Statistische Modellierung III

-Kategoriale Regression-

Dr. Martin Scharpenberg

MSc Medical Biometry/Biostatistics

WiSe 2019/2020

Setup

- Betrachten nun Daten mit einer kategorialen (nominal oder ordinal) Zielvariable mit mehr als 2 Kategorien
- Werden die so erhaltenen Daten auf eine Multinomialverteilung zurückführen
- Dies erlaubt uns dann die bereits bekannte Theorie der GLM anzuwenden
- Werden bei der Modellwahl zwischen nominalskalierten und ordinalskalierten Merkmalen unterscheiden

Datenbeispiele

Infektionen nach Kaiserschnittgeburten

- Dieses Beispiel bereits bei binärer Regression betrachtet
- Es gab jedoch 2 Arten von Infektionen, die wir nun getrennt betrachten wollen
- Erhalten eine Zielvariable Y mit drei Kategorien: "Infektion vom Typ I", "Infektion vom Typ II" und "keine Infektion"
- Keine Ordnung zwischen den Infektionstypen $\rightarrow$ nominalskaliertes Merkmal
- Kovariablen wie vorher NPLAN (Kaiserschnitt ungeplant, ja=1/nein=0), RISK (Vorliegen von Risikofaktoren, ja=1/nein=0) und ANTIB (Gabe einer Antibiotika-Prophylaxe, ja=1/nein=0)

Infektionen nach Kaiserschnittgeburten - Daten

		Kaiserschnitt geplant			Kaiserschnitt nicht geplant		
		Infektion			Infektion		
		I	II	nein	I	II	nein
Antibiotika	Risikofaktor	0	1	17	4	7	87
	kein Risikofaktor	0	0	2	0	0	0
keine Antibiotika	Risikofaktor	11	17	30	10	13	3
	kein Risikofaktor	4	4	32	0	0	9

Aus Fahrmeir, Kneib und Lang (2009)

Daten zur Lungenfunktion

- Daten texanischer Industriearbeiter
- Zielvariable Y beschreibt die Ergebnisse eines Atmungstests mit Kategorien "normal", "grenzwertig" und "abnormal"
- Haben also ordinalskaliertes Merkmal
- Die Kovariablen sind "Alter" und "Rauchverhalten"

Daten zur Lungenfunktion

Alter	Rauchverhalten	Testergebnis		
		nomral	grenzwertig	abnormal
< 40	kein Raucher	577	27	7
	früherer Raucher	192	20	3
	derzeitiger Raucher	682	46	11
40-59	kein Raucher	164	4	0
	früherer Raucher	145	15	7
	derzeitiger Raucher	245	47	27

Aus Fahrmeir, Kneib und Lang (2009)

Datenstruktur und Verteilungsmodelle

Darstellung der Zielvariablen

- Gehen von Zielvariable aus, die c geordnete oder ungeordnete Kategorien hat, die wir mit den Zahlen 1 bis c codieren
- Y hat damit Werte in $\{1, \ldots, c\}$
- Wollen die Wahrscheinlichkeiten $\pi_r = P(Y = r)$, in Abhängigkeit von den Kovariablen modellieren und dann schätzen
- Stelle dazu die Zielvariable multivariat dar, mit kategorien-spezifischen Zielvariablen

$$Y_r = \begin{cases} 1 \text{ falls } Y = r, \\ 0 \text{ sonst} \end{cases}, \qquad \text{für } r = 1, \ldots, q \text{ mit } q = c - 1$$

Darstellung der Zielvariablen

- Beachte: Wir definieren nur $c-1$ Variablen bei c Kategorien
- Der Zufallsvektor $\mathbf{Y} = (Y_1, \ldots Y_q)^T$ codiert den Zustand Y vollständig, denn für jedes $r < c$ gilt

$$Y = r \quad \Longleftrightarrow \quad Y_r = 1 \text{ und } Y_s = 0 \text{ für alle } s \neq r$$

und

$$Y = c \quad \Longleftrightarrow \quad Y_r = 0 \text{ für alle } r = 1, \ldots, c-1$$

- Die ausgelassene Kategorie c wird als *Referenzkategorie* bezeichnet
- In welchem Sinne dies eine Referenz ist: später

Darstellung der Zielvariablen

- Wir können allgemein auch jede andere Kategorie als Referenzkategorie bestimmen
- In dieser Vorlesung bleiben wir aber bei der Kategorie c
- Mit dieser Codierung gilt nun

$$\pi_r = P(Y = r) = P(Y_r = 1) \quad \text{für} \quad r < c$$

- Und

$$P(Y = c) = 1 - \pi_1 - \cdots - \pi_q = 1 - \sum_{s=1}^{q} \pi_s \quad \text{für} \quad q = c - 1$$

- π_c muss also nicht bestimmt oder geschätzt werden, sondern ergibt sich aus $\pi_1, \ldots, \pi_q$

Verteilung der Zielvariablen

- Die Dichte des Datenvektors $\mathbf{Y} = (Y_1, \ldots, Y_q)^T$ für eine Einzelbeobachtung ist

$$f(\mathbf{Y}|\boldsymbol{\pi}) = \pi_1^{Y_1} \cdots \pi_q^{Y_q} \cdot (1 - \pi_1 - \cdots - \pi_q)^{1-Y_1-\cdots-Y_q}$$

- Dabei ist

$$\boldsymbol{\pi} = (\pi_1, \ldots, \pi_q) \in \Theta = \{\boldsymbol{\pi} \in [0,1]^q : \sum_{r=1}^{q} \pi_q \leq 1\}$$

  der Vektor der Wahrscheinlicheiten $\pi_r = P(Y_r = 1)$ für $r = 1, \ldots, q$ (mit $q = c - 1$)

- Bei m stochastisch unabhängigen und identisch verteilten Individuen definieren wir für $r = 1, \ldots, q$ die Variable Y_r als Anzahl der Individuen, für die Kategorie r beobachtet wurde

Verteilung der Zielvariablen

- Der Datenvektor $\mathbf{Y} = (Y_1, \ldots, Y_q)$ fasst die Beobachtungen der gesamten Stichprobe (oder Gruppe) zusammen
- Er hat den Träger

$$\mathbb{T} = \{\, \mathbf{y} = (y_1, \ldots, y_q)^T \;:\; y_r \in \{0, \ldots, m\} \text{ für alle } r \text{ mit } \sum_{r=1}^{q} y_r \leq m \,\}$$

- Und die Dichte

$$f(\mathbf{y}|\boldsymbol{\pi}) = \frac{m!}{y_1! \ldots y_c!} \cdot \pi_1^{y_1} \cdots \pi_q^{y_q} \cdot \pi_c^{y_c},$$

  mit sich aus $\boldsymbol{\pi}$ ergebendem $\pi_c = 1 - \pi_1 - \ldots - \pi_q$ und aus sich aus $\mathbf{y}$ ergebenem $y_c = m - y_1 - y_2 - \ldots - y_q$

Verteilung der Zielvariablen

- Es handelt sich um eine Dichte bzgl. des Zählmaßes, d.h. $f(\mathbf{y}|\boldsymbol{\pi})$ ist die Wahrscheinlichkeit $P(\mathbf{Y} = \mathbf{y})$
- Man nennt die Verteilung auf $\mathbb{T}$ mit dieser Dichte bzw. Wahrscheinlichkeitsfunktion die *Multinomilaverteilung*
- Wir schreiben kurz $\mathbf{Y} \sim MN(m, \pi)$
- Erwartungswertvektor und Kovarianzmatrix eines Zufallsvektors $\mathbf{Y} \sim MN(m, \pi)$ sind

$$E(\mathbf{Y}) = m\pi = \begin{pmatrix} m\pi_1 \\ \vdots \\ m\pi_q \end{pmatrix} \quad \text{und} \quad Cov(\mathbf{Y}) = m \begin{pmatrix} \pi_1(1-\pi_1) & \dots & -\pi_1\pi_q \\ \vdots & \ddots & \vdots \\ -\pi_q\pi_1 & \dots & \pi_q(1-\pi_q) \end{pmatrix}$$

Verteilung der Zielvariablen

- Oft werden auch die relativen Häufigkeiten

$$\bar{\mathbf{Y}} = \left(\frac{Y_1}{m}, \ldots, \frac{Y_q}{m}\right)^T = (\bar{Y}_1, \ldots, \bar{Y}_q)^T$$

betrachtet

- Es folgt

$$E(\bar{\mathbf{Y}}) = \pi = \begin{pmatrix} \pi_1 \\ \ldots \\ \pi_q \end{pmatrix} \quad \text{und} \quad Cov(\bar{\mathbf{Y}}) = \frac{1}{m} \begin{pmatrix} \pi_1(1-\pi_1) & \ldots & -\pi_1\pi_q \\ \vdots & \ddots & \vdots \\ -\pi_q\pi_1 & \ldots & \pi_q(1-\pi_q) \end{pmatrix}$$

Multinomialverteilung als (mehrparametrige) Exponentialfamilie

- Es lässt sich zeigen, dass

$$f(\mathbf{y}|\boldsymbol{\pi}) = \exp\left(\mathbf{y}^T\boldsymbol{\theta} - m\,b(\boldsymbol{\theta}) - c(m, \mathbf{y})\right)$$

mit q-dimensionalem kanonischem Parameter

$$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)^T = \left(\log\left(\frac{\pi_1}{\pi_c}\right), \ldots, \log\left(\frac{\pi_q}{\pi_c}\right)\right)^T,$$

und kumulanten-Funktion

$$b(\boldsymbol{\theta}) = \log\left(1 + \sum_{s=1}^{q} e^{\theta_s}\right)$$

sowie $c(m, \mathbf{y}) = -\log[m!/(y_1! \ldots y_c!)]$

Multinomialverteilung als (mehrparametrige) Exponentialfamilie

- Die Dichten der Multinomialverteilungen bilden also ebenfalls eine Exponentialfamilie
- Allerdings mit einem q-dimensionalen $\mathbf{y}$ und einem q-dimensionalen kanonischen Parameter $\boldsymbol{\theta}$
- Man kann zeigen, dass

$$\pi_r(\boldsymbol{\theta}) = e^{\theta_r}\pi_c = e^{\theta_r} / \left(1 + \sum_{s=1}^{q} e^{\theta_s}\right)$$

- Die Kanonischen Parameter θ_r bestimmen also π_r als Vielfaches der Referenzwahrscheinlichkeit $\pi_c = 1/\left(1 + \sum_{s=1}^{q} e^{\theta_s}\right)$

Multinomialverteilung als (mehrparametrige) Exponentialfamilie

- In Analogie zur eindimensionalen, einparametrigen Exponentialfamilie lässt sich zeigen, dass

$$E(\mathbf{Y}) = m\frac{\partial\, b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \quad \text{und} \quad Cov(\mathbf{Y}) = m\,\frac{\partial^2\, b(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = m\left(\frac{\partial^2\, b(\boldsymbol{\theta})}{\partial \theta_r \partial \theta_s}\right)_{r,s=1}^{q}$$

- Darüber hinaus ist

$$\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\pi}} = \Sigma^{-1} \qquad \text{mit} \qquad \Sigma = \Sigma(\boldsymbol{\pi}) = \begin{pmatrix} \pi_1(1-\pi_1) & \dots & -\pi_1\pi_q \\ \vdots & \ddots & \vdots \\ -\pi_q\pi_1 & \dots & \pi_q(1-\pi_q) \end{pmatrix}$$

- $\Sigma = \Sigma(\boldsymbol{\pi})$ ist die Kovarianzmatrix eines $MN(1, \boldsymbol{\pi})$-verteilten Zuvallsvektors, also einer Einzelbeobachtung und damit die mehrdimensionale Verallgemeinerung der Varianzfunktion

Einzelbeobachtungen und gruppierte Daten

- Für Einzelbeobachtungen haben wir die folgende Datenstruktur:

$$\begin{array}{c} \text{Einheit } 1 \\ \vdots \\ \text{Einheit i} \\ \vdots \\ \text{Einheit n} \end{array} \begin{pmatrix} \mathbf{Y}_1^T = (Y_{11}, \ldots, Y_{1q}) \\ \vdots \\ \mathbf{Y}_i^T = (Y_{i1}, \ldots, Y_{iq}) \\ \vdots \\ \mathbf{Y}_n^T = (Y_{n1}, \ldots, Y_{nq}) \end{pmatrix} \begin{pmatrix} x_{11} \ldots x_{1k} \\ \vdots \\ x_{i1} \ldots x_{ik} \\ \vdots \\ x_{n1} \ldots x_{nk} \end{pmatrix}$$

- Gehen von n stochastisch unabhängigen Zufallsvektoren $\mathbf{Y}_i \sim MN(1, \boldsymbol{\pi}_i)$ aus, wobei der Wahrscheinlichkeitsvektor $\boldsymbol{\pi}_i$ vom Individuum i abhängen kann

Einzelbeobachtungen und gruppierte Daten

- Für gruppierte Daten haben wir die folgende Datenstruktur:

$$\begin{array}{c} \text{Gruppe 1} \\ \vdots \\ \text{Gruppe l} \\ \vdots \\ \text{Gruppe n} \end{array} \begin{pmatrix} m_1 \\ \vdots \\ m_l \\ \vdots \\ m_n \end{pmatrix} \begin{pmatrix} \mathbf{Y}_1^T = (Y_{11}, \dots, Y_{1q}) \\ \vdots \\ \mathbf{Y}_l^T = (Y_{l1}, \dots, Y_{lq}) \\ \vdots \\ \mathbf{Y}_n^T = (Y_{n1}, \dots, Y_{nq}) \end{pmatrix} \begin{pmatrix} x_{11} \dots x_{1k} \\ \vdots \\ x_{l1} \dots x_{lk} \\ \vdots \\ x_{m1} \dots x_{mk} \end{pmatrix}$$

- Hier nehmen wir an, dass $\mathbf{Y}_l = (Y_{l1}, \dots, Y_{lq}) \sim MN(m_l, \boldsymbol{\pi}_l)$
- Das entspricht den absoluten Häufigkeiten einer Gruppe von m_l stochastisch unabhängigen Einzelbeobachtungen, die alle einer $MN(1, \boldsymbol{\pi}_l)$-Verteilung folgen
- Nehmen also denselben Wahrscheinlichkeitsvektor $\boldsymbol{\pi}_l$ für alle Individuen der Gruppe l an

Modelle für ungeordnete Kategorien

Modelle für ungeordnete Kategorien

- Im Folgenden wollen wir

$$\pi_{ir} = P(Y_i = r|\mathbf{x}_i) = P(Y_{ir} = 1|\mathbf{x}_i)$$

  in Abhängigkeit von Kovariablenvektoren $\mathbf{x}_i$ modellieren und schätzen

- Nehmen zunächst an, dass die Kategorien der Zielvariable entweder ungeordnet sind oder wir die Ordnung nicht ausnutzen möchten

- Wir betrachten nun Strukturmodelle für $\boldsymbol{\pi}_i = (\pi_{i1}, \ldots, \pi_{iq})$, die diesen Wahrscheinlichkeitsvektor mit dem Kovariablenvektor $\mathbf{x}_i = (x_{i1}, \ldots, x_{ik})$ verknüpfen

Merkategoriale Logit-Modelle

- Im *mehrkatgorialen Logit-Modell* wird folgender Ansatz gemacht:

$$\pi_{ir} = P(Y_i = r|\mathbf{x}_i) = \frac{\exp\{\mathbf{x}_i\beta_r\}}{1 + \sum_{s=1}^{q} \exp\{\mathbf{x}_i\beta_s\}}, \qquad r = 1, \ldots, q$$

- Für die Referenzkategorie c gilt dann

$$\pi_{ic} = P(Y_i = c|\mathbf{x}_i) = 1 - \pi_{i1} - \ldots - \pi_{iq} = \frac{1}{1 + \sum_{s=1}^{q} \exp\{\mathbf{x}_i\beta_s\}}.$$

- Eine äquivalente Darstellung ist

$$\log(\pi_{ir}/\pi_{ic}) = \mathbf{x}_i\beta_r \quad \text{bzw.} \quad \pi_{ir}/\pi_{ic} = \exp\{\mathbf{x}_i\beta_r\},$$

wobei aus $\pi_{ir} = \pi_{ic}$ automatisch $\beta_r = \mathbf{0}$ folgt

Merkategoriale Logit-Modelle – Bemerkungen

- Mit dem mehrkategorialen Logit-Modell unterstellen wir also lineare Zusammenhänge zwischen den kanonischen Parametern $\theta_{ir} = \log(\pi_{ir}/\pi_{ic})$ und dem Kovariablenvektor $\mathbf{x}_i$
- Dies ist analog zum binären logistischen Modell, nur dass nun der kanonische Parameter mehrdimensional ist: $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{iq})^T$
- Haben für jede Kategorie entsprechend eigene Regressionskoeffizienten:

$$\beta_r = (\beta_{r1}, \ldots, \beta_{rk})^T, \quad \theta_{ir} = \mathbf{x}_i \beta_r, \quad r = 1, \ldots, q$$

- Also sind die Parameter β_r sowie die Prädiktoren $\eta_{ir} = \mathbf{x}_i \beta_r = \beta_{r1} + x_{i2}\beta_{r2} + \ldots + x_{ik}\beta_{rk}$ kategorien-spezifisch

Merkategoriale Logit-Modelle – Bemerkungen

- Der Prädiktor $\eta_{ir} = \mathbf{x}_i \beta_r$ ist der Logarithmus des relativen Risikos π_{ir}/π_{ic} zwischen der Kategorie r und der Referenzkategorie c
- Beim binären Modell ($c = 2$ und $q = 1$), ist die 0-Kategorie die Referenzkategorie und die Log-Odds $\log\left(\pi/(1-\pi)\right)$ ist der Logarithmus des relativen Risikos zwischen Kategorie 1 und Kategorie 0
- Das binäre Modell ist in diesem Sinne also ein Spezialfall des mehrkategorialen Modells

Merkategoriale Logit-Modelle – Interpretation der Regressionskoeffizienten

- $\beta_{rj} > 0$ bedeutet, dass $\pi_{ir}/\pi_{ic} = e^{\mathbf{x}_i\beta_r} = e^{\beta_{r2}}e^{\beta_{r2}x_{i2}} \cdots e^{\beta_{rk}x_{ik}}$ in x_{ij} steigt, und zwar um den Faktor $e^{\beta_{rj}} > 1$, wenn sich x_{ij} um eine Einheit erhöht
- Ein positives $\beta_{rj} > 0$ bedeutet aber nicht unbedingt, dass π_{ir} bei Erhöhung von x_{ij} steigt!
- Es gilt nämlich:

$$\pi_{ir} = \frac{e^{\mathbf{x}_i\beta_r}}{1 + \sum_{s=1}^{q} e^{\mathbf{x}_i\beta_s}} = \frac{1}{e^{-\mathbf{x}_i\beta_r} + \sum_{s=1}^{q} e^{\mathbf{x}_i(\beta_s - \beta_r)}}$$

steigt oder bleibt konstant in x_{ij} nur wenn $\beta_{rj} \geq 0$ und $\beta_{sj} \geq \beta_{rj}$ für alle $s \neq r$

Merkategoriale Logit-Modelle – Interpretation der Regressionskoeffizienten

- Ansonsten, d.h. wenn $\beta_{rj} < 0$ oder $\beta_{sj} - \beta_{rj} < 0$ für mindestens ein $s \neq r$, verhält sich π_{ir} nicht monoton in x_{ij}

- Um den Effekt von Änderungen in x_{ij} auf π_{ir} zu verstehen, kann man für die anderen Kovariablen typische Werte (z.B. die Mittelwerte) einsetzen und π_{ir} über x_{ij} plotten

Datenbeispiel Kaiserschnittgeburten

Kaiserschnittgeburten – Modell

- Definiere $\mathbf{Y}_i = (Y_{i1}, Y_{i2})^T$ so, dass $Y_{i1} = 1$ und $Y_{i2} = 0$ bei einer Infektion vom Typ 1, $Y_{i2} = 1$ und $Y_{i1} = 0$ bei einer Infektion vom Typ 2 und $Y_{i1} = Y_{i2} = 0$ bei keiner Infektion
- Fall "keine Infektion" ist also Referenzkategorie
- Die Kovariablen sind NPLAN, RISK, ANTIB, die alle Werte in $\{0, 1\}$ annehmen
- Das Strukturmodell ist

$$\log \frac{P(\text{ Infektion vom Typ r })}{P(\text{ keine Infektion })} = \beta_{r1} + \beta_{r2} \cdot \text{NPLAN} + \beta_{r3} \cdot \text{RISK} + \beta_{r4} \cdot \text{ANTIB}$$

bzw.

$$\frac{P(\text{ Infektion vom Typ r })}{P(\text{ keine Infektion })} = e^{\beta_{r1}} (e^{\beta_{r2}})^{\text{NPLAN}} (e^{\beta_{r3}})^{\text{RISK}} (e^{\beta_{r4}})^{\text{ANTIB}}$$

Kaiserschnittgeburten – Interpretation der Koeffizienten

- Beim Vorliegen von Risikofaktoren, d.h. RISK=1, verändert sich das Verhältnis zwischen P(Infektion vom Typ r) zu P(keine Infektion) um den Faktor $e^{\beta_{r3}}$
- Dieser Faktor kann als Odds-Ratio interpretiert werden

$$e^{\beta_{r3}} = \frac{P(\text{ Infektion vom Typ } r \mid \text{RISK} = 1)}{P(\text{ keine Infektion} \mid \text{RISK} = 1)} \Big/ \frac{P(\text{ Infektion vom Typ } r \mid \text{RISK} = 0)}{P(\text{ keine Infektion} \mid \text{RISK} = 0)},$$

- Da P(keine Infektion) nicht die Gegenwahrscheinlichkeit von P(Infektion vom Typ r) ist handelt sich hier eher um das Verhältnis zweier relativer Risiken als um ein Odds-Ratio

Kaiserschnittgeburten – Ergebnis

Infektion vom Typ 1			Infektion vom Typ 2		
	β	$\exp(\beta)$		β	$\exp(\beta)$
Intercept	-2.621	0.072	Intercept	-2.560	0.077
NPLAN	1.174	3.235	NPLAN	0.996	2.707
ANTIB	-3.520	0.030	ANTIB	-3.087	0.046
RISK	1.829	6.228	RISK	2.195	8.980

Aus Fahrmeir, Kneib, Lang (2009)

Kaiserschnittgeburten – Ergebnis

- Den geschätzten Koeffizienten entsprechend senkt eine Antibiotika-Prophylaxe das relative Risiko für Infektionen beiden Typs
- Risikofaktoren und ungeplante Kaiserschnitte erhöhen die relativen Risiken

Latente Nutzenmodelle

Latente Nutzenmodelle

- Das mehrkategoriale Logit-Modell und andere Modelle für nominal-skalierte Variablen können aus dem sog. *latenten Nutzenmodell* abgeleitet werden
- Dabei wird jeder Kategorie $r \in \{1, \ldots, c\}$ ein unbeobachteter, zufallsbehafteter Nutzen U_r zugeordnet
- Wir nehmen an, dass

$$Y = \text{argmax}_{r=1}^{c} U_r,$$

  wobei Y die beobachtete Kategorie ist
- In Worten: es realisiert sich immer die Kategorie mit dem größten (unbeobachteten) Nutzen

Latente Nutzenmodelle

- Jeder Beobachtungseinheit $i = 1, \ldots, n$ und Kategorie $r = 1, \ldots, q$ wird ein Nutzen zugeordnet, wobei angenommen wird, dass

$$U_{ir} = \tilde{\eta}_{ir} + \epsilon_{ir}$$

- Dabei ist $\tilde{\eta}_{ir}$ von den Kovariablen deterministisch abhängig und ϵ_{ir} ein stochastisch unabhängiger Störterm
- $\epsilon_{ir} \sim F$ für eine vorgegebene Verteilungsfunktion F
- Die Wahl von F und $\tilde{\eta}_{ir}$ legt dann das Modell fest

Mehrkategoriales Logit-Modell als latentes Nutzenmodell

- Wählen wir $F(x) = \exp(-\exp(-x))$ und sind $\epsilon_{i1}, \ldots, \epsilon_{iq}$ stochastisch unabhängig, dann gilt

$$P(Y_i = r) = \frac{\exp(\tilde{\eta}_{ir})}{\sum_{s=1}^{c} \exp(\tilde{\eta}_{is})} = \frac{\exp\{\tilde{\eta}_{ir} - \tilde{\eta}_{ic}\}}{1 + \sum_{s=1}^{c-1} \exp\{\tilde{\eta}_{is} - \tilde{\eta}_{ic}\}} = \frac{e^{\eta_{ir}}}{1 + \sum_{s=1}^{c-1} e^{\eta_{ir}}}$$

- Während die Prädiktoren $\tilde{\eta}_{ir}$ im ersten Term nicht eindeutig sind, da Zähler und Nenner mit einer beliebigen Zahl multipliziert werden können, sind $\eta_{ir} = \tilde{\eta}_{ir} - \tilde{\eta}_{ic}$ im letzten Term eindeutig

- Zum Erreichen von Eindeutigkeit muss also eine Kategorie als Referenzkategorie gewählt werden (oben Kategorie c)

Mehrkategoriales Probit-Modell

- Andere Verteilungen für ϵ_{ir} führen zu anderen Modellen
- Wenn z.B. $\epsilon_{ir} \sim N(0,1)$ und $\epsilon_{i1}, \ldots, \epsilon_{iq}$ stochastisch unabhängig sind, dann ergibt sich das sogen. *mehrkategoriale (unabhängige) Probit-Modell*
- Bei nur zwei Kategorien erhalten wir im Wesentlichen (d.h. bis auf bekannte Umrechnungsfaktoren) das bekannte binäre Probit-Modell
- Wenn $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{iq})^T \sim N(0, \Lambda)$ für ein nicht-diagonale Kovarianzmatrix Λ, dann spricht man von einem *multivariaten Probit-Modell*

Erweiterung um kategorienspezifische Kovariablen

- Oft gibt es Kovariablen, die nicht nur vom Individuum sondern auch von der beobachteten Kategorie abhängen
- Entsprechend sei $w_{ir} \in \mathbb{R}^l$ ein von der Kategorie r abhängiger Kovariablenvektor
- Beispiel: Fahrpreis von öffentlichen Transportmitteln, deren Wahl Y (und die Gründe dafür) in einer Studie untersucht werden sollen
- Der Fahrpreis hängt offensichtlich vom gewählten Verkehrsmittel r und vom Individuum i ab, denn Kinder, Schüler, Rentner etc. haben Sonderpreise
- Unter Hinzunahme solcher kategoriespezifischen Kovariablen könnten wir nun

$$\eta_{ir} = \mathbf{x}_i \beta_r + (w_{ir} - w_{ic})\gamma \quad \text{für} \quad \gamma \in \mathbb{R}^l, \qquad r = 1, \ldots, q \quad (q = 1 - c)$$

ansetzen

Erweiterung um kategorienspezifische Kovariablen

- Der Regressionskoeffizientenvektor γ hängt nicht von der Kategorie ab
- Man spricht dann von einem *globalen Regressionskoeffizienten*
- Bei Annahme von unabhängigen, extremwertverteilten Störgrößen ϵ_i führt das zu

$$\pi_{ir} = \frac{\exp\{\mathbf{x}_i\beta_r + (w_{ir} - w_{ic})\gamma\}}{1 + \sum_{s=1}^{q} \exp\{\mathbf{x}_i\beta_r + (w_{ir} - w_{ic})\gamma\}}.$$

Ordinale Modelle

Ordinale Modelle

- Wir betrachten nun Modelle für ordinalskalierte Zielvariablen Y mit Kategorien $1, \ldots, c$
- Nehmen an, dass die Codierung geordnet ist
- Das heißt Kategorie 2 ist "größer" (also je nach Kontext: stärker, schwerer, besser, etc.) als Kategorie 1 und Kategorie 3 ist "größer" als 2 und 1, usw.
- Ziel ist es die Ordnung (und die in ihr entaltene Information) für eine sparsamere Parametrisierung auszunutzen

Das kumulative und Schwellenwert-Modell

- Im Folgenden sei U eine nicht beobachtbare, latente Variable (z.B. die unbeobachtete Schädigung der Lunge)
- Weiter sei für jede Beobachtungseinheit i

$$U_i = -\mathbf{x}_i\beta + \epsilon_i,$$

  wobei β der Vektor der Regressionkoeffizienten und ϵ_i eine Störvariable mit Verteilungsfunktion F sind

- U_i hängt nun nicht mehr von der Kategorie ab (nur noch vom Individuum)
- Nur noch einen globalen Regressionskoeffizienten $\beta \Rightarrow$ deutlich weniger Parameter als in vorigen Modellen

Verknüpfung zwischen Y_i und U_i

- Wir gehen nun davon, dass

$$Y_i = r \qquad \Longleftrightarrow \qquad \theta_{r-1} < U_i \leq \theta_r, \qquad r = 1, \ldots, q \ ,$$

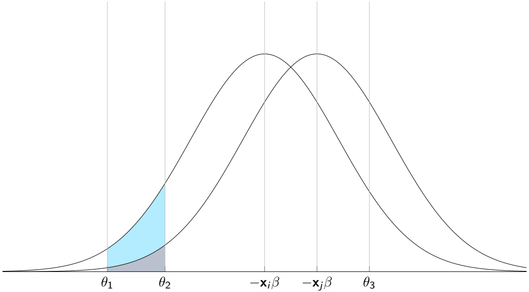
für unbekannte Parameter $-\infty = \theta_0 < \theta_1 < \ldots < \theta_c = \infty$

- Man sieht leicht ein, dass

$$Y_i \leq r \qquad \Longleftrightarrow \qquad U_i \leq \theta_r$$

- Die beiden Beziehungen sind sogar äquivalent

Verknüpfung zwischen Y_i und U_i – Illustration



Schwellenwerte und Dichten der latenten Variablen

Bemerkungen zur Eindeutigkeit der Parameter

- Mit $x_{i1} = 1$ für alle i (also wenn das Modell einen Intercept hat) ist $Y_i = r$ äquivalent zu
$$\theta_{r-1} < -\beta_1 - x_{i2}\beta_2 - \ldots - x_{ik}\beta_k + \epsilon_i < \theta_r$$
- Dies ist aber wiederum äquivalent zu
$$\tilde{\theta}_{r-1} < -\tilde{\beta}_1 - x_{i2}\beta_2 - \ldots - x_{ik}\beta_k + \epsilon_i < \tilde{\theta}_r$$
mit $\tilde{\theta}_r = \theta_r - a$ und $\tilde{\beta}_1 = -\beta_1 - a$ für beliebiges $a \in \mathbb{R}$
- Die Parameter $\beta_1, \theta_1, \ldots, \theta_q$ sind nicht eindeutig durch die Ereignisse $Y = r$ bestimmt (wir können sie "verschieben")
- Die Wahrscheinlichkeiten sind daher durch $\beta_1, \theta_1, \ldots, \theta_q$ überparametrisiert

Bemerkungen zur Eindeutigkeit der Parameter

- Um Eindeutigkeit zu erzielen, müssen die Parameter daher eingeschränkt werden
- Es gibt es zwei natürliche Ansätze Eindeutigkeit der Parameter zu erzielen:
  a) wir setzen $\theta_1 = 0$, woraus sich eindeutige $\beta_1, \theta_2, \ldots, \theta_q$ ergeben, oder
  b) wir definieren $\beta_1 = 0$, wodurch $\theta_1, \theta_2, \ldots, \theta_q$ eindeutig werden.
- Meist wird Einschränkung a) gewählt

Das kumulative Modell

- Aus den Überlegungen $U_i = -\mathbf{x}_i\beta + \epsilon_i$ und $Y_i \leq r \Leftrightarrow U_i \leq \theta_r$ folgt

$$P_{\mathbf{x}_i}(Y_i \leq r) = P(U_i \leq \theta_r) = P(\epsilon_i \leq \theta_r + \mathbf{x}_i\beta) = F(\theta_r + \mathbf{x}_i\beta), \qquad r = 1, \ldots, q$$

  wobei F die Verteilungsfunktion von ϵ_i ist

- Man nennt dies das *kumulative Modell*

- Es folgt $P_{\mathbf{x}_i}(Y_i = 1) = F(\theta_1 + \mathbf{x}_i\beta)$ und

$$P_{\mathbf{x}_i}(Y_i = r) = F(\theta_r + \mathbf{x}_i\beta) - F(\theta_{r-1} + \mathbf{x}_i\beta) \quad \text{für} \quad r = 2, \ldots, q$$

- Je nach Wahl von F erhält man verschiedene Modelle (siehe kommende Folien)

Das kumulative Logit-Modell

- Beim *kumulativen Logit-Modell* wählt man für F die Verteilungsfunktion der logistischen Verteilung $F(\eta) = e^\eta/(1+e^\eta)$
- Damit ist

$$P_{\mathbf{x}_i}(Y_i \leq r) = \frac{\exp\{\theta_r + \mathbf{x}_i\beta\}}{1+\exp\{\theta_r + \mathbf{x}_i\beta\}} \qquad \Longleftrightarrow \qquad \log \frac{P_{\mathbf{x}_i}(Y_i \leq r)}{P_{\mathbf{x}_i}(Y_i > r)} = \theta_r + \mathbf{x}_i\beta$$

- Modellieren also Log-Odds der kumulativen Wahrscheinlichkeiten $P_{\mathbf{x}_i}(Y_i \leq r)$
- Ändern sich mit $\mathbf{x}_i$ alle um denselben (kategorien-unabhängigen) Faktor $e^{\mathbf{x}_i\beta}$
- Man spricht im Englischen daher auch vom *Proportional Odds Model*

Das kumulative Extremwert-Modell

- Beim kumulativen Extremwert-Modell verwenden wir die Extremwertverteilung $F(x) = 1 - \exp(-\exp(x))$
- Damit ist

$$P_{\mathbf{x}_i}(Y_i \leq r) = 1 - \underbrace{\exp(-\exp(\theta_r + \mathbf{x}_i\beta))}_{P_{\mathbf{x}_i}(Y_i > r)} \iff \log(-\log[P_{\mathbf{x}_i}(Y_i > r)]) = \theta_r + \mathbf{x}_i\beta$$

- Betrachten den Parameter

$$\nu_r(\mathbf{x}_i\beta) = -\log[P_{\mathbf{x}_i}(Y_i > r)] = e^{\theta_r} e^{\mathbf{x}_i\beta},$$

- Für $\mathbf{x}_i$ und $\tilde{\mathbf{x}}_i$ gilt dann

$$\nu_r(\mathbf{x}_i\beta)/\nu_r(\tilde{\mathbf{x}}_i\beta) = e^{(\mathbf{x}_i - \tilde{\mathbf{x}}_i)\beta}$$

- Dies ist unabhängig von der Kategorie r

Das kumulative Extremwert-Modell

- Das bedeutet: die negativ logarithmischen Überschreitungswahrscheinlichkeiten ν_r ändern sich mit x_{ij} um einen von r unabhängigen Faktor
- Wenn sich z.B. x_{ij} um eine Einheit ändert, dann ist dies der Faktor e^{β_j}
- Ganz allgemein gilt, dass sich

$$\nu_r(\mathbf{x}_i\beta) = \exp\{F^{-1}[P(Y_i \leq r|\mathbf{x}_i)]\} = e^{\theta_r} e^{\mathbf{x}_i\beta}$$

mit x_{ij} um einen von r unabhängigen Faktor ändert

Das Kumulative Extremwert-Modell als diskretes Hazard-Modell

- Man kann zeigen, dass im kumulativen Extremwert-Modell gilt:

  $$P_{\mathbf{x}_i}(\, Y_i = r \mid Y_i \geq r \,) = F(\Delta_r + \mathbf{x}_i \beta),$$

  mit $\Delta_r = \log\{e^{\theta_r} - e^{\theta_{r-1}}\}$

- D.h., dass die diskrete Hazard $P_{\mathbf{x}_i}(\, Y_i = r \mid Y_i \geq r \,)$ dem gleichen Modell folgt, wie die kumulative Wahrscheinlichkeit $P_{\mathbf{x}_i}(Y_i \leq r)$, allerdings mit Schwellenwerten Δ_r statt θ_r (β ist identisch)

- Man kann zeigen, dass sich aus den diskreten Hazards die Wahrscheinlichkeiten $P_{\mathbf{x}_i}(Y_i = r)$ vollständig rekonstruieren lassen:

  $$P_{\mathbf{x}_i}(Y_i = r) = P_{\mathbf{x}_i}(Y_i = r | Y_i \geq r) \prod_{s=1}^{r-1} \{1 - P_{\mathbf{x}_i}(Y_i = s | Y_i \geq s)\}$$

Sequentielle Modelle (Hazard-Modelle)

- Betrachten nun Modelle, bei denen direkt die diskrete Hazard modelliert wird:

$$P_{\mathbf{x}_i}(Y_i = r | Y_i \geq r) = F(\theta_r + \mathbf{x}_i\beta)$$

- F ist eine vorab festgelegte Responsefunