigkeiten definieren. Da die Wahrscheinlichkeiten π_{ij} nicht beobachtet werden können, müssen sie geschätzt werden mit den relativen Häufigkeiten der Stichprobe $p_{ij} = n_{ij}/n$.

Variable X	Variable Y					Randverteilung
	Y_1	$\dots$	Y_j	$\dots$	Y_J	von X
X_1	π_{11} $(\pi_{1\|1})$	$\dots$	π_{1j} $(\pi_{j\|1})$	$\dots$	π_{1J} $(\pi_{J\|1})$	π_{1+} (1)
$\vdots$	$\vdots$	$\dots$	$\vdots$	$\dots$	$\vdots$	$\vdots$
X_i	π_{i1} $(\pi_{1\|i})$	$\dots$	π_{ij} $(\pi_{j\|i})$	$\dots$	π_{iJ} $(\pi_{J\|i})$	π_{i+} (1)
$\vdots$	$\vdots$	$\dots$		$\dots$	$\vdots$	$\vdots$
X_I	π_{I1} $(\pi_{1\|I})$	$\dots$	π_{Ij} $(\pi_{j\|I})$	$\dots$	π_{IJ} $(\pi_{J\|I})$	π_{I+} (1)
Randverteilung von Y	π_{+1}	$\dots$	π_{+j}	$\dots$	π_{+J}	$\pi_{++} = 1$

Tabelle 1: Kontigenztabelle mit den gemeinsamen (bedingten) Zellwahrscheinlichkeiten für (X, Y)

Bei der symmetrischen Fragestellung steht die gemeinsame Verteilung der Variablen im Mittelpunkt. Daher war es bisher sinnvoll, davon auszugehen, dass alle zu untersuchenden Variablen Zufallsvariablen sind. Möchte man nun, wie im Fall der Logit-Modelle, den Einfluß von erklärenden Variablen z.B. X auf eine unabhängige Response-Variable z.B. Y untersuchen, spielt die gemeinsame Verteilung beider Variablen keine Rolle mehr. Denn dabei gehen wir von der Response-Variable als Zufallsvariable aus und betrachten die erklärenden Variablen als fest und nicht stochastisch. Für eine bestimmte Ausprägung von X hat Y somit eine bedingte Wahrscheinlichkeitsverteilung.

Im Mittelpunkt von Logit-Modellen steht die Untersuchung, wie sich die Wahrscheinlichkeitsverteilung von Y ändert, wenn sich die erklärende Variable X verändert. Zu diesem Zweck lassen sich die Wahrscheinlichkeiten wieder in einer Kontigenztabelle (wie in Tabelle 1) darstellen. Nur besteht hier der Unterschied, dass anstelle von $\{\pi_{ij}\}$ nun die bedingten Wahrscheinlichkeiten $\{\pi_{j|i}\}$ (in der Tabelle in Klammern gesetzt) verwendet werden. Die Wahrscheinlichkeiten sind bedingt auf die jeweilige Ausprägung von X, wobei $\pi_{j|i}$ die Wahrscheinlichkeit einer Ausprägung von Y in Spalte j bedeutet, gegeben X ist in Zeile i ausgeprägt. Die Wahrscheinlichkeiten in jeder Zeile addieren sich zu 1

$$\sum_j \pi_{ij} = 1.$$

Die Logit-Modelle sind Gegenstand der folgenden Kapitel. Nachfolgend soll jedoch zuerst das loglineare Modell betrachtet werden. Auch hier kann die Untersuchung der

gemeinsamen Verteilung mit Hilfe von bedingten Wahrscheinlichkeiten vorgenommen werden. Die bedingten Wahrscheinlichkeiten von Y, gegeben X ergeben sich auch wie folgt

$$\pi_{j|i} = \pi_{ij}/\pi_{i+}. \tag{1}$$

Die verschiedenen Möglichkeiten der Verwendung der Wahrscheinlichkeiten korrespondieren mit unterschiedlichen Erhebungskonzepten (siehe Anhang Kapitel 6.1). Hier soll vom Multinomialen Stichprobenmodell ausgegangen werden.

Die beiden Variablen sind statistisch unabhängig, wenn gilt:

$$\pi_{ij} = \pi_{i+}\pi_{+j} \qquad \forall\, i, j. \tag{2}$$

Und für die erwarteten absoluten Häufigkeiten $\{m_{ij} = n\pi_{ij}\}$ ergibt sich

$$m_{ij} = n\pi_{i+}\pi_{+j} \qquad \forall\, i, j. \tag{3}$$

Beim loglinearen Modell wird nun ein Modell für die erwarteten Häufigkeiten aufgestellt. Durch Logarithmieren läßt sich bei Unabhängigkeit der Variablen ein additives Modell herleiten

$$\log m_{ij} = \log n + \log \pi_{i+} + \log \pi_{j+}. \tag{4}$$

Damit ist die erwartete Häufigkeit eine Funktion des i-ten Zeileneffektes und des j-ten Spalteneffektes.

In Parameterschreibweise erhalten wir das loglineare Modell der Unabhängigkeit:

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y. \tag{5}$$

Zur Identifizierung der Parameter müssen noch bestimmte Bedingungen eingeführt werden. Führt man die *zero-sum* Bedingungen ein, ergibt sich

$$\sum \lambda_i^X = \sum \lambda_j^Y = 0.$$

Sind die Variablen nicht vollständig unabhängig, ist bei der Schätzung des Modells ein Parameter, der die Abweichung von die Unabhängigkeit beinhaltet erforderlich. Es

läßt sich zeigen, dass sich das vollständige oder saturierte loglineare Modell zu

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY} \tag{6}$$

ergibt.

Wobei folgende Nebenbedingungen als *zero-sum* Bedingungung eingeführt wird:

$$\sum_i \lambda_{ij}^{XY} = \sum_j \lambda_{ij}^{XY} = 0.$$

Außerdem läßt sich zeigen, dass sich für das Multinomiale und das Produkt-multinomiale Stichprobenmodell das identische loglineare Modell herleiten läßt. Worauf bei der Interpretation zuachten ist, da jedes Stichprobenmodell von verschiedenen Nebenbedingungen ausgeht. Bei den in dieser Arbeit betrachteten loglinearen Modellen werden wir immer vom Produkt-multinomialen Stichprobenmodell ausgehen (siehe Anhang Kapitel 6.1).

Das loglineare Modell läßt sich auch für den Fall erweitern, dass mehr als zwei Variablen untersucht werden sollen. Analog zum zuvor beschriebenen Weg, läßt sich für drei Variablen folgendes Modell aufstellen:

$$\log m_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}. \tag{7}$$

Mit dem letzten Term kommt jedoch noch der Einfluß der 3-Faktorinteraktion hinzu. Dieses Modell umfasst alle möglichen Assoziationsbeziehungen zwischen den drei Variablen. Aufgrund fehlender Signifikanz kann man beim Schätzen beliebige Parameter vernachlässigen, um so zu einem besser angepaßten Modell zu gelangen. Hinsichtlich einer ausführlichen Beschreibung des loglinearen Modells sei zum Beispiel auf Agresti (1990) verwiesen.

2 Die Logit-Transformation

Im Gegensatz zum loglinearen Modell soll nun eine kategoriale Variable als Abhängige (Response) betrachtet werden. Diese soll durch den Einfluss von verschiedenen unabhängigen Variablen, die unterschiedliches Skalenniveau haben können, erklärt werden. Ziel ist hierbei ein statistisches Modell für den Erwartungswert der Response-

Variable, die nun die einzigste Zufallsvariable des Modells ist. Dabei drücken die bedingten Wahrscheinlichkeiten die Beziehung zwischen der Response-Variable und den erklärenden Variablen aus. Die Erklärenden sind nicht stochastisch, sondern fest vorgegeben. Tabelle 1 zeigt diese Beziehung für den zweidimensionalen Fall. Wie bereits im vorigen Kapitel beschrieben, soll mit dem Modell erklärt werden, wie sich die Wahrscheinlichkeitsverteilung der Response-Variable ändert, wenn sich die erklärenden Variablen verändern.

In diesem Kapitel wird zunächst die Notwendigkeit gezeigt, die bedingten Wahrscheinlichkeiten zu tranformieren und sogenannte Logits zu bilden. Desweiteren fließen bei der Bildung der Logits die Eigenschaften der Response-Variable mit ein. So werden bei der Logitbildung für ordinale Responses die besonderen Skaleneigenschaften der Variable berücksichtigt. Der einfachste Fall stellt zunächst eine Response-Variable dar, die nur zwei Ausprägungen haben kann.

2.1 Binäre Response-Variablen [3]

Es sei die Variable Y eine binäre Zufallsvariable. Das heißt, sie kann nur zwei Ausprägungen annehmen. So können wir uns eine Variable vorstellen, die in einer einfachen Zufallsstichprobe im Erfolgsfall den Wert 1 und im Mißerfolgsfall den Wert 0 annimmt. Für die unabhängigen Realisationen der Zufallsvariable $Y_i (i = 1, ..., n)$ ergeben sich somit folgende Werte

$$y_i = \begin{cases} 1 & \text{Erfolg} \\ 0 & \text{Mißerfolg} \end{cases}$$

Die Wahrscheinlichkeiten für eine binär kodierte Variable gibt die Bernoulliverteilung an. Für den Erfolgsfall folgt daraus $P(Y_i = 1) = \pi$ und für den Mißerfolg ist $P(Y_i = 0) = 1 - \pi$.

Folgt Y_i einer Bernoulliverteilung mit dem Parameter π_i, ist die Wahrscheinlichkeitsverteilung für Y_i wie folgt bestimmt:

$$P(Y_i = y_i) = f(y_i, \pi_i) = \pi_i^{y_i}(1 - \pi_i)^{1-y_i}, \tag{8}$$

[3] vergleiche zum Beispiel Agresti(1990), Rönz (2001), u.a.

für $y_i = 0, 1$ und $i = 1, ..., n$. Der Erwartungswert einer Bernoulli-verteilten Zufallsvariable ist bekanntermaßen $E[Y_i] = \pi_i$ und die Varianz ist $Var[Y_i] = 1 - \pi_i$.

Um den Einfluß einer Reihe von Erklärenden auf die Response-Variable zu modellieren, muß bei der Stichprobe das Auftreten der Erfolges anhand von erklärenden Variablen beobachtet werden. Die Ausprägungen der erklärenden Variablen lassen sich zu diesem Zweck am besten in k verschiedene Gruppen zusammenfassen. Jede Gruppe repräsentiert somit alle statistischen Einheiten mit derselben Ausprägung in allen erklärenden Variablen und wird mit $x_i (i = 1, ..., k)$ bezeichnet. Die bedingten Wahrscheinlichkeiten für das Auftreten des Erfolges in der Gruppe k sind dann $\pi(x_i)$.

Der Erwartungswert der binären Response muß mit den erklärenden Variablen verknüpft werden. Wählt man für den bedingten Erwartungswert $E[Y|X] = \pi(x)$ der Response eine einfache lineare Regressionsgleichung, erhält man

$$E[Y|X] = \pi(x) = \beta' x_i \ .$$

Das *Lineare Wahrscheinlichkeits Modell* weist jedoch eine Reihe von strukturellen Mängeln auf. Wahrscheinlichkeiten können nur Werte zwischen 0 und 1 annehmen. Die Vorhersagegleichung des Modells würde aber auch Wahrscheinlichkeiten kleiner 0 oder größer 1 voraussagen, also aus dem gesamten Wertbereich der reellen Zahlen. Außerdem beruht das lineare Modell auf Annahmen über die normalverteilte Fehlerterme mit Mittelwert 0 und konstanter Varianz .

Schaut man sich die Fehlerterme genauer an, die sich aus

$$\begin{aligned} y &= E[Y|X] + \epsilon \\ y &= \pi(x) + \epsilon \end{aligned}$$

ergeben, würde man für den Erfolgsfall $y = 1$ folgenden Fehlerterm erhalten:

$$\epsilon = 1 - \pi(x) \text{ mit der Wahrscheinlichkeit } P(\epsilon) = \pi(x)$$

und im Mißerfolgsfall $y = 0$:

$$\epsilon = 0 - \pi(x) \text{ mit der Wahrscheinlichkeit } P(\epsilon) = 1 - \pi(x)$$

Der Erwartungswert und die Varianz des Fehlerterms ergeben sich somit zu:

$$E\left[\epsilon\right] = \left[1 - \pi\left(x\right)\right]\pi\left(x\right) - \pi\left(x\right)\left[1 - \pi\left(x\right)\right] = 0$$

$$Var\left[\epsilon\right] = \pi\left(x\right)\left[1 - \pi\left(x\right)\right].$$

Man sieht also, dass hier keine normalverteilten Fehlerterme vorliegen und die Annahme einer konstanten Varianz verletzt wird. Im allgemeinen wird man bei der Modellierung von Wahrscheinlichkeiten in Abhängigkeit von Erklärenden auch von einem nichtlinearen Zusammenhang ausgehen. Zum Beispiel möchte man die Erfolgswahrscheinlichkeit für den Kauf eines Autos in Abhängikeit vom jährlichen Einkommen einer Familie modellieren. In diesem Fall würde man erwarten, dass ein Anstieg des jährlichen Einkommens um 10.000 Euro unterschiedliche Effekte auf die Wahrscheinlichkeit einer Kaufentscheidung hat. Wenn das jährlich Einkommen einer Familie bei 1 Mio. Euro liegt, würde man einen kleineren Effekt auf die Wahrscheinlichkeit erwarten, als wenn das Einkommen bei 30.000 Euro liegt.

Um die oben beschriebenen Mängel zu beseitigen muß vorher eine Tranformation der Wahrscheinlichkeiten vorgenommen werden.

Ein erster Schritt ist die Einführung von sogenannten Odds

$$odds_i = \frac{\pi_i}{1 - \pi_i}. \tag{9}$$

Wie leicht zu erkennen ist, können die Odds nun jeden positiven Wert annehmen. Das Odds ist im binären Fall das Verhältnis der Wahrscheinlichkeit eines Erfolges zur Wahrscheinlichkeit eines Misserfolges.

Im zweiten Schritt wird der Wertebereich auch auf den Bereich der negativen Zahlen ausgeweitet und so die Restriktion auf den Bereich der positiven Zahlen beseitigt. Dieser Schritt beseht einfach im Logarithmieren der Odds (der natürliche Logarithmus wird der Einfachhalber mit log angegeben) . Dadurch erhält man die sogenannten Log-Odds, die auch als Logits bezeichnet werden

$$L_i = \eta_i = logit(\pi_i) = \log\left[\frac{\pi_i}{1 - \pi_i}\right]. \tag{10}$$

Jetzt ist der Wertebereich der transformierten Wahrscheinlichkeiten wie gefordert auf alle reellen Zahlen ausgeweitet. Da die Logits den Anforderungen eines linearen Mo-

dells genügen, können diese anstatt der Wahrscheinlichkeiten zur Modellbildung benutzt werden. Man erhält ein Modell der logistischen Regression, in dem η_i die Verknüpfung (Link-Funktion) mit den Vorhersagevariablen angibt.

Wird die Erfolgswahrscheinlichkeit immer kleiner, also wenn sie gegen Null geht, bewegen sich auch die Odds gegen Null. Die Logits nehmen dabei negative Werte an und gehen gegen -∞. Werden anderseits die Wahrscheinlichkeiten immer größer, also bewegen sie sich gegen 1, gehen die Odds gegen +∞. Wobei auch die Logits gegen +∞ gehen. Beträgt die Erfolgswahrscheinlichkeit genau $\frac{1}{2}$, sind die Odds genau 1 und die Logits genau 0. Damit wird ersichtlich, dass bei Wahrscheinlichkeiten kleiner $\frac{1}{2}$ die Logits negativ werden und bei Wahrscheinlichkeiten größer $\frac{1}{2}$ die Logits positiv sind.

Durch die inverse Transformation kann man die Wahrscheinlichkeiten auch als Verteilungsfunktion modellieren:

$$\pi_i(\eta_i) = F(\eta_i) = P(Y = 1|x_i) = \frac{\exp \eta_i}{1 + \exp \eta_i}. \tag{11}$$

Diese Verteilung wird auch als logistische Verteilung bezeichnet. Setzt man andere Verteilungsfunktionen für die Wahrscheinlichkeiten ein, so sind auch andere Transformationen der Wahrscheinlichkeiten möglich. Darauf soll hier aber nicht weiter eingegangen werden. Nach der Logit-Transformation lassen sich die Odds, also dem Verhältnis der Wahrscheinlichkeit des Erfolgs zur Wahrscheinlichkeit des Mißerfolgs wie folgt darstellen:

$$odds_i = \frac{\pi_i}{1 - \pi_i} = \exp(\eta_i). \tag{12}$$

2.2 Kategoriale Response-Variablen

Auch bei Response-Variablen, die verschiedene diskrete Werte annehmen können, ist es möglich, die Wahrscheinlichkeiten des Auftretens einer Kategorie in Abhängigkeit von erklärenden Variablen zu modellieren. Wie schon im binären Fall ist zunächst eine Transformation der Wahrscheinlichkeiten notwendig. Während es im binären Fall jedoch nur zwei Wahrscheinlichkeiten je Gruppe der Erklärenden gab, existieren im multinomialen Fall so viele Wahrscheinlichkeiten je Gruppe, wie es Kategorien für die Response-Variable gibt.

Der erste Schritt bei der Logit-Transformation ist zunächst auf die Bildung der Odds gerichtet, also dem Verhältnis zweier Wahrscheinlichkeiten. Nun braucht man bei mehr-

kategorialen Response-Variablen nicht alle Wahrscheinlichkeiten zueinander ins Verhältnis zu setzten. Es genügt die Wahl einer Kategorie der Response-Variable als Basis. Zu dieser Basis werden alle Wahrscheinlichkeiten ins Verhältnis gesetzt. Es ist im allgemeinen gebräuchlich, die letzte Kategorie der Response-Variable als Basis zu wählen.

Es sei Y eine kategoriale Response-Variable mit J verschiedenen Ausprägungen. Die Erklärenden werden in i Gruppen unterteilt. Dann ist es möglich, für jede Gruppe der Erklärenden $J-1$ verschiedene Verhältnisse der Wahrscheinlichkeit des Eintretens der Kategorie j zur Wahrscheinlichkeit des Eintretens der J-ten Referenzkategorie zu bilden. Die Logits sind somit

$$L_{ij} = \eta_{ij} = \log\left[\frac{\pi_{ij}}{\pi_{iJ}}\right] \qquad j = 1, ..., J-1. \tag{13}$$

Die $J-1$ Logits zur Referenzkategorie können auch benutzt werden, um die Logits für alle möglichen Paare (g, k) von Response-Kategorien zu bilden. Es gilt die folgende Beziehung

$$\log\left(\frac{\pi_{ig}}{\pi_{ik}}\right) = \log\left(\frac{\pi_{ig}}{\pi_{iJ}}\right) - \log\left(\frac{\pi_{ik}}{\pi_{iJ}}\right). \tag{14}$$

Man sieht, dass die $J-1$ Logits zur Referenzkategorie J ausreichen, um alle möglichen Logits zu bestimmen.

Für die inverse Transformation der Wahrscheinlichkeiten kann (13) auch wie folgt geschrieben werden

$$\pi_{ij} = \pi_{iJ} \exp\{\eta_{ij}\}. \tag{15}$$

Wie schon in Kapitel 1.1 beschrieben wurde, addieren sich die bedingten Wahrscheinlichkeiten für jede Kategorie von X zu 1, $\sum_j \pi_{ij} = 1$. Um diese Bedingung zu erfüllen und für alle j gültig zu machen, setzt man $\eta_{iJ} = 0$ und damit $\exp\{\eta_{iJ}\} = 1$. Addiert man nun alle Wahrscheinlichkeiten zu

$$\sum_j^J \pi_{ij} = 1 = \pi_{iJ} \sum_j^J \exp\{\eta_{ij}\} \qquad j = 1, ..., J \tag{16}$$

erhalten wir die Formel für die bedingte Wahrscheinlichkeit der Referenzkategorie

$$\pi_{iJ} = P(Y = J|x_i) = \frac{1}{\sum_j^J \exp\{\eta_{ij}\}} = \frac{1}{1 + \sum_j^{J-1} \exp\{\eta_{ij}\}}. \tag{17}$$

Für die Wahrscheinlichkeiten der anderen Kategorien folgt damit

$$\pi_{ij} = F(\eta_{ij}) = P(Y = j|x_i) = \frac{\exp\{\eta_{ij}\}}{\sum_j^J \exp\{\eta_{ij}\}} = \frac{\exp\{\eta_{ij}\}}{1 + \sum_j^{J-1} \exp\{\eta_{ij}\}} \; j = 1, ..., J-1. \tag{18}$$

Wie hier zu sehen ist, lassen sich die Wahrscheinlichkeiten wieder als Verteilungsfunktion darstellen. Bei der Logit-Transformation der bedingten Wahrscheinlichkeiten für eine kategoriale Response-Variable erfüllen die $\pi_{ij}(i = 1, ...J)$ die multinomiale logistische Verteilung.

2.3 Ordinale Response - Variablen

Bei ordinalen Response-Variablen haben die Kategorien eine natürliche Rangfolge. Die Rangordnung der Kategorien einer Response-Variable kann dazu benutzt werden, zusätzliche Aussagen über die Wahrscheinlichkeiten des Eintretens der Kategorien zu treffen. So kann man sich zum Beispiel für die Wahrscheinlichkeit des Eintretens der ersten j Kategorien interessieren. Oder man möchte Aussagen über das Eintreten der letzten $J - j$ Kategorien treffen. Anders als bei den bisher betrachteten nominalen kategorialen Response-Variablen kann es bei ordinalen Response-Variablen durchaus sinnvoll sein, mehr als zwei Wahrscheinlichkeiten für das Eintreten der Kategorien zu betrachten. Natürlich ist es auch möglich für ordinale Response-Variablen Logits zu einer Referenz-Kategorie zu bilden. Nachfolgend wird dargestellt, wie bei der Transformation der Wahrscheinlichkeiten der ordinale Charakter der Response in die Logits mit einfließt. Unter den vielfältigen Möglichkeiten Logits für ordinale Variablen zu bilden, hat Agresti (1990) drei verschiedene Möglichkeiten dargestellt. Auf den Index für die Erklärenden soll hier verzichtet werden. L_j steht im folgenden für das j-te Logit.

Kumulative Logits (Cumulative Logits)

Für diese Version der Logit-Transformation benutzt man kumulierte Wahrscheinlichkeiten. Kumulierte Wahrscheinlichkeiten seien wie folgt definiert: $F_j = \pi_1 + ... + \pi_j$. Es soll das Verhältnis der Wahrscheinlichkeit des Eintretens der ersten j Kategorien $P(Y \leq j|x)$ zur Wahrscheinlichkeit des Eintretens der letzten $J - j$ Kategorien $P(Y > j|x)$ gebildet werden. Das Logit lautet dann

$$L_j = \eta_j = \log\left[\frac{P(Y \leq j|x)}{P(Y > j|x)}\right] = \log\left[\frac{F_j}{1-F_j}\right] = \log\left[\frac{\pi_1 + \ldots + \pi_j}{\pi_{j+1} + \ldots + \pi_J}\right] \quad j = 1, \ldots, J-1. \tag{19}$$

Für eine Response mit J geordneten Kategorien gibt es somit $J-1$ Möglichkeiten, kumulierte Wahrscheinlichkeiten zueinander ins Verhältnis zu setzten. Die kumulativen Logits nutzen dabei alle J Kategorien zur Formulierung jedes einzelnen Logits. Man kann sehen, dass dies eine Form eines binären Logits ist, bei der die Kategorien 1 bis j zu einer Kategorie und die Kategorien $j+1$ bis J zu einer anderen Kategorie zusammengefaßt werden.

Continuation-Ratio Logits

Hier betrachten wir das Verhältnis der Wahrscheinlichkeit des Eintretens der Kategorie j $P(Y = j|x)$ zur Wahrscheinlichkeit des Eintretens aller Kategorien höher j $P(Y > j|x)$

$$L_j = \log\left[\frac{P(Y = j|x)}{P(Y > j|x)}\right] = \log\left[\frac{\pi_j}{\pi_{j+1} + \ldots + \pi_J}\right] \qquad j = 1, \ldots, J-1 \tag{20}$$

oder aber die Wahrscheinlichkeit des Eintretens der Kategorie $j+1$ $P(Y = j+1|x)$ zur Wahrscheinlichkeit des Eintretens aller Kategorien darunter $P(Y < j+1|x)$.

$$L_j^* = \log\left[\frac{P(Y = j+1|x)}{P(Y < j|x)}\right] = \log\left[\frac{\pi_{j+1}}{\pi_1 + \ldots + \pi_j}\right] \qquad j = 1, \ldots, J-1. \tag{21}$$

Nachbarschafts-Logits (Adjacent-Categories Logits)

Für zwei jeweils aufeinander folgende Kategorien lassen sich auch die sogannten Nachbarschafts-Logits bilden. Wie schon bei den nominal-kategorialen Logits werden zwei Wahrscheinlichkeiten ins Verhältnis gesetzt. Nur drückt die Tatsache, dass die Kategorien aufeinander folgen, hier den besonderen ordinalen Charakter der Response aus.

$$L_j = \log\left[\frac{\pi_j}{\pi_{j+1}}\right] \qquad j = 1, \ldots, J-1 \tag{22}$$

3 Die Logistische Regression

Nachdem die Wahrscheinlichkeiten in eine geeignete Form transformiert wurden, können sie in Abhängigkeit von erklärenden Variablen modelliert werden. Das binäre und multinomiale Logit-Modell soll zunächst so dargestellt werden, dass es sowohl für kategoriale als auch für metrische Erklärende angepasst werden kann. Für kategoriale Variablen ist zusätzlich eine Kodierung mit Kontrastvariablen notwendig. Diese dient der Vermeidung einer Überparametrisierung. Es soll außerdem gezeigt werden, dass man für die Logit-Modelle eine *zero-sum* Restriktion für die Parameter einführen kann. Schließlich kann die Existenz einer engen Beziehung zwischen loglinearen und den Logit Modellen aufgezeigt werden. Die Korrespondenz zwischen beiden Modellarten führt dazu, dass man kategoriale Erklärende in nominale und ordinale Erklärende unterscheiden kann. Später soll gezeigt werden, daß es Modelle für Nachbarschafts-Logits gibt, die sich als loglineare Modelle für ordinale Variablen darstellen lassen.

3.1 Das binäre Logit-Modell

In einer einfachen Zufallsstichprobe vom Umfang n wird die Anzahl der Erfolge einer binären Response-Variable Y für k verschiedene Gruppen von Kovariablen erhoben (der Laufindex i steht für die verschiedenen Gruppen mit $i = 1, ..., k$). Alle Beobachtungen sind damit unabhängig. Für jede Gruppe der Kovariablen sei eine konstante Erfolgswahrscheinlichkeit angenommen. Die Anzahl der Erfolge in jeder der k Gruppen sei binomial verteilt (siehe Anhang Kapitel 6.1).

Dann läßt sich für die bedingten Wahrscheinlichkeiten, gegeben der Ausprägung von X das binäre Logit-Modell anpassen. Die systematische Struktur sei durch die lineare Modellierung der Logits, wie sie im vorigem Kapitel für binäre Response-Variablen beschrieben wurden, gegeben

$$L_i = logit\left(\pi\left(x_i\right)\right) = \log\left[\frac{\pi(x_i)}{1-\pi(x_i)}\right] = \beta' x_i. \tag{23}$$

β ist hier ein Parametervektor, der zu schätzen ist. Die einzelnen Koeffizienten in diesem Vektor β_j repräsentieren die Veränderung der Logits der Wahrscheinlichkeiten, wenn sich eine X-Variable um eine Einheit verändert, während die anderen Erklärenden konstant gehalten werden.

Die Odds für einen Erfolg sind dann

$$\frac{\pi(x_i)}{1-\pi(x_i)} = \exp\{\beta' x_i\}\,. \tag{24}$$

Womit wir ein multiplikatives Modell für die Odds erhalten. Verändert sich der j-te Wert von X um eine Einheit, wobei alle anderen konstant gehalten bleiben, werden die Odds mit $\exp\{\beta_j\}$ multipliziert. Dies zeigt sich wie folgt:

$$\beta' x_i + \beta_j \text{ die Veränderung um eine Einheit von } \mathrm{x_j}$$

$$\text{führt zur Veränderung der odds : } \exp\{\beta' x_i\} \text{ mal } \exp\{\beta_j\}\,.$$

Für die Erfolgswahrscheinlichkeiten ergibt sich folgende nichtlineare Funktion

$$\pi(x_i) = \frac{\exp\{\beta' x_i\}}{1+\exp\{\beta' x_i\}} \tag{25}$$

Wie bereits dargelegt, ist dies die logistische Verteilungsfunktion.

3.2 Das multinomiale Logit-Modell

In einer einfachen Zufallsstichprobe vom Umfang n wird die Anzahl der Erfolge einer binären Response-Variable Y für k verschiedene Gruppen von Kovariablen erhoben (der Laufindex i steht für die verschiedenen Gruppen mit $i = 1, ..., k$). Alle Beobachtungen sind damit unabhängig. Für jede Gruppe der Kovariablen sei eine konstante Erfolgswahrscheinlichkeit angenommen. Die Anzahl der Erfolge in jeder der k Gruppen sei binomialverteilt (siehe Anhang 6.1).

Sind diese Annahmen erfüllt, läßt sich für die bedingten Wahrscheinlichkeiten ein multinomiales Logit-Modell schätzen. Wir gehen hier von den im vorigen Kapitel beschriebenen Logits für kategoriale Response-Varaiblen, mit der letzten Kategorie als Referenz-Kategorie aus. Die Logits sollen wieder einem linearen Modell folgen

$$L_{ij} = \eta_{ij} = \log\left[\frac{\pi_j(x_i)}{\pi_J(x_i)}\right] = \alpha_j + \beta_j' x_i \qquad j = 1, ..., J-1 \tag{26}$$

Im Vergleich zum binären Logit-Modell ist hier eine Konstante a_j explizit angegeben und β_j ist ein Vektor von Regressionskoeffizienten. Insgesamt gibt es $J-1$ multi-

nomiale Logit-Gleichungen, für die ein Parametervektor zu schätzen ist. Die Wahrscheinlichkeiten ergeben sich dann zu

$$\pi_j(x_i) = \frac{\exp\left(\beta_j' x_i\right)}{\sum_h^J \exp\left(\beta_j' x_i\right)} \qquad j = 1, ..., J-1 \tag{27}$$

Man errinnere sich an die Nebenbedingung, die für die Identifikation nötig ist. Für die Bedingung $\eta_{iJ} = 0$ wird $\beta_J = 0$ gesetzt.

3.3 Zur Korrespondenz zwischen dem Logit-Modell und dem loglinearen Modell[4]

Während loglineare Modelle die erwartete Häufigkeit in Assoziations- und Interaktionstermen zwischen allen Variablen ausdrücken, betrachten die Logit-Modelle die strukturellen Eigenschaften zwischen den erklärenden Variablen nicht. Obwohl die Logit-Modelle nur die Effekte der Erklärenden auf die Response-Variable wiederspiegeln, gibt es jedoch eine Beziehung zwischen beiden Modellarten. Denn es besteht die Möglichkeit das jeweilige Logit-Modell aus dem loglinearen Modell zu erhalten. Dieses loglineare Modell muß jedoch dieselbe Struktur für Assoziation zwischen der Response und den erklärenden Variablen haben. Darüber hinaus sollte die allgemeinste Interaktion für die Beschreibung der Beziehung zwischen den erklärenden Variablen und der Response-Variablen in dem Modell enthalten sein.

Um die Korrespondenz zwischen dem Logit- und dem loglinearen Modell besser zu verdeutlichen, sei hier ein binäres Logit- Modell für den dreidimensionalen Fall angegeben. Es sei Z eine Response-Variable mit k Kategorien und X und Y seien zwei erklärende Variablen mit jeweils i und j Kategorien. Sind beide Erklärenden nominal, ergibt sich folgende Darstellung des Logit-Modells

$$\log\left(\frac{\pi_{ij2}}{\pi_{ij1}}\right) = \log\left(\frac{m_{ij2}}{m_{ij1}}\right) = \alpha + \tau_i^X + \tau_j^Y. \tag{28}$$

Mit den zusätzlichen *zero-sum* Bedingungen für die Parameter $\sum \tau_i^X = \sum \tau_j^Y = 0$. In dieser Darstellung ist die Response-Variable Z mit den nominalen Variablen X und Y assoziiert. Ziel ist ein loglineares Modell zu finden, dass dieser Assoziation gerecht

[4] Siehe Agresti(1984)...

wird. Deshalb sollten die Parameter λ_{ik}^{XZ} und λ_{jk}^{YZ} im loglinearen Modell (vergleiche Kapitel 1.1) enthalten sein. Außerdem ist noch der Assoziationsterm zwischen den Erklärenden λ_{ij}^{XY}einzuschließen, um zu erreichen, dass die Randhäufigkeiten für alle X -Y Kombinationen den beobachteten entsprechen. Dies dient dazu, die Randhäufigkeiten von (X, Y) als fix anzusehen, um damit das Produkt-Multinomiale Stichprobenmodell zu reproduzieren (vergleiche Anhang Kapitel 6.1) , was man allgemein bei Logit-Modellen annimmt. Gehen wir also von folgendem loglinearen Modell aus:

$$\log m_{ijk} = \mu + \lambda_i^X \ + \ \lambda_j^Y + \lambda_k^Z \ + \ \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} \ + \ \lambda_{jk}^{YZ} \tag{29}$$

Das Logit-Modell (28) kann auch als Differenz dargestellt werden und das loglineare Modell (29) eingesetzt werden

$$\begin{aligned} \log\left(\frac{m_{ij2}}{m_{ij1}}\right) &= \log(m_{ij2}) - \log(m_{ij1}) \\ &= \left[\mu + \lambda_i^X \ + \ \lambda_j^Y + \lambda_2^Z \ + \ \lambda_{ij}^{XY} + \lambda_{i2}^{XZ} \ + \ \lambda_{j2}^{YZ}\right] \\ &\quad - \left[\mu + \lambda_i^X \ + \ \lambda_j^Y + \lambda_1^Z \ + \ \lambda_{ij}^{XY} + \lambda_{i1}^{XZ} \ + \ \lambda_{j1}^{YZ}\right] \\ &= \left(\lambda_2^Z \ - \lambda_1^Z\right) + \left(\lambda_{i2}^{XZ} - \lambda_{i1}^{XZ}\right) + \left(\lambda_{j2}^{YZ} - \lambda_{j1}^{YZ}\right) \end{aligned} \tag{30}$$

Man beachte, dass aufgrund der dichotomen Response-Variable Z und der *zero-sum* Restriktion folgendes gilt: $\lambda_2^Z \ = -\lambda_1^Z$ und $\lambda_{i2}^{XZ} = -\lambda_{i1}^{XZ}$ und $\lambda_{j2}^{YZ} = -\lambda_{j1}^{YZ}$. Letztendlich erhält man das Logit-Modell

$$\log\left(\frac{m_{ij2}}{m_{ij1}}\right) = \left(2\lambda_2^Z\ \right) + \left(2\lambda_{i2}^{XZ}\right) + \left(2\lambda_{j2}^{YZ}\right) \tag{31}$$

Zur Korrespondenz seien außerdem folgende Beziehungen gegeben:

$2\lambda_{i2}^{XZ}$ = τ_i^X : i-ter Effekt von X auf die Logit von Z

$2\lambda_{j2}^{YZ}$ = τ_j^Y : j-ter Effekt von Y auf die Logit von Z

$2\lambda_2^Z$: α

λ_{ij}^{XY}: der Assoziationparameter zwischen den Erklärenden fällt in der Differenz raus, die Assoziation wird im Logit-Modell nicht erklärt.

Für weitere Parameterinterpretationen sei auf Agresti (1984) verwiesen.

3.4 Logit-Modelle für ordinale Response-Variablen[5]

Möchte man Modelle spezifieren, die alle Informationen nutzten, die ordinale Response-Variablen bieten, ist es wichtig, auf die Unterschiede zu Modellen für nominale Variablen oder Variablen höherer Skalen zu achten. Denn wie Tutz betont:

> „Modelle, die für nominales Skalenniveau zulässig sind, sind es auch für höheres, z.B. ordianles Skalenniveau. Umgekehrt kann ein für ordinales Skalenniveau zulässiges Modell auch für kategoriales Skalenniveau zulässig sein. Interessant sind jedoch ordinale Modelle, für die das nicht der Fall ist. Sie erst machen wirklich Gebrauch von der ordinalen Struktur der Variablen ...”[6]

Folgt man Tutz und beschränkt sich auf Modelle, die die Rangfolge in der Response-Variablen berücksichtigen, ergeben die oben beschriebenen Logits für ordinale Variablen eine recht gute Grundlage, solche Modelle aufzustellen. Seien jetzt $\{\pi_1(x), ..., \pi_J(x)\}$ Response-Wahrscheinlichkeiten bedingt auf die Ausprägung von x, also einer Anzahl von erklärenden Variablen.

3.4.1 Nachbarschafts-Modelle (Adjacent-Categories-Models)

Die Nachbarschafts-Logits wurden bereits im vorigen Kapitel dargestellt.

$$L_j = \log\left[\frac{\pi_j(x)}{\pi_{j+1}(x)}\right] \qquad j = 1, \ldots, J-1. \tag{32}$$

Wie bereits bemerkt, sind alle $\begin{pmatrix} J \\ 2 \end{pmatrix}$ Paare von möglichen Response-Kategorien durch die Logits zur Referenzkategorie J determiniert. Die Nachbarschafts-Logits lassen sich somit mit den Referenzkategorie-Logits

$$L_j^* = \log\left[\frac{\pi_j(x)}{\pi_J(x)}\right], \qquad j = 1, ..., J-1 \tag{33}$$

[5] vergleiche Agresti(1984), Agresti(1990),Fahrmeier & Tutz (1994,2001) und Tutz (1990)

[6] Tutz(1990), S.71; siehe auch ebenda für eine detaillierte Betrachtung von Invarianzeigenschaften von ordinalen Regressionsmodellen.

schätzen. Möchte man nun ein Nachbarschaftsmodell schätzen, so dass

$$L_j = a_j + \beta' x, \tag{34}$$

nutzt man die Beziehung $L_j^* = L_j + L_{j+1} + ... + L_{J-1}$ aus, um folgendes multinomiale Logit-Modell zu erhalten

$$\begin{aligned} L_j^* &= \textstyle\sum_{k=j}^{J-1} \alpha_k + \beta'(J-j)x, \qquad j = 1, ..., J-1 \\ &= \alpha_j^* + \beta' u_j \qquad j = 1, ..., J-1, \end{aligned} \tag{35}$$

mit $u_j = (J-j)x$.

Es soll nun gezeigt werden, wie loglineare Modelle für ordinale Variablen sich mit Hilfe von Nachbarschafts-Logit-Modellen darstellen lassen.

Das *Uniform Association Model* für ordinale Erklärende

Wir gehen hier von einer zweidimensionalen Kontigenztabelle (siehe Kapitel 1) aus, und betrachten zunächst das loglineare Modell (man beachte das Produkt-multinomiale Stichprobenmodell, siehe Anhang 6.1). Es sei Y eine ordinale Variable in den Spalten und X eine ordinale Variable in den Zeilen. Um ein angemessenes Modell zu erhalten, muß den Variablen sogenannte Scores zugeordnet werden, die die Rangordnung der Variablen reflektieren. Nimmt man nun hier die Scores $\{u_i\}$ für die Zeilen und $\{v_j\}$ für die Spalten einer zweidimensionalen Kontingenztabelle, muß für die Scores die Bedingung der Rangfolge gelten: $u_1 \leq u_2 \leq ... \leq u_I$ bzw. $v_1 \leq v_2 \leq ... \leq v_J$. Das loglineare Modell für die zu erwartenden logarithmierten Zellhäufigkeiten sei nun folgendes

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y + \beta u_i v_j. \tag{36}$$

Dieses Modell ist nicht saturiert. Man muß zur Identifizierung der Parameter auch hier wieder die Bedingungen $\sum \lambda_i^X = \sum \lambda_j^X = 0$ einführen. In diesem loglinearen Modell gibt β die Stärke der Assozisation zwischen beiden Variablen an. Und unabhängig von der Anzahl der Zeilen und Spalten soll hier nur ein einziger Parameter diese Assoziation ausdrücken. Im Unabhängigkeits-Modell wäre $\beta = 0$.

Zur besseren Interpretation des Modells sollen die Scores jeweils den gleichen Abstand

zueinander haben. So dass gilt:

$$u_2 - u_1 = u_3 - u_2 = ... = u_I - u_{I-1}$$

$$v_2 - v_1 = v_3 - v_2 = ... = v_J - v_{J-1}$$

Dieses Ergebnis erhält man zum Beispiel, wenn die Scores die Nummer der Kategorie wiedergeben. So dass $\{u_i = i\}$ und $\{v_j = j\}$ ist. Der Abstand der Scores ist gleichmäßig immer eins. Manchmal ist es auch ganz hilfreich, die Scores linear zu transformieren, um die Scores zu standardisieren.

$$\begin{array}{rclcl} \sum u_i \pi_{i+} & = & \sum v_j \pi_{+j} & = & 0 \\ \sum u_i^2 \pi_{i+} & = & \sum v_j^2 \pi_{+j} & = & 1 \end{array}$$

Dieses Modell wird *linear-by-linear association model* genannt. Zu weiteren Ausführungen sei auf Agresti (1990) verwiesen.

Zur Interpretation des Parameters β in diesem Modell kann die Darstellung mit Hilfe von lokalen *odds ratios* θ_{ij} verwendet werden. Für ein beliebiges Paar von benachbarten Zeilen (Kategorien von X) und ein beliebiges Paar benachbarter Spalten (Kategorien von Y) gilt dann,

$$\log(\theta_{ij}) = \log\left(\frac{m_{ij} m_{i+1,j+1}}{m_{i,j+1} m_{i+1,j}}\right) = \beta\,(u_{i+1} - u_i)\,(v_{j+1} - v_j)\,. \tag{37}$$

In dem Spezialfall, bei dem die Abstände zwischen den Scores gleichmäßig sind, sind die lokalen *odds ratios* alle gleich. Dieses Modell wird in der Literatur auch als *uniform association model* bezeichnet (Vergleiche Goodmann(1979)). Beträgt der Abstand zwischen den Scores genau Eins, gilt für alle

$$\log \theta_{ij} = \beta \tag{38}$$

und alle

$$\theta_{ij} = \exp(\beta)\,. \tag{39}$$

Woraus sich die einfache Interpretation von β als das gemeinsame lokale *log odds ratio* ergibt. Aus der Wahl von Unit-Scores ergibt sich somit auch ein konstantes lokales log odds ratio für alle jeweils benachbarten Zellwahrscheinlichkeiten.

Dieses Ergebnis erlaubt es uns nun, Aussagen über die Wahrscheinlichkeit eine höhere (niedrigere) Kategorie von Y zu erhalten, wenn bei X eine höhere (niedrigere) Kategorie auftritt. Zu diesem Zweck kann man aus dem *linear-by-linear association model* ein Logit-Modell bilden, indem man Y als Response-Variable in den Spalten und X als Erklärende in den Zeilen betrachtet. Es liegt nahe, für dieses Logit-Modell die Nachbarschafts-Logits zu verwenden. Man erhält für die Nachbarschafts-Logits folgende Formulierung

$$\log\left(\frac{\pi_{j+1|i}}{\pi_{j|i}}\right) = \log\left(\frac{m_{i,j+1}}{m_{ij}}\right) = \left(\lambda_{j+1}^{Y} - \lambda_{j}^{Y}\right) + \beta\left(v_{j+1} - v_j\right) u_i. \tag{40}$$

Das *uniform association model* mit Score-Abstand 1 vereinfacht sich zu ,

$$\log\left(\frac{\pi_{j+1|i}}{\pi_{j|i}}\right) = \alpha_j + \beta u_i, \tag{41}$$

wobei $a_j = \lambda_{j+1}^{Y} - \lambda_{j}^{Y}$ ist. In diesem Nachbarschafts-Logit-Modell tritt also für alle benachbarten Response-Kategorien simultan derselbe lineare Logit-Effekt auf. Die Interpretation ist einfach: die Odds, als Maß dafür um wieviel wahrscheinlicher das Eintreten der Kategorie $j+1$ anstelle der Kategorie j ist, werden mit $\exp(\beta)$ multipliziert, wenn sich die Erklärende X um eine Einheit ändert. Die Verwendung von Unit-Scores führt somit zu der impliziten Annahme, dass der Effekt von X auf alle $J-1$ Logits von Y gleich ist.

In diesem Modell sind also die Wahrscheinlichkeitsverteilungen der Y bedingt auf die Ausprägung der Variable X stochastisch geordnet. Das bedeutet, es ist wahrscheinlicher, eine höhere Kategorie von Y zu erhalten mit höherer Kategorie von X wenn $\beta > 0$. Und es ist wahrscheinlicher, eine niedrigere Kategorie von Y zu erhalten mit höher Kategorie von X, wenn $\beta < 0$.

Das Zeilen- Effekt-Modell (*Row-Effects-Model*) für nominale Erklärende

Die geordneten Scores $v_1 \leq v_2 \leq \ldots \leq v_J$ werden jetzt nur noch auf die Spaltenvariable angewandt (Y ist ordinal). Die Zeilen sollen nun einer nominalen Erklärenden (X) zugeordnet sein. Man braucht also keine geordnete Scores für die Erklärende. Nehmen wir wieder ein loglineares Modell wie oben als *linear-by-linear* Modell (man beachte das Produkt-multinomiale Stichprobenmodell; siehe Anhang 6.1). Die geord-

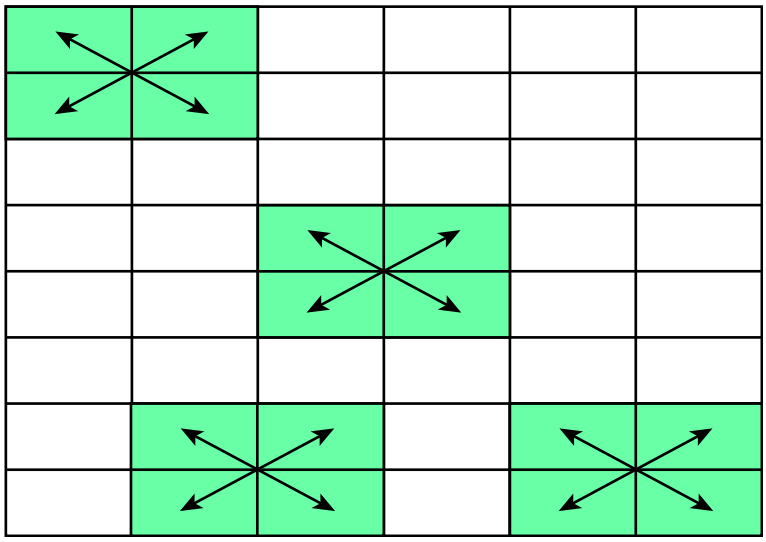

Abbildung 1: Konstante *odds-ratios* im *uniform association model*

neten Werte $\{\beta u_i\}$ sind jedoch durch die ungeordneten Parameter $\{\mu_i\}$ ersetzt worden. Somit erhalten wir folgendes Zeilen-Effekt-Modell:

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y + \mu_i v_i \tag{42}$$

Die $\{v_j\}$ sind feste Konstanten und die Parameter $\{\mu_i\}$ werden auch als Zeilen-Effekte bezeichnet. Zur Identifizierung der Parameter braucht man auch hier wieder folgende Bedingungen $\sum \lambda_i^X = \sum \lambda_j^Y = \sum \mu_i = 0$. Interpretiert man die $\{\mu_i\}$ mit Hilfe der *odds ratio*'s und geht von Unit-Scores $\{v_j\}$ aus, ergibt sich für die *log odds ratio*'s für beliebige Paare von Zeilen (Kategorien der nominalen Variablen X) h und l und benachbarte Spalten (Kategorien der ordinalen Variablen Y) j und $j+1$

$$\log\left(\frac{m_{hj}m_{l,j+1}}{m_{h,j+1}m_{lj}}\right) = \mu_l - \mu_h. \tag{43}$$

Für benachbarte Spalten oder Kategorien der ordinalen Variablen Y, sind die *log odds ratios* somit identisch. Und dies gilt auch für alle möglichen $J-1$ Paare von benachbarten Kategorien der ordinalen Variablen. Das hat den Vorteil, dass man nun für verschiedene Kategorien von X Aussagen über die Wahrscheinlichkeit treffen kann, höher oder niedriger auf der Skala von Y auszufallen.

Die Differenz $\{\mu_l - \mu_h\}$ beschreibt nämlich den Unterschied zwischen den Wahrscheinlichkeitsverteilungen von Y bedingt auf die jeweilige Ausprägung (Zeile) von

X. Wenn $\mu_l = \mu_h$, ist Y in beiden Zeilen (beiden Kategorien von X) identisch verteilt. Wenn $\mu_l > \mu_h$, ist die Verteilung von Y stochastisch in Zeile l höher als in Zeile h, was bedeutet, dass es für Zeile l wahrscheinlicher ist, Beobachtungen auf höherer Rangstufe von Y zu erhalten als für Zeile h. Mit steigendem μ_i, neigt die Response also auf höherer Rangstufe auszufallen.

Nimmt man Y nun als ordinale Response-Variable und leitet aus dem loglinearen Modell wieder ein Logit-Modell her:

$$\log\left(\frac{\pi_{j+1|i}}{\pi_{j|i}}\right) = \log\left(\frac{m_{i,j+1}}{m_{ij}}\right) = \left(\lambda_{j+1}^{Y} - \lambda_{j}^{Y}\right) + \mu_i\left(v_{j+1} - v_j\right). \tag{44}$$

Für Unit-Scores in den Spalten ergibt sich

$$\log\left(\frac{\pi_{j+1|i}}{\pi_{j|i}}\right) = \alpha_j + \mu_i. \tag{45}$$

Somit erhält man ein einfaches Logit-Modell für eine ordinale Response-Variable und eine nominale Erklärende. Wobei natürlich dieses Modell auch für ordinale Erklärende angewandt werden kann. Die Zeilen-Effekte μ_i geben an wieviel wahrscheinlicher es ist, eine Response in Kategorie $j+1$ anstatt in Kategorie j zu erhalten. Der Effekt in Zeile i oder der i-ten Kategorie von X ist identisch für alle Paare von benachbarten Responses. Plottet man diese Logits gegen i, verlaufen die Plots für jede Kategorie von Y parallel. Goodman (1983) bezeichnet dieses Modell auch als *parallel odds model.*

3.4.2 Das kumulative Logit-Modell

Modelle für kumulative Logits gehen zurück auf Edward und Thurtsone (1952). Die kumulativen Logits in Abhängigkeit von den Erklärenden x seien nun wie folgt definiert:

$$L_j = \log\left[\frac{P(Y \leq j|x)}{P(Y > j|x)}\right] = \log\left[\frac{F_j(x)}{1-F_j(x)}\right] = \log\left[\frac{\pi_1(x)+\ldots+\pi_j(x)}{\pi_{j+1}(x)+\ldots+\pi_J(x)}\right] \quad j = 1, \ldots, J-1. \tag{46}$$

Wie bereits dargelegt, können diese Logits auch als Logit-Modell für binäre Response-Variablen geschätzt werden. Da es $J-1$ Möglichkeiten gibt, diese Logits zu formen, müßten $J-1$ Modelle angepaßt werden. Hier soll beschrieben werden, wie alle $J-1$ Logits simultan geschätzt werden können.

Das einfachste Modell ergibt sich, wenn die Response-Variable simultan unabhängig von allen Erklärenden ist

$$L_j = \alpha_j \qquad j = 1, \ldots, J-1. \tag{47}$$

Die $\{\alpha_j\}$ werden auch Schwellen-Parameter (*treshold* oder *cutpoint*) genannt. Diese sind nichtfallend in j, da das kumulative Logit eine monton steigende Funktion von $F_j(x)$ ist. Und $F_j(x) = P(Y \leq j|x)$ ist für ein festes x eine monoton steigende Funktion in j.

Das Proportional Odds Model

Berücksichtigt man die Effekte der Erklärenden, ergibt sich folgendes Modell

$$L_j(x) = a_j + \beta' x \qquad j = 1, \ldots, J-1 \tag{48}$$

In diesem Modell ist ein gemeinsamer Vektor von Koeffizienten $\beta = (\beta_1, ..., \beta_I)$ für alle j enthalten. Eine interessante Eigenschaft dieses Modells ergibt sich aus dem Vergleich von zwei Ausprägungen der erklärenden Variable x. Setzt man die kumulativen Odds für beide Ausprägungen ins Verhältnis und bildet ein kumulatives *odds ratio*, erhält man:

$$L(x_1) - L(x_2) = \log\left[\frac{P\left(Y \leq j \mid x_1\right)/P\left(Y > j \mid x_1\right)}{P\left(Y \leq j \mid x_2\right)/P\left(Y > j \mid x_2\right)}\right] = \beta'\left(x_1 - x_2\right). \tag{49}$$

Der Logarithmus des kumulativen *odds ratios* ist also proportional zum Abstand der Ausprägungen der Erklärenden. Deswegen werden diese Modelle auch *proportional odds model* genannt (McCullagh (1980)). Das *odds ratio* hängt nicht mehr von den Ausprägungen von Y ab, sondern nur noch von $(x_1 - x_2)$. Diese Eigenschaft wird auch strikte stochastische Ordnung (*strict stochastical ordering*) genannt. Dies bedeutet einfach, dass wenn die kumulativen Logits für x_1 größer als für x_2 sind, dies für jede Kategorie der Response Y gilt. Oder anders interpretiert: das Odds, als Maß dafür, um wieviel wahrscheinlicher das Eintreten der Kategorien j und darunter gegenüber dem Eintreten der Kategorien $j+1$ und darüber ist, ist $\exp\left[\beta'\left(x_1 - x_2\right)\right]$ mal höher bei $x = x_1$ als bei $x = x_2$.

Es sei β_i der Parameter, der den Effekt der i-ten Ausprägung von x (also x_i) auf die kumulative Logit beschreibt. Ist $\beta_i > 0$ steigt jedes kumulative Logit an, wenn x_i größer wird. Damit steigt aber auch die kumulierte Wahrscheinlichkeit an. Da wir es hier mit kumulierten Wahrscheinlichkeiten zu tun haben, fällt aber auch ein großer Teil der Wahrscheinlichkeit auf die unteren Kategorien von Y. Das heißt, für hohe Werte von x tendiert Y zu eher niedrigeren Kategorien. Um dies zu umgehen, ersetzt man β durch $-\beta$ und erhält nun folgendes Modell:

$$L_j(x) = a_j - \beta' x \qquad j = 1, \ldots, J-1 \tag{50}$$

Ist $\beta_i > 0$ wird Y nun zu höheren Kategorien tendieren, für höhere Werte von x.

Es ist jedoch noch nicht geklärt, weshalb man für verschiedene j den gemeinsamen Effekt β annimt. Um dies zu zeigen, soll das kumulative Logit-Modell nun mit Hilfe des Schwellenwert-Ansatzes hergeleitet werden.

Der Schwellenwert-Ansatz

Beim Schwellenwert-Ansatz wird angenommen, dass die Kategorien der beobachteten Variablen Y Ausdruck einer unbeobachtbaren latenten Variablen oder einer stetigen Response Variablen Y^* ist. Zwischen der beobachteten Variablen $Y \in \{1, \ldots, J\}$ und Y^* gelte folgender Zusammenhang

$$Y = j \quad \Leftrightarrow \quad \alpha_{j-1} < Y^* \leq \alpha_j \qquad j = 1, ..., J \ ,$$

wobei $-\infty = \alpha_0 < \alpha_1 ... < \alpha_J = \infty$. Also beobachten wir für Y eine Response in Kategorie j, wenn Y^*in das j-te Intervall fällt. Die j Intervalle sind durch die Schwellwerte $\alpha_j, ..., \alpha_{J-1}$ bestimmt. Des Weiteren sei x ein Vektor von erklärenden Variablen. Dann kann Y^* damit wie folgt bestimmt werden:

$$Y^* = \beta' x + \epsilon \ , \tag{51}$$

wobei $\beta = (\beta_1, ..., \beta_I)$ ein Vektor von Koeffizienten ist und ϵ eine Zufallsvariable mit der Verteilungsfunktion G.

Die beobachtete Variable Y kann nun durch folgendes Modell bestimmt werden:

$$F_j(x) = P(Y \leq j \mid x) = P(Y^* \leq \alpha_j) = G(\alpha_j - \beta' x) \tag{52}$$

In dieser Form nennt man das Modell auch Schwellenwert-Modell. Wählt man als Verteilungsfunktion von ϵ die logistische Verteilungsfunktion, erhält man das kumulative Logit Modell:

$$P(Y \leq j \mid x) = \frac{\exp(\alpha_j - \beta' x)}{1 + \exp(\alpha_j - \beta' x)}, \qquad j = 1, ..., J-1. \tag{53}$$

Damit lassen sich folgende Formulierung für die Logits und die Odds ableiten lassen:

$$\log \left\{ \frac{P(Y \leq j \mid x)}{P(Y > j \mid x)} \right\} = \alpha_j - \beta' x \qquad j = 1, ..., J-1 \tag{54}$$

$$\frac{P(Y \leq j \mid x)}{P(Y > j \mid x)} = \frac{P(Y \leq j \mid x)}{1 - P(Y \leq j \mid x)} = \exp(\alpha_j - \beta' x) \qquad j = 1, ..., J-1. \tag{55}$$

Das hier dargestellte kumulative Logit-Modell nach dem Schwellenwert-Ansatz basiert auf der Annahme, dass die Erklärenden Variablen einen Einfluß auf die Veränderung in der latenten Skala haben, ohne dass der Einfluss der Erklärenden auf die Schwellenwerte $(\alpha_1, ..., \alpha_J)$ berücksichtigt wird. In einer erweiterten Version des Modells läßt sich jedoch auch die Abhängikeit der Schwellenwerte von Erklärenden modellieren. Dazu sei auf Fahrmeir & Tutz (1994) verwiesen.

Das Modell kann nun mit allen $J-1$ kumulativen Logits simultan geschätzt werden. Es sei betont, dass Modelle die auf der Annahme der proportionalen Odds beruhen, keine äquivalente Darstellung durch loglineare Modelle haben. Die Erklärenden Variablen können im kumulativen Logit-Modell sowohl metrisch als auch kategorial sein.

Es ist aber durchaus möglich, für ein *proportional odds model* ein *uniform association model* anzupassen, indem wieder die Scores u_i für eine ordinale Variable verwendet werden. Benutzt man dagegen ungordnete Parameter μ_i für nominale Erklärende, erhält man auch hier wieder ein *row effects model.* (Für Details siehe Agresti (1990)). In Agresti(1984) ist dagegen noch eine Version des *uniform association model* zu finden, in dem die Scores als Abweichungen vom Mittelwert gebildet werden, so dass sie

die Form $(u_i - \bar{u})$ annehmen. All diese Modelle können noch dahingehend erweitert werden, dass nicht nur eine Erklärende enthalten ist. Außerdem lassen sich mit diesen Modellen noch Interaktioneffekte modellieren.

3.4.3 Sequentielle Modelle oder *Continuation Ratio Models*

Variablen können auch in ordinalen Kategorien vorliegen, wenn die Ausprägungen voneinander abhängig sind. Die Abhängigkeit kann zum Beispiel darin bestehen, dass eine Kategorie nur zur Ausprägung gelangen kann, wenn zu einem vorhergehenden Zeitpunkt bereits die vorhergehenden Kategorien zur Ausprägung gelangt sind. Fahrmeier & Tutz (1994,2001) geben ein Bespiel das auf Holmes & Williams (1954) zurückgeht. Dabei wurde bei Kindern die Größe der Mandeln in drei Kategorien erfaßt. Diese Kategorien sind „vorhanden und nicht vergrößert", „vergrößert", „stark vergrößert". Es ist leicht einzusehen, dass ein Kind nur „stark vergrößerte" Mandeln haben kann, nachdem es bereites zu einem vorhergehenden Zeitpunkt bereits „vergrößerte" Mandeln hatte.

Für ordinale Variablen, bei denen die Rangfolge sich aus der zeitlichen Abfolge ihrer Ausprägungen ergibt, benutzt man sogenannte sequentielle Logit-Modelle. Auch hier kann wieder eine latente Variable als der ordinalen Variable zu grundeliegende Größe betrachtet werden. Die latente Variable soll wieder die Form $Y_j^* = -\beta' x + \epsilon$ haben. Die Darstellung soll hier vereinfachend mit $-\beta$ vorgenommen werden. ϵ folgt wieder der Verteilungsfunktion G. Der erste Schritt des Response-Mechanismus sei folgend bestimmt:

$$Y = 1 \qquad \Leftrightarrow \qquad Y_1^* \leq \alpha_1,$$

wobei α_1 wieder ein Schwellwert ist. Wenn $Y_1^* \leq \alpha_1$ stopt der Prozeß. In diesem Beispiel könnte Y_1^*eine latente Tendenz ausdrücken, dass die Mandeln eine normale Größe haben. Normale Mandeln $(Y = 1)$ sind vorhanden, solange keine vergrößerten Mandeln auftreten $(Y \geq 2)$. Der Prozeß setzt sich fort wenn $Y_1^* > \alpha_1$. Der nächste Schritt wäre:

$$Y = 2 \quad \text{gegeben} \quad Y \geq 2 \quad \Leftrightarrow \quad Y_2^* \leq \alpha_2 \quad \text{usw.}$$

Im Fall der Mandeln bei kleinen Kindern würde Y_2^* hierbei die unbeobachtete Tendenz

ausdrücken, dass die Mandeln weiter anwachsen, wenn die Mandeln schon vergrößert sind. Allgemein kann für Y auch geschrieben werden:

$$Y = j \quad \text{gegeben} \quad Y \geq j \quad \Leftrightarrow \quad Y_j^* \leq \theta_j$$

oder

$$Y > j \quad \text{gegeben} \quad Y \geq j \quad \Leftrightarrow \quad Y_j^* > \theta_j$$

$j = 1, ..., J - 1$. Also wird hier der Übergang von einer Kategorie j zu der Kategorie $j + 1$ dargestellt, gegeben, dass Kategorie j bereits eingenommen wurde. Das heißt, dass ein Übergang nur stattfindet, wenn die latente Variable über einem bestimmten Schwellenwert, der für die spezielle Kategorie charakteristisch ist, liegt. Im Gegensatz zu dem kumulativen Modell findet hier also ein bedingter Übergang statt. Bedingt bedeutet hier immer, dass die vorhergehende Kategorie bereits eingenommen wurde. Also gegeben eine Kategorie ist zur Ausprägung gelangt, muß entschieden werden, ob dieser Prozeß endet oder in eine höhere Kategorie übergeht. Ist Letzteres der Fall, würde dies die letztendliche Kategorie darstellen, die als einzigste auch beobachtbar ist.

Aus der Beschreibung des Mechanismus der durch die Modellierung abgebildet werden soll, kommt man somit zu einem sequentiellen Modell mit der Verteilungsfunktion G:

$$P(Y = j | Y \geq j, x) = G(\alpha_j + \beta' x), \tag{56}$$

$j = 1, ..., J$ mit $\alpha_k = \infty$. Die Wahrscheinlichkeiten seien mit:

$$P(Y = j | x) = G(\alpha_j + \beta' x) \prod_{i=1}^{j-1} \{1 - G(\alpha_j + \beta' x)\} \tag{57}$$

gegeben. $j = 1, ..., J$ und $\prod_{i=1}^{0} \{\bullet\} = 1$. Dieses Modell wird auch *continuation ratio model* genannt. Für die Parameter der Schwellenwerte sind im Gegensatz zum kumulativen Modell keine Anforderungen an die Ordnung nötig. Nehmen wir nun für G die logistische Verteilungsfunktion, dann erhalten wir das sequentielle Logit-Modell

$$P(Y = j | Y \geq j, x) = \frac{\exp(\alpha_j + \beta' x)}{1 + \exp(\alpha_j + \beta' x)}. \tag{58}$$

Für die hier verwendeten „continuation-ratio¨-Logits ergibt sich:

$$L_j = \log \left\{ \frac{P(Y = j|x)}{P(Y > j|x} \right\} = \alpha_j + \beta' x. \tag{59}$$

Für weitere Details sei auf Fahrmeir & Tutz (1994) verwiesen.

3.4.4 Zwei-Schritt (Two-Step) -Modelle

Nachfolgend sei noch auf sogenannte Two-Step- oder Zwei-Schritt-Modelle verwiesen. Bei vielen Untersuchungen, bei den ordinale Variablen auftreten, kann man dem ordinalen Charakter der Daten durch die Modellierung mit Hilfe des kumulativen und des sequentiellen Modells gerecht werden. Oft tritt bei diesen Daten aber auch die Möglichkeit auf, dass sich die Kategorien auf natürliche Weise in zwei Hälften teilen lassen.

Nehmen wir das Beispiel der Einschätzung der Wirtschaftslage. Die Befragten könnten angeben, ob sie die Lage „sehr gut”, „gut”, „teils/teils”, „schlecht” oder „sehr schlecht” einschätzen. Diese Kategorien kann man auch zu „postiv”, „neutral”, „negativ” zusammenfassen. Wobei die neue Kategorie „positiv” aus den alten Kategorien „sehr gut” und „gut” bestehen würde. Die Kategorie „neutral” wäre die alte Kategorie „teils/teils” und „negativ” würde sich aus „schlecht” und „sehr schlecht” ergeben. Man würde somit die alte Response-Skala nach der Kategorie 2 und der Kategorie 3 teilen. Als Resultat erhält man drei Mengen von realtiv homogenen Responses.

Bei Zwei-Schritt-Modellen faßt man in einem ersten Schritt die Kategorien zu relativ homogenen Gruppen (z.B. hier zu „positiv”) zusammen. Dabei sind die Gruppen selber heterogen („positiv”, „neutral” und „negativ”). In einem zweiten Schritt betrachet man dann die Responses innerhalb dieser homogenen Gruppen.

Es sei Y eine ordinale Response-Variable, die in die Kategorien $1, ..., J$ eingeteilt ist. Diese J Kategorien werden nun in T verschiedene Gruppen $S_1, ..., S_T$ zusammengefaßt. Wenn m_t die Nummer der letzten Kategorie einer Gruppe darstellt, sind die S_t wie folgt definiert $S_t = \{m_{t-1} + 1, ..., m_t\}$ mit $m_0 = 0$ und $m_T = J$.

Im ersten Schritt sollen die Gruppen zuerst mit einem kumulativen Modell geschätzt werden. Wenn man wieder annimmt, dass Y eine latente Variable $Y_0^* = -\beta_0' x + \epsilon$

zugrundeliegt, ergibt sich:

$$Y \in S_t \quad \Leftrightarrow \quad \alpha_{t-1} < Y_o^* \leq \alpha_t,$$

wobei ϵ wieder einer Verteilungsfunktion G folgt, die in diesem Fall wieder die logistische Verteilungsfuntkion sein soll. In einem zweiten Schritt folgt dann die Schätzung ebenfalls mit einem kumlativen Modell. Dabei ist zu beachten, dass die Wahrscheinlichkeiten des zweiten Schrittes bedingt sind auf die Ausprägung der Gruppen. Die zugrunde liegende latente Variable sei mit $Y_t^* = -\beta_t' x + \epsilon_t$ gegeben und ϵ_t folgt einer logistischen Verteilung. Der Response-Mechanismus sei wie folgt bestimmt:

$$Y = j | Y \in S_t \quad \Leftrightarrow \quad \alpha_{j-1,t} < Y_t^* \leq \alpha_{jt}.$$

Mit der zusätzlichen Annahme, dass die Noise-Variablen unabhängig sind, erhalten wir folgendes Modell:

$$\begin{aligned} P\left(Y \in V_t | x\right) &= G\left(\alpha_t + \beta_0' x\right), \\ P\left(Y \leq j | Y \in S_t, x\right) &= G\left(\alpha_{jt} + \beta_j' x\right), \end{aligned} \tag{60}$$

wobei

$$V_t = S_1 \bigcup ... \bigcup S_t, \alpha_1 < ... < \alpha_{T-1}, \alpha_T = \infty, \alpha_{t,m_{t-1}+1} < \quad ... < \alpha_{t,m_t-1}, \alpha_{t,m_t} = \infty, \quad t = 1, ..., T.$$

In beiden Schritten wurde hier das kumulative Modell angewandt. Es ist aber auch möglich im zweiten Schritt ein sequentielles Modell anzupassen. Man würde dann ein sogenanntes kumulativ-sequentielles Modell erhalten. Diese Art von Modellen wäre sinnvoll, wenn man zum Beispiel die Kategorien einer Response in drei Gruppen teilt. Die zweite Gruppe könnte dabei eine sogenannte Start-Gruppe bilden, die erste und dritte Gruppe wären jeweils Veränderungen in unterschiedliche Richtungen.

Die Zwei-Schritt-Modelle haben den Vorteil, dass bei jedem Schritt unterschiedliche Parameter einbezogen werden. So kann sich der Einfluß der erklärenden Variablen auf die Response-Variable in jedem Schritt verändern. Untersucht man z.B. nur den Einfluß auf die Response-Gruppe „positiv” so können die erklärenden Variablen einen ganz anderen Einfluß ausüben, als wenn man den Einfluß auf die Kategorien innerhalb der Gruppe untersucht. Für weitere Details sei auf Fahrmeir & Tutz (1994,2001)

verwiesen.

4 Das Schätzen und Testen von kumulativen Logit-Modellen

4.1 Die Schätzmethoden

Für das Testen von Logit-Modellen stehen zwei Methoden zur Verfügung. Zum einen die gewichtete Methode der kleinsten Quadrate (WLS) und zum anderen die Maximum-Likelihood-(ML)-Schätzung. Die WLS-Schätzung hat den Vorteil, dass sie in der Standardform zur Durchführung der Schätzung vorliegt und sich besonders eignet, wenn kein teures kommerzielles Softwarepaket zur Verfügung ist. In diesem Fall sollte es möglich sein, die WLS manuell zu programmieren.

4.2 Gewichtete Kleinste Quadrate - Schätzung (Weighted Least Squares - WLS)

Man habe eine Response-Variable, die c Kategorien annehmen kann. Diese resultieren aus s unabhängigen Beobachtungen einer einfachen Zufallstichprobe, mit jeweils $n_1$bis n_s Beobachtungen.

$$\pi' = (\pi_{(1)}, \ldots, \pi_{(s)})$$

$$\text{mit } \pi_{(i)} = (\pi_{1(i)}, \ldots, \pi_{c(i)})$$

Damit haben wir die bedingten Wahrscheinlichkeiten zur Ausprägung i der Erklärenden.

p sei die beobachtete relative Häufigkeit , die zu π korrespondiert. Die Kovarianz-Matrix von p ist die $(cs \times sc)$ - Matrix $V(\pi)$.

$$V(\pi) = \begin{pmatrix} V_1 & & 0 \\ & \ddots & \\ 0 & & V_s \end{pmatrix},$$

wobei

$$n_i V_i = \begin{pmatrix} \pi_{1(i)}(1-\pi_{1(i)}) & -\pi_{1(i)}\pi_{2(i)} & \dots & -\pi_{1(i)}\pi_{c(i)} \\ -\pi_{2(i)}\pi_{1(i)} & \pi_{2(i)}(1-\pi_{2(i)}) & \dots & -\pi_{2(i)}\pi_{c(i)} \\ \vdots & \vdots & \ddots & \vdots \\ -\pi_{c(i)}\pi_{c(i)} & -\pi_{c(i)}\pi_{2(i)} & \dots & \pi_{c(i)}(1-\pi_{c(i)}) \end{pmatrix}$$

Die Response-Funktion setzt sich aus den Funktionen der Wahrscheinlichkeiten

$$f_m(\pi),\ m = 1, \dots, n \leq s(c-1) \tag{61}$$

zusammen.

Es ist notwendig, die Annahme zu machen, dass eine stetige zweite Ableitung für die die Response-Funktionen

$$F(\pi)' = [f_1(\pi), ..., f_n(\pi)]$$

vorhanden ist. Die Hessesche-Matrix:

$$Q(\pi) = \left(\frac{\partial f_m(\pi)}{\partial \pi_{j(i)}}\right) \qquad m = 1, ..., n \tag{62}$$

für alle cs-Kombiantionen (i, i) darf aber nicht singulär sein. Dazu muß noch angenommen werden, dass die f_m linear unabhängig sind, so dass Q den Rang n hat. Die lineare Response-Funktion sei wie folgt:

$$F(\pi) = A\pi \text{ für eine Matrix A} \tag{63}$$

$$\text{und } Q(\pi) = A$$

Bei Logit-Modellen kann man von folgender Response-Funktion ausgehen:

$$F(\pi) = K \log(A\pi) \tag{64}$$

für Matrizen K und A. Im Logit Modell sei $Q(\pi)$:

$$Q(\pi) = KD^{-1}A, \tag{65}$$

wobei D eine Diagonal-Matrix mit den Elementen von $A\pi$ auf der Hauptdiagonalen ist. Es soll ein lineares Modell für die Funktion $F(\pi)$ geschätzt werden:

$$F(\pi) = X\beta$$

Die X-Matrix ist eine $(n \times v)$ Design-Matrix, von bekannten Konstanten mit Rang v. β sei der Vektor der Parameter.

Der WLS-Schätzer von β ergibt sich aus dem Vektor, der folgenden Term minimiert:

$$(F(p) - X\beta)'S^{-1}(F(p) - X\beta) \tag{66}$$

Der Schätzer b für β ist dann folgender:

$$b = \left(X'S^{-1}X\right)^{-1} X'S^{-1}F(p) \tag{67}$$

Die asymptotische Kovarianz-Matrix von b ist $(X'S^{-1}X)^{-1}$.

Soll nun das kumulative Logit-Modell als Zeilen-Effekt-Modell geschätzt werden, nehmen wir die Logits

$$L_{j(i)} = \alpha_j + \tau_i \qquad i = 1, 2, 3 \; j = 1, 2 \tag{68}$$

$$\text{mit } \sum \tau_i = 0$$

Die Responsefunktion $F(\pi)$ besteht hierbei aus sechs Logits:

$$F(\pi)' = \left(L_{1(1)},\, L_{2(1)},\, L_{1(2)},\, L_{2(2)},\, L_{1(3)},\, L_{2(3)}\right)$$

Für das Modell $F = X\beta$ mit $\beta' = (\alpha_1, \alpha_2, \tau_1, \tau_2)$ würde man folgende X - Matrix annehmen:

$$X = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & -1 & -1 \\ 0 & 1 & -1 & -1 \end{pmatrix}$$

Man beachte, dass $\tau_3 = -\tau_1 - \tau_2$.

Es seien $L_i = (L_{1(i)}, L_{2(i)})$ die zwei Logits, die in der Zeile i der 3x3 Tabelle gebildet wurden. Außerdem sei $\pi' = (\pi'_{(1)}, \pi'_{(2)}, \pi'_{(3)})$ wobei $\pi'_{(i)} = (\pi'_{1(i)}, \pi'_{2(i)}, \pi'_{3(i)})$. Damit kann L_i in Termen von $\pi_{(i)}$ ausgedrückt werden:

$$L_i = K^* \log\left(A^* \pi_{(i)}\right),$$

mit

$$A^* = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \text{und} \qquad K^* = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}.$$

Das gesamte Modell $F = X\beta$ kann nun folgendermaßen ausgedrückt werden:

$$K \log(A\pi) = X\beta$$

Dabei sind K und A die Kronecker-Produkte $K = I \otimes K^*$, $A = I \otimes A^*$und I ist die 3×3 Identitäts-Matrix. C sei eine $r_1 x c_1$-Matrix und die D eine $r_2 x c_2$-Matrix. Damit ergibt sich als Kronecker-Produkt $C \otimes D$ die $r_1 r_2 x c_1 c_2$-Matrix

$$C \otimes D = \begin{pmatrix} c_{11}D & c_{12}D & \dots \\ c_{21}D & c_{22}D & \dots \\ \vdots & \vdots & \end{pmatrix}$$

Zum Schätzen des Modells nehmen wir an, dass eine einfache Zufallsstichprobe, die einem multinomialen Stichprobenmodell folgt, vorliegt. Desweiteren braucht man noch die geschätzte Kovarianz-Matrix $V(p)$. Die geschätzte asymptotische Kovarianz-Matrix des Vektors der beobachten Logits $F(p)$ ist $S = QVQ'$ wobei $Q = KD^{-1}A$ ist. Und D ist die Diagonal-Matrix mit den Elementen der Vektors Ap auf der Hauptdiagonalen.

Die C-Matrix für den Unabhängigkeitstest (Nullhypothese $C\beta = 0$):

$$C = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Das kumulative Logit-Modell mit ordinaler Erklärender kann so ausgedrückt werden:

$$L_{j(i)} = a_j + \beta(u_i - \bar{u})$$

Es soll wieder $F = X\beta$ als das in Frage kommende Modell geschätzt werden. $F' = (L_{1(1)}, L_{2(1)}, L_{1(2)}, L_{2(2)}, L_{1(3)}, L_{1(3)}, L_{1(4)}, L_{2(4)})$ und $\beta' = (\alpha_1, \alpha_2, \beta)$. Für das *uniform association model* $(u_i = i)$ ist die X-Matrix:

$$X = \begin{pmatrix} 1 & 0 & -1,5 \\ 0 & 1 & -1,5 \\ 1 & 0 & -0,5 \\ 0 & 1 & -0,5 \\ 1 & 0 & 0,5 \\ 0 & 1 & 0,5 \\ 1 & 0 & 1,5 \\ 0 & 1 & 1,5 \end{pmatrix}$$

4.3 Das verallgemeinerte lineare Modell

Das Konzept der verallgemeinerten linearen Modell (*generalized linear models*) geht auf Nelder &Wedderburn (1972) zurück. Folgende Eigenschaften müssen für diese Modelle erfüllt sein:

1. Die Response-Variablen Y_i sind unabhängige Zufallsvariablen mit einer Verteilung aus der exponentiellen Familie. Die Verteilung muß für alle Y_i gleich sein und in der kanonischen Form vorliegen.

2. Die Daten werden durch ein lineares Modell beschrieben $\eta = X\beta$.

3. Der Erwartungswert der Response-Variablen μ_i ist durch eine Link-Funktion mit dem linearen Modell $\eta_i = g(\mu_i)$ verknüpft.

Es kann gezeigt werden, dass die Multinomialverteilung, die wir bei den kumulativen Logit-Modellen für die Response-Variable annehmen, eine Verteilung aus der exponentiellen Familie ist (vergleiche z.B. Rönz (2001)). Des Weiteren gehen wir immer davon aus, dass die Beobachtungen aus einer einfachen Zufallsstichprobe stammen und damit unabhängig sind. Wie bereits beschrieben, stellt bei kumulativen Logit-Modellen die Logit-Transformation die Verknüpfung des Erwartungswertes der Response mit dem linearen Modell dar.

Zur Schätzung von verallgemeinerten linearen Modellen kann die Maximum Likelihood Methode verwendet werden. McCullagh(1980) hat die Newton-Raphson-Methode für die ML-Schätzung auf Modelle angewandt, zu denen auch das kumulative Logit-Modell gehöhrt. Dies soll nun im nächsten Abschnitt dargestellt werden.

4.4 Die ML-Schätzung für kumulative Logit-Modelle

Man habe eine Zufallsstichprobe vom Umfang $(n_1, ..., n_s)$. Die Stichprobe folge einem Produkt-multinomialen Stichprobenmodell. Das heißt, es gibt s Kombinationen der Kategorien für die Response, die c Kategorien hat. Das Modell sei hier $F = X\beta$. Die F enthalten die $s(c-1)$ wahren Logits. Die X-Matrix ist eine Design-Matrix der Größe $s(c-1) \times v$. Die $\beta' s$ enthalten die Parameter $\beta = (\beta_1, ..., \beta_v)$.

Die Logits lassen sich auch mit Hilfe der Verteilungsfunktionen schreiben:

$$L_{j(i)} = \log\left[\left(1 - F_{j(i)}\right)/F_{j(i)}\right], \tag{69}$$

wobei $F_{j(i)} = \pi_{1(i)} + ... + \pi_{j(i)}$. Die kumulative Wahrscheinlichkeit sei gegeben durch

$$Z_{j(i)} = \frac{(n_{i1} + ... + n_{ij})}{n_i}$$

und X sei:

$$X = \begin{pmatrix} X^{(1)} \\ \vdots \\ X^{(s)} \end{pmatrix}.$$

Die $X^{(i)} = \left(x_{jk}^{(i)}\right)$ repräsentieren die $c-1$ Zeilen der Design-Matrix, die die Logits auf der Faktorstufe i enthalten. Aus McCullagh(1980) folgt die Ableitung der log-likelihood

$$\begin{aligned} q_k &= \frac{\partial \log L}{\partial \beta_k} \\ &= \sum_{i=1}^{s} n_i \left[\sum_{j=1}^{c-1} \frac{\left(Z_{j(i)}F_{j+1(i)} - Z_{j+1(i)}F_{j(i)}\right)}{F_{j(i)}\left(F_{j+1(i)} - F_{j(i)}\right)a_{ijk}}\right] \end{aligned} \tag{70}$$

$$a_{ijk} = F_{j(i)}(1 - F_{j(i)})x_{jk}^{(i)} - F_{j(i)}(1 - F_{j+1(i)})x_{j+1,k}^{(i)}$$

Der Erwartungswert der 2. Ableitung ist

$$\begin{aligned} h_{kl} &= E\left[\frac{\partial^2 \log L}{\partial\beta_k \partial\beta_l}\right] \\ &= -\sum_{i=1}^{s} u_i \left\{ \sum_{j=1}^{c-1} \left[\frac{F_{j+1(i)}}{F_{j(i)}\left(F_{j+1(i)} - F_{j(i)}\right)}\right] a_{ijk} a_{ijl} \right\} \end{aligned} \tag{71}$$

Diese Werte braucht man, um den Newton-Raphson Allgorithmus anzuwenden. Dieser stellt eine Methode da, mit der man die Lokation finden kann, bei der eine Funktion maximiert wird. Sie wird vorallem bei nichtlinearen Funktionen angewandt, bei denen einfache Methoden nicht mehr ausreichen um das Maximum einer Funktion zu ermitteln.

Bei der Anwendung der Newton-Raphson-Methode geht man von Anfangswerten aus, die mehr oder weniger geraten werden müssen. Die Funktion wird nun in der Umgebung des Anfangswertes mit Hilfe einer Taylor-Entwicklung zweiten Grades approximiert.

Soll nun der Parameter $\hat{\beta}$ eine Funktion $g(\beta)$ maximieren, bildet man die Taylor-Entwicklung um einen Anfangswert $\beta^{(t)}$ (wobei t den jeweiligen Iterationschritt angibt und im ersten Schritt $t = 0$ ist):

$$Q^{(t)}(\beta) = g\left(\beta^{(t)}\right) + q^{(t)'}\left(\beta - \beta^{(t)}\right) + \left(\frac{1}{2}\right)\left(\beta - \beta^{(t)}\right)' H^{(t)}\left(\beta - \beta^{(t)}\right) \tag{72}$$

q ist hier wie schon erwähnt, der Vektor der ersten Ableitungen und H ist die Hessesche-Matrix, die die zweiten Ableitungen enthält. Die Taylor-Entwicklung kann nun nach β abgeleitet und Null gesetzt werden. Man erhält folgende Gleichung:

$$\frac{\partial Q^{(t)}}{\partial \beta} = q^{(t)} + H^{(t)}\left(\beta - \beta^{(t)}\right) = 0$$

Den Wert, den man nun erhält, nimmt man als den Wert des nächsten Iterationschrittes. Somit ergibt sich für

$$\beta^{(t+1)} = \beta^{(t)} - \left(H^{(t)}\right)^{-1} q^{(t)}. \tag{73}$$

Um dieses $\beta^{(t+1)}$ wird nun im nächsten Schritt eine Taylor-Entwicklung approximiert. Eine Annahme ist, dass die Matrix $H^{(t)}$ nicht singulär ist. Die Grundidee hinter der Newton-Raphson-Methode ist, dass man sich durch die unterschiedlichen Vorzeichen

der Quotienten[7] der ersten und zweiten Ableitung, sich dem wahren Maximum weiter annähert. Sind die Veränderungen zwischen den Iterationschritten minimal genug (eigenen Vorgaben entsprechend) bricht man den Allgorithmus ab.

Nimmt nun die ersten und zweiten Ableitungen, wie sie oben bestimmt wurden, läßt sich das Newton-Raphson-Verfahren auf das Modell der kumulativen Logits anwenden. Die ersten und zweiten Ableitungen, werden dabei mit Hilfe der geschätzen kumulativen Wahrscheinlichkeiten im Iterationschritt t gebildet: $\left\{F_{j(i)}^{(t)}\right\}$. Die Schätzgleichung für das β läßt sich dann auch folgendermaßen schreiben:

$$\beta^{(t+1)} = -\left(H^{(t)}\right)^{-1} r^{(t)} \tag{74}$$

wobei

$$\begin{aligned} r_t \;=\; & \textstyle\sum_{i=1}^{s} n_i \left\{ \sum_{j=1}^{c-1} \left[F_{j+1(i)}/F_{j(i)}(F_{j+1(i)} - F_{j(i)}\right] a_{ijk} \right. \\ & \times \left[F_{j(i)}\left(1 - F_{j(i)}\right)\log\left(L_{j(i)}\right) - F_{j(i)}\left(1 - F_{j+1(i)}\right)\log\left(L_{j+1(i)}\right)\right. \\ & \left.\left. -Z_{j(i)} + F_{j(i)} Z_{j+1(i)}/F_{j+1}\right]\right\} \end{aligned} \tag{75}$$

Die Anfangswerte $\beta_1, ..., \beta_{c-1}$ sollten monoton fallend sein. Da die Log-Likelihood-Funktion konkav ist, sollte nach drei bis fünf Schritten eine gute Konvergenz erreicht sein.

4.5 Hypothesenüberprüfung in kumulativen Logit-Modellen

Hier soll die Modellüberprüfung nicht beschrieben werden. Dazu sei zum Beipiel auf Rönz (2001) verwiesen. Jedoch wird im nächsten Kapitel bei der Interpretation eines SPSS-Outputs darauf kurz eingegangen.

5 Anwendung auf die Perzeption der Wirtschaftslage

Rönz hat in einem Aufsatz[8] die Perzeption der Wirtschaftslage in der Bundesrepublik Deutschland zwischen 1991 und 1996 untersucht. Dort wurden für verschiedene

[7] exakterweise kann man hier nicht von dem Vorzeichen eines Quotienten sprechen, da hier die Inverse einer Matrix mit einem Verkor multipliziert wird. Man müßte eher die Definitheit derMatrix betrachten.

[8] siehe Rönz (1999)

ordinale Response-Variablen multinomiale Logit-Modelle geschätzt. Ausgehend von den dort vorgestellten Ergebnissen, soll hier der ordinale Charakter der Variablen in die Analyse mit einfließen und kumulative Logit-Modelle für die Response-Variablen geschätzt werden. Die Daten stammen aus dem kumulierten ALLBUS[9]. Untersucht wurden Daten aus den ALLBUS-Umfragen für 1991, 1992, 1994 und 1996. Befragt wurden in Privathaushalten lebende Personen mit einem Alter von mindestens 18 Jahren zum Zeitpunkt der Datenerhebung. Zusätzlich berücksichtigt die Analyse nur die Antworten von deutschen Staatsbürgern. Folgende Fragen wurden gestellt:

1. Wie beurteilen Sie ganz allgemein die heutige wirtschaftliche Lage in Deutschland ?
2. Wie beurteilen Sie Ihre heutige persönliche wirtschaftliche Lage ?
3. Wie wird die wirtschaftliche Lage in Deutschland in einem Jahr sein ?
4. Wie wird Ihre persönliche wirtschaftliche Lage in einem Jahr sein ?

Die Inhalte dieser Fragen wurden als Response-Variablen Y1-Y4 definiert. Die Beantwortung der Fragen konnte in folgenden Kategorien vorgenommen werden:

	für die Fragen 1 und 2		für die Fragen 3 und 4
1 -	sehr gut	1 -	wesentlich besser als heute
2 -	gut	2 -	etwas besser
3 -	teils gut/ teils schlecht	3 -	gleichbleibend
4 -	schlecht	4 -	etwas schlechter
5 -	sehr schlecht	5 -	wesentlich schlechter

Die Response-Variablen sind also in fünf Kategorien ausgeprägt. Tabelle 2 zeigt die relativen Häufigkeiten des Auftretens der Response-Kategorien.

Untersucht wurde der Einfluß folgender erklärenden Variablen, die in den ALLBUS-Umfragen enthalten sind:

X1 Erhebungsgebiet

[9] Alle inhaltlichen Ausführungen zum ALLBUS beziehen sich auf: „Allgemeine Bevölkerungsumfrage der Sozialwissenschaften, ALLBUS 1980-94", Codebuch, ZA-Nr 1795, Zentralarchiv für Empirische Sozialforschung an der Universität zu Köln, Zentrum für Umfragen, Methoden und Analysen Mannheim (ZUMA). Die vorgenannten Institutionen tragen keine Verantwortung für die Verwendung der Daten in dieser Arbeit.

1 - alte Bundesländer (West)
2 - neue Bundesländer (Ost)

X2 Geschlecht
1 - männlich
2- weiblich

X3 Haushaltsgröße
1 - 1 Personenhaushalt
2 - 2 Personenhaushalt
3 - 3 Personenhaushalt
4 - 4 und mehr Personenhaushalt

X4 Familienstand
1 - verheiratet
2 - getrennt, geschieden, verwitwet
3 - ledig

X5 Alter
1 - 18 bis 29 Jahre
2 - 30 bis 49 Jahre
3 - 50 und älter

X6 Nettoeinkommen
1 - 0 bis 800
2 - 800 bis 1500
3- 1500 bis 4000
4 - 4000 und mehr

Für weitere Details sei auf Rönz (1999) verwiesen. Dort ist auch ein grafische Analyse der Veränderung der einzelnen Response-Variablen über die Jahre enthalten. Zum Beispiel zeigt Abb. 2 eine deutliche Verschlechterung der Einschätzung der Wirtschaftlage 1996 gegenüber 1991 an.

5.1 Resultate der bivariaten kumulativen Logit-Modelle

Als ordinales Modell sollen hier kumulative Logits angepaßt werden. Wir gehen hier vom bivariaten Fall aus, d.h. daß die Response-Variablen nur von einer Erklärenden abhängen. Zum Schätzen von kumulativen Logit-Modellen kann zum Beispiel die Soft-

	1991	1992	1994	1996	1991	1992	1994	1996
Kategorien	Y1				Y2			
sehr gut	12,8	4,7	0,2	0,8	3,6	3,5	2,8	1,9
gut	46,1	40	11,4	11,8	47,9	49,4	51,9	47,6
teils teils	37,8	46,9	52	48,9	34,6	34,5	31,2	39,4
schlecht	2,6	7,4	30,5	31,9	11	9,5	11,3	8,6
sehr schlecht	0,7	1,1	5,9	6,6	2,9	3,1	2,9	2,5
Gültig	2965	3409	3264	3271	3005	3432	3284	3284
Fehlend	52	47	30	19	12	24	10	6
Gesamt	3017	3456	3294	3290	3017	3456	3294	3290
Kategorien	Y3				Y4			
wesentlich besser	9,6	2	0,7	0,3	3,5	2,6	1,6	0,8
etwas besser	42	19,1	27,6	10	29,9	21,3	17,8	13
gleichbleibend	38,5	43,6	43,1	39,3	55,2	60,7	67,2	66,1
etwas schlechter	9,1	30,8	25,2	43,2	9,9	13,5	12,1	17,9
wesentlich schlechter	0,8	4,4	3,3	7,3	1,4	1,9	1,3	2,2
Gültig	2907	3340	3212	3215	2904	3325	3218	3206
Fehlend	110	116	82	75	113	131	76	84
Gesamt	3017	3456	3294	3290	3017	3456	3294	3290

Tabelle 2: Häufigkeitsverteilungen der Response-Variablen; relative Häufigkeiten in %.

ware SPSS verwendet werden. Diese bietet seit der Version 10 die Möglichkeit ordinale Regressionsmodelle mit Hilfe der Prozedur „Polytomous Logit Universal Models" (PLUM) zu schätzen[10]. Als Alternative könnte auch GOLDMineR (Graphical Ordinal Logit Display on Monotonic Regression) verwendet werden. Jedoch werden hier verallgemeinerte Logit-Modelle als *monotonic regression models* geschätzt, die in dieser Arbeit nicht beschrieben wurden[11].

Bevor wir zu den Resultaten kommen, soll hier der SPSS-Output der Schätzung des bivariaten Modells mit X1=Erhebungsgebiet als Erklärender für 1991 interpretiert werden. Interpretiert wurden nur die Ergebnisse für X1 und X2. Für die restlichen Variablen werden nur die Ergebnisse der Schätzung präsentiert.

Das kumulative Logit-Modell, dass hier geschätzt werden soll, ist folgendes:

$$L_j = a_j - \beta_1 x_1$$

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	121,077			
Final	61,896	59,181	1	,000

Link function: Logit.

Der Chi-Square-Wert beruht auf der Differenz der Deviance eines Modells, das nur

[10] siehe Bühl & Zöfel (2000)

[11] siehe Magidson (1998); das Programm bietet die Möglichkeit die Logits gegen die Erklärende zu plotten, so wie sie in Magidson (1996) zu sehen sind

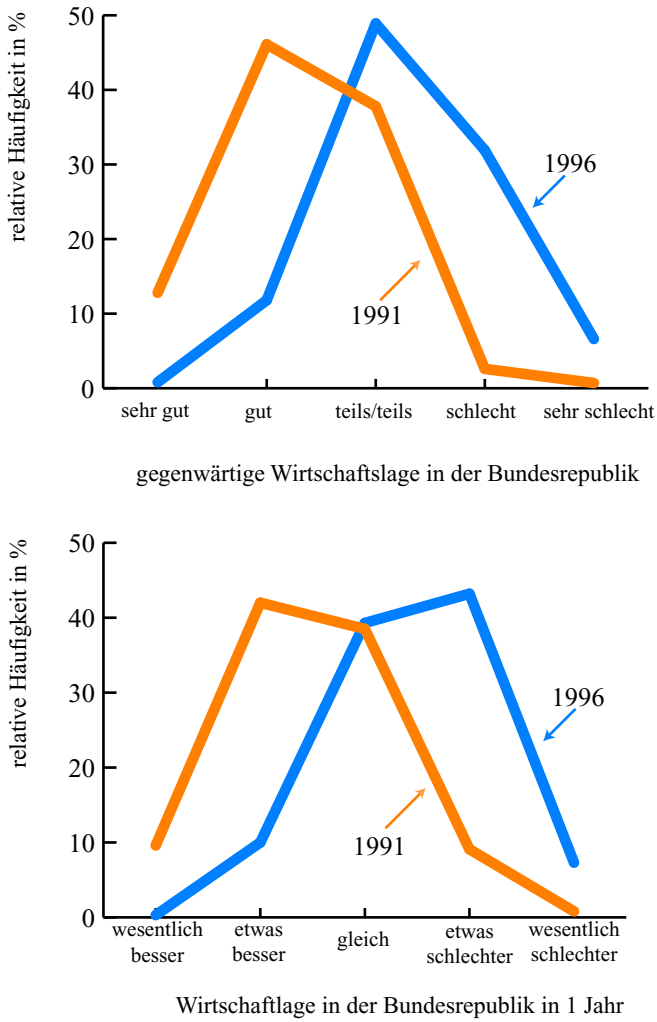

Abbildung 2: Einschätzung der Wirtschaftslage in der BRD

die α_j enthält und der Deviance eines Modells bei dem der Parameter β_1 vorhanden sind. Es zeigt damit an, ob sich durch die Aufnahme des Parameters eine signifikante Verbesserung der Modellinformation ergibt. Hier zeigt sich eine signifikante Verbesserung (p<0,001) des Modells durch den Einschluss des Faktors.

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	13,961	3	,003
Deviance	13,899	3	,003

Link function: Logit.

Jedoch ist die Anpassungsgüte in diesem Fall nicht besonders gut. Der Chi-Quadrat-Test nach Pearson zeigt hier einen signifikanten Wert an. Was bedeutet, dass sich die

beobachteten Zellhäufigkeiten signifikant von den erwarteten Häufigkeiten unterscheiden. Zu ähnlichen Aussagen gelangt man, wenn man sich die Bestimmtheitsmaße anschaut.

Pseudo R-Square

Cox and Snell	,020
Nagelkerke	,022
McFadden	,009

Link function: Logit.

Es ist zu erkennen, dass durch die ordinale Regression nur ein sehr geringer Anteil der Variation der Response-Variable erklärt wird.

Parameter Estimates

			Estimate	Std. Error	Wald	df	Sig.
Threshold	sehr gut	[Y1_91 = 1]	-2,212	,068	1059,380	1	,000
	gut	[Y1_91 = 2]	9,853E-02	,050	3,851	1	,050
	teils teils	[Y1_91 = 3]	3,135	,106	871,681	1	,000
	schlecht	[Y1_91 = 4]	4,713	,221	455,741	1	,000
Location	West	[X1_91=1]	-,536	,070	58,626	1	,000
	Ost	[X1_91=2]	0(a)	,	,	0	,

Link function: Logit.

a This parameter is set to zero because it is redundant.

Für jede Kategorie der abhängigen Variable ist hier ein Parameter angegeben. Wobei der Parameter der letzten Kategorie redundant ist, da dieser die Referenzkategorie ist. Den Einfluß der erklärenden Variablen, geben alleine die Lageschätzer (Location) wieder. Es zeigt sich, dass X1 einen signifikanten Einfluss auf die Response-Variable hat. Auch die Schwellenschätzer sind signifikant. Dies kann man dem Wald-Test entnehmen, der zum 5% Niveau durchgeführt wurde.

Der negative Lageschätzer für „West" gibt an, dass diese Kategorie im Sinne einer niedrigeren Kategorie der Response-Variable wirkt. Dass bedeutet, dass wenn der Befragte aus den alten Bundesländern kam, er eher die Wirtschaftslage positv eingeschätzt hat. Das Odds des Eintretens der unteren Kategorien von Y1 gegenüber dem Eintreten der höheren Kategorien von Y1, ist

$$(\exp\left[-0,536(x_{west} - x_{ost})\right] = \exp\left[0,536\right] \approx 1,7)$$

mal höher bei Befragten aus den alten Bundesländern als bei den Befragten aus den neuen Bundesländern. Damit war es 1,7 mal wahrscheinlicher eine eher positivere Einschätzung der Wirtschaftlage zu erhalten, wenn der Befragte aus dem Westen der Bundesrepublik kam, als wenn der Befragte aus dem Osten der Bundesrepublik stammte. Dieses Ergebnis gilt unabhängig der Ausprägung der Response-Variable.

Zur Verdeutlichung der Eigenschaft der *proportional odds* lassen sich aus den Parameterwerten die geschätzten kumulierten Wahrscheinlichkeiten für die Kategorien der Response-Variable berechnen (In SPSS besteht die Möglichkeit sich die geschätzten Zellwahrscheinlichkeiten ausgeben zu lassen). Für einen Befragten aus den alten Bundesländern ergibt sich die Wahrscheinlichkeit die Wirtschaftlage „gut" und besser einzuschätzen:

$$L_{gut} = \alpha_{gut} - \beta_{west} = 0,099 + 0,536 = 0,635$$

$$\exp(L_{gut}) = \exp(0,636) = 1,887$$

$$\pi_{gut} = \frac{1,887}{2,887} = 0,65$$

Die folgende Tabelle beruht auf dem SPSS-Output. Man sieht hier leichte Rundungsabweichungen.

		Erhebungsgebiet 1991			
		West		Ost	
		geschätzte Zellwahrscheinlichkeit	geschätzte kum. Zellwahrscheinlichkeit	geschätzte Zellwahrscheinlichkeit	geschätzte kum. Zellwahrscheinlichkeit
Wirtschaftslage in Dtl.	sehr gut	0,16	0,16	0,10	0,10
	gut	0,50	0,66	0,43	0,53
	teils teils	0,32	0,98	0,43	0,96
	schlecht	0,02	1,00	0,03	0,99
	sehr schlecht	0,01	1,01	0,01	1,00

Die kumulierte Wahrscheinlichkeit von 0,66 gibt an, dass ein Befragter aus dem Westen die Wirtschaftslage mit 66 prozentiger Wahrscheinlichkeit „gut" oder „sehr gut" einschätzt. Bildet man ein Odds und setzt es zum selben Odds bei der Katgorie Ost ins Verhältnis

$$\frac{0,66/0,35}{0,53/0,47} \approx 1,7 \ ,$$

erhält man den konstanten Proportionalitätsfaktor der Odds für alle Schwellenwerte.

In den folgenden Tabellen sind die Ergebnisse der Schätzuungen enthalten. Für alle Tabellen gilt, dass fettgedruckte Werte eine Signifikanz zum 5% Niveau anzeigen. Unterstrichene Werte sind zum 10% Niveau signifikant.

	1991	1992	1994	1996	1991	1992	1994	1996
			Y1				Y2	
$\hat{\alpha}_1$	**-2,212**	**-3,017**	**-5,83**	**-4,820**	**-4,166**	**-4,145**	**-3,846**	**-4,239**
$\hat{\alpha}_2$	**0,099**	**-0,216**	**-1,844**	**-1,928**	**-0,590**	**-0,610**	-0,084	**-0,320**
$\hat{\alpha}_3$	**3,135**	**2,379**	**0,75**	**0,475**	**1,326**	**1,313**	**1,543**	**1,803**
$\hat{\alpha}_4$	**4,713**	**4,512**	**2,975**	2,651	**3,039**	**2,841**	**3,271**	**3,387**
$\hat{\beta}_1$	**-0,536**	0,000	**0,283**	0,010	**-1,357**	**-1,090**	**-0,407**	**-0,454**
			Y3				Y4	
$\hat{\alpha}_1$	**-1,797**	**-3,063**	**-4,641**	**-5,728**	**-3,028**	**-3,08**	**-3,864**	**-4,849**
$\hat{\alpha}_2$	**0,648**	**-0,36**	**-0,673**	**-2,018**	**-0,359**	**-0,543**	**-1,184**	**-1,876**
$\hat{\alpha}_3$	**2,930**	**1,864**	**1,189**	**0,136**	**2,445**	**2,443**	**2,124**	**1,338**
$\hat{\alpha}_4$	**5,624**	**4,459**	**3,65**	**2,700**	**4,623**	**4,678**	**4,605**	**3,775**
$\hat{\beta}_1$	**1,187**	**1,707**	**0,398**	**0,233**	**0,649**	**0,971**	**0,361**	-0,064

Tabelle 4: Schätzwerte für X1 als Erklärende

Erhebungsgebiet (X1)

Das Ergebnis zeigt, dass die Befragten aus den alten Bundesländern die Wirtschaftlage (sowohl die Deutschlands, als auch ihre persönliche) tendenziell positiv einschätzen, wobei für Y1 nur die Hälfte der Parameter signifikant sind. Dem kann entgegengehalten werden, dass sie über die Jahre hinweg immer eine Verschlechterung der Wirtschaftlage erwartet haben. Sinkende Parameterwerte über die Jahre zeigen an, dass sich die Unterschiede zwischen Ost und West in der Einschätzung der Wirtschaftlage über die Jahre verringert haben.

Geschlecht (X2)

	1991	1992	1994	1996	1991	1992	1994	1996
			Y1				Y2	
$\hat{\alpha}_1$	**-2,139**	**-3,165**	**-5,976**	**-4,855**	**-3,331**	**-3,383**	**-3,578**	**-3,947**
$\hat{\alpha}_2$	**0,159**	**-0,353**	**-1,990**	**-1,963**	0,026	0,045	**0,167**	-0,048
$\hat{\alpha}_3$	**3,183**	**2,252**	**0,593**	**0,441**	**1,792**	**1,867**	**1,782**	**2,057**
$\hat{\alpha}_4$	**4,762**	**4,386**	**2,813**	**2,616**	**3,470**	**3,368**	**3,504**	**3,637**
$\hat{\beta}_1$	**-0,424**	**-0,292**	0,067	-0,058	-0,067	**-0,152**	-0,041	-0,058
			Y3				Y4	
$\hat{\alpha}_1$	**-2,318**	**-3,921**	**-5,038**	**-5,927**	**-3,394**	**-3,713**	**-4,116**	**-4,916**
$\hat{\alpha}_2$	-0,011	**-1,362**	**-1,070**	**-2,218**	**-0,775**	**-1,242**	**-1,445**	**-1,939**
$\hat{\alpha}_3$	**2,138**	**0,563**	**0,783**	-0,068	**1,974**	**1,625**	**1,843**	**1,282**
$\hat{\alpha}_4$	**4,803**	**3,033**	**3,236**	**2,491**	**4,163**	**3,853**	**4,324**	**3,718**
$\hat{\beta}_1$	**-0,159**	-0,098	**-0,278**	-0,102	**-0,181**	**-0,173**	-0,053	**-0,213**

Tabelle 5: Schätzwerte für X2 als Erklärende

Beim Geschlecht als der erklärenden Größe, zeigt sich, dass Männer geneigt sind, die Wirtschaftlage eher postiv einzuschätzen. Außerdem haben sie auch eher positivere Erwartungen für die künftige Entwicklung der wirtschaftlichen Situation. Sowohl bei der Wirtschaftslage Deutschlands, als auch bei der persönlichen wirtschaftlichen Lage sind die Parameter negativ. Es sind aber nur knapp die Hälfte der Parameter signifikant. Die Unterschiede zwischen Männer und Frauen sind relativ konstant über die Jahre.

Haushaltsgröße (X3)

	1991	1992	1994	1996	1991	1992	1994	1996
			Y1				Y2	
$\hat{\alpha}_1$	**-1,920**	**-2,985**	**-5,964**	**-4,676**	**-3,296**	**-3,317**	**-3,548**	**-4,016**
$\hat{\alpha}_2$	**0,357**	**-0,181**	**-1,978**	**-1,783**	0,065	0,110	**0,204**	-0,110
$\hat{\alpha}_3$	**3,364**	**2,418**	**0,609**	**0,624**	**1,833**	**1,933**	**1,824**	**2,002**
$\hat{\alpha}_4$	**4,941**	**4,551**	**2,828**	**2,803**	**3,514**	**3,436**	**3,550**	**3,583**
$\hat{\beta}_1$	0,028	0,179	0,040	**0,215**	0,171	0,154	0,182	0,078
$\hat{\beta}_2$	-0,050	0,018	0,004	**0,228**	-0,119	-0,091	-0,133	**-0,257**
$\hat{\beta}_3$	0,041	-0,042	0,177	0,171	0,069	-0,029	0,152	-0,039
			Y3				Y4	
$\hat{\alpha}_1$	**-2,284**	**-3,798**	**-4,951**	**-5,863**	**-3,120**	**-3,359**	**-4,045**	**-4,692**
$\hat{\alpha}_2$	0,023	**-1,238**	**-0,988**	**-2,156**	**-0,490**	**-0,884**	**-1,373**	**-1,715**
$\hat{\alpha}_3$	**2,172**	**0,688**	**0,860**	-0,006	**2,278**	**1,997**	**1,917**	**1,505**
$\hat{\alpha}_4$	**4,837**	**3,160**	**3,312**	**2,553**	**4,468**	**4,228**	**4,397**	**3,942**
$\hat{\beta}_1$	-0,017	**0,223**	-0,093	-0,029	**0,272**	**0,420**	0,041	0,057
$\hat{\beta}_2$	0,020	0,031	-0,147	0,012	**0,454**	**0,410**	0,115	**0,244**
$\hat{\beta}_3$	**-0,209**	0,094	0,031	0,060	0,006	**0,242**	-0,006	0,101

Tabelle 6: Schätzwerte für X3 als Erklärende

Familienstand (X4)

	1991	1992	1994	1996	1991	1992	1994	1996
			Y1				Y2	
$\hat{\alpha}_1$	**-1,946**	**-3,073**	**-5,948**	**-4,690**	**-3,452**	**-3,483**	**-3,561**	**-4,152**
$\hat{\alpha}_2$	**0,333**	**-0,271**	**-1,963**	**-1,797**	-0,087	-0,046	**0,192**	**-0,236**
$\hat{\alpha}_3$	**3,342**	**2,325**	**0,623**	**0,610**	**1,686**	**1,785**	**1,815**	**1,887**
$\hat{\alpha}_4$	**4,920**	**4,458**	**2,842**	**2,788**	**3,368**	**3,291**	**3,543**	**3,473**
$\hat{\beta}_1$	-0,081	-0,076	0,043	0,140	**-0,253**	**-0,296**	-0,081	**-0,368**
$\hat{\beta}_2$	0,149	-0,040	**0,219**	0,322	0,113	0,125	**0,363**	0,139
			Y3				Y4	
$\hat{\alpha}_1$	**-2,343**	**-3,886**	**-4,911**	**-5,712**	**-2,951**	**-3,231**	**-3,548**	**-4,201**
$\hat{\alpha}_2$	-0,038	**-1,327**	**-0,952**	**-2,004**	**-0,323**	**-0,747**	**-0,854**	**-1,196**
$\hat{\alpha}_3$	**2,109**	**0,596**	**0,895**	**0,148**	**2,437**	**2,148**	**2,489**	**2,090**
$\hat{\alpha}_4$	**4,775**	**3,066**	**3,346**	**2,711**	**4,628**	**4,380**	**4,977**	**4,532**
$\hat{\beta}_1$	-0,101	-0,023	-0,037	**0,227**	**0,437**	**0,473**	**0,708**	**0,854**
$\hat{\beta}_2$	-0,221	0,010	0,011	0,143	**0,500**	**0,732**	**0,856**	**0,748**

Tabelle 7: Schätzwerte für X4 als Erklärende

Alter (X5)

	1991	1992	1994	1996	1991	1992	1994	1996
			Y1				Y2	
$\hat{\alpha}_1$	**-1,873**	**-2,991**	**-5,973**	**-4,840**	**-3,201**	**-3,208**	**-3,548**	**-3,736**
$\hat{\alpha}_2$	**0,406**	**-0,189**	**-1,987**	**-1,947**	**0,160**	**0,223**	**0,196**	**0,178**
$\hat{\alpha}_3$	**3,415**	**2,407**	**0,597**	**0,462**	**1,927**	**2,048**	**1,812**	**2,301**
$\hat{\alpha}_4$	**4,993**	**4,540**	**2,816**	**2,650**	**3,607**	**3,550**	**3,534**	**3,885**
$\hat{\beta}_1$	**0,201**	0,079	0,004	**-0,195**	**0,249**	**0,290**	0,007	**0,356**
$\hat{\beta}_2$	0,012	0,032	0,098	0,064	0,126	0,134	0,022	**0,342**
			Y3				Y4	
$\hat{\alpha}_1$	**-2,208**	**-3,846**	**-4,816**	**-5,878**	**-3,862**	**-4,178**	**-4,586**	**-5,190**
$\hat{\alpha}_2$	0,096	**-1,287**	**-0,855**	**-2,169**	**-1,188**	**-1,661**	**-1,866**	**-2,171**
$\hat{\alpha}_3$	**2,242**	**0,637**	**0,992**	-0,019	**1,651**	**1,311**	**1,537**	**1,140**
$\hat{\alpha}_4$	**4,907**	**3,107**	**3,444**	**2,549**	**3,848**	**3,555**	**4,024**	**3,581**
$\hat{\beta}_1$	0,047	0,063	0,103	-0,133	**-1,046**	**-1,113**	**-1,136**	**-1,101**
$\hat{\beta}_2$	0,052	0,043	**0,146**	0,061	**-0,667**	**-0,653**	**-0,493**	**-0,266**

Tabelle 8: Schätzwerte für X5 als Erklärende

Nettoeinkommen (X6)

	1991	1992	1994	1996	1991	1992	1994	1996
		Y1				Y2		
$\hat{\alpha}_1$	**-1,770**	**-3,065**	**-5,743**	**-4,446**	**-3,110**	**-2,923**	**-3,044**	**-2,650**
$\hat{\alpha}_2$	**0,522**	**-0,259**	**-1,755**	**-1,525**	**0,396**	**0,582**	**0,783**	**1,496**
$\hat{\alpha}_3$	**3,542**	**2,341**	**0,838**	**0,820**	**2,252**	**2,471**	**2,455**	**3,546**
$\hat{\alpha}_4$	**5,120**	**4,475**	**3,062**	**2,925**	**3,954**	**3,995**	**4,201**	**5,099**
$\hat{\beta}_1$	0,336	0,131	**0,472**	0,282	**0,766**	**0,816**	**1,057**	**1,591**
$\hat{\beta}_2$	**0,310**	-0,049	**0,292**	**0,396**	**0,807**	**1,091**	**1,145**	**2,088**
$\hat{\beta}_3$	-0,073	-0,152	**0,306**	0,344	**-0,418**	0,176	**0,406**	**1,427**
		Y3				Y4		
$\hat{\alpha}_1$	**-2,406**	**-4,158**	**-4,706**	**-5,873**	**-3,480**	**-3,935**	**-4,224**	**-4,569**
$\hat{\alpha}_2$	-0,079	**-1,583**	**-0,742**	**-2,273**	**-0,857**	**-1,454**	**-1,552**	**-1,460**
$\hat{\alpha}_3$	**2,088**	**0,368**	**1,109**	-0,032	**1,895**	**1,433**	**1,740**	**1,686**
$\hat{\alpha}_4$	**4,758**	**2,852**	**3,562**	**2,435**	**4,081**	**3,660**	**4,221**	**4,075**
$\hat{\beta}_1$	-0,138	**-0,286**	**0,368**	-0,107	-0,200	**-0,387**	-0,197	0,137
$\hat{\beta}_2$	**-0,453**	**-0,685**	0,175	-0,090	**-0,309**	**-0,548**	-0,112	**0,558**
$\hat{\beta}_3$	0,123	-0,116	**0,202**	-0,076	**-0,085**	**-0,250**	-0,168	0,336

Tabelle 9: Schätzwerte für X6 als Erklärende

6 Anhang

6.1 Stichprobenmodelle und -verteilungen

Es seien n_i (mit $i = 1, ..., N$) die beobachteten Zellhäufigkeiten der N Zellen einer $I \times J$-Kontigenztabelle. Für N gelte $N = IJ$. Die Zellhäufigkeiten n_i seien außerdem Zufallsvariablen. Dann sind die erwarteten Zellhäufigkeiten $m_i = E(n_i)$.

Das Poisson Stichprobenmodell

Beim Poisson-Stichprobenmodell nimmt man an, dass die einzelnen Zellhäufigkeiten n_i unabhängige Zufallsvariablen sind, die einer Poisson-Verteilung folgen. Die Verteilungsfunktion der Poisson-Verteilung ist folgende

$$\frac{\exp(-m_i)\, m_i^{n_i}}{n_i!} \qquad \text{für } n_i = 0, 1, 2, ... \tag{76}$$

Der Erwartungswert und die Varianz sind gleich: $E[n_i] = Var[n_i] = m_i$. Die Verteilung hängt nur von einem Parameter m_i ab. Die gemeinsame Verteilung aller $\{n_i\}$ ist das Produkt der Wahrscheinlichkeiten (76) für alle N Zellen. Damit ist der Gesamtstichprobenumfang $n = \sum n_i$ auch eine Zufallsvariable, die einer Poisson-Verteilung folgt. Das Poisson-Stichprobenmodell wird immer dann angewandt, wenn die Anzahl der Beobachtungen nicht von vornherein festgelegt werden soll. Zum Beispiel, wenn man das Auftreten eines Ereignisses über eine Zeitspanne beobachten möchte.

Das Multinomiale Stichprobenmodell

Beim Multinomialen Stichprobenmodell ist der Gesamtstichprobenumfang n von vornherein festgelegt. Damit sind die einzelnen Zellhäufigkeiten keine Zufallsvariablen mehr. Da sie niemals größer n sein können, sind sie damit bedingt auf den Gesamtstichprobenumfang $n = \sum n_j$. Die bedingte Wahrscheinlichkeit, n_i Beobachtungen in der Zelle i (für alle $i = 1, ..., N$ gleichzeitig) zu erhalten, ist dann wie folgt bestimmt:

$$\frac{\prod_i \left[\exp(-m_i)\, m_i^{n_i}/n_i!\right]}{\exp\left(-\sum m_j\right)\left(\sum m_j\right)^n/n!} = \left(\frac{n_i!}{\prod_i n_i!}\right) \prod_i \pi_i^{n_i}.$$

Dies ist die Multinomialverteilung mit den Parametern n für den Gesamtstichprobenumfang und π_i für die Zellwahrscheinlichkeiten. Dabei ist $\pi_i = m_i/(\sum m_j)$ und kann durch die Maximum-Likelihood Schätzer $p_i = n_i/n$ geschätzt werden.

Als Sonderfall ergibt sich die Binomialverteilung

$$\begin{pmatrix} n \\ n_i \end{pmatrix} \pi_i^{n_i} (1 - \pi_i)^{n-n_i},$$

in der die Wahrscheinlichkeiten π_i die Erfolgswahrscheinlichkeit für ein binäres Ereignis darstellt. Der Erwartungwert von n_i ist $E[n_i] = n\pi_i$ und die Varianz $Var[n_i] = n\pi_i(1 - \pi_i)$.

Das Produkt-multinomiale Stichprobenmodell

Bei einer Stichprobe soll eine Variable Y eine kategoriale Response mit j Kategorien sein, die durch eine Erklärende X erklärt werden soll. Oder man betrachtet den symmetrischen Fall für Y und X, wobei eine Variable X als Faktor angesehen wird. X habe i Kategorien. Dann läßt sich die Grundgesamtheit in i verschiedene sich gegenseitig ausschließende Segmente teilen, die durch die Kategorien von X bestimmt sind. Nimmt man aus jedem Segment unabhängig voneinander eine Stichprobe vom festen Umfang n_{i+}und beobachtet die Ausprägung von Y, erhält man folgende bedingte Wahrscheinlichkeiten $\{\pi_{1|i}, ..., \pi_{J|i}\}$ für das Eintreten von Y gegeben die i-te Ausprägung von X. Die Zellhäufigkeiten n_{ij} haben dann, folgende Multinomiale Verteilung:

$$\left(\frac{n_{i+}!}{\prod_j n_{ij}!}\right) \prod_j \pi_{j|i}^{n_{ij}}$$

Die gemeinsame Verteilung aller n_{ij} ergibt sich als Produkt der Wahrscheinlichkeiten für alle Segmente, daher auch die Bezeichnung Produkt-Multinomial.

Literatur

[1] AGRESTI, A. (1984). *Analysis of ordinal categorial data*. John Wiley & Sons New York.

[2] AGRESTI, A. (1990). *Categorial data analysis.* John Wiley & Sons New York.

[3] ANDERSON, J.A. AND PHILLIPS R.R. (1981). Regression, discrimination and measurment models for orderd categorial variables. *Applied Statistics* **30**: 22-31.

[4] BÜHL,A. , ZÖFEL,P. (2000). *SPSS Version 10. Einführung in die moderne Datenanalyse unter Windows*. Addison-Wesley. München

[5] EDWARDS, A.L. , THURSTONE, L.L. (1952). An Internal Consistency Check for Scale Values Determined by the Method of Successive Intervals". *Psychometrika*, **82**: 205-213.

[6] FAHRMEIR, L. , TUTZ, G. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer-Verlag New York.

[7] GOODMANN, L. A. (1979). Simple models for the Analysis of Association in cross-classifications having ordered categories. *J. Amer. Statist. Assoc.* **74**: 537-552.

[8] GOODMANN, L. A. (1983). The Analysis of dependence in cross-classifications having ordered categories, using log-linear models for frequencies and log-linear models for odds. *Biometrics* **39**: 149-160.

[9] MAGIDSON, J. (1996). Maximum Likelihood assessment of clinical trials based on an ordered categorial reponse. *Drug Information Journal.* Vol.**30**: 143-170.

[10] MAGIDSON, J. (1998). *GOLDMineR 2.0 User's Guide*. Statistical Innovations Inc. Belmont

[11] MCCULLAGH P. (1980). Regression Model for Ordinal Data (with discussion). *Journal of the Royal Statistical Society*, **B 42**: 109-127.

[12] NELDER, J. A. , WEDDERBURN, W. R. M (1972). Generalized Linear Models. *Journal of the Royal Satistical Society*, **A 135**: 370-284.

[13] 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.

[14] RÖNZ, B. (2001). *Generalisierte Lineare Modelle*. Vorlesungs-Script. Humboldt-Universität zu Berlin, Wirtschaftswissenschaftliche Fakultät, Institut für Statistik und Ökonometrie.

[15] RÖNZ, B. , STROHE, H. G. (Hrsg.) (1994). *Lexikon Statistik*. Gabler Wiesbaden.

[16] STIRLING, W. D. (1984). Fitting linear models to ordinal responses. *British Journal of Math. and Stat. Psychology* **37**: 263-270.

[17] TERZA, J. V. (1985): Ordinal Probit: a generalisation. *Commun. Statist.- Theor. Meth.* **14**: 1-11.

[18] TUTZ, G. (1990). M*odelle für kategoriale Daten mit ordinalen Skalenniveau: parametrische und nichtparametrische Ansätze*, Vandenhoeck & Ruprecht Verlag Göttingen.

[19] WILLIAMS, O.D. AND GRIZZLE, J.E. (1972). Analysis of contigency tables having ordered response categories. *J. Amer. Statist. Assoc.* **67**: 55-63.

Erklärung zur Urheberschaft

Hiermit erkläre ich, dass ich die vorliegende Arbeit allein und nur unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe.

Andre Kühnlenz

Berlin, 6. Februar 2002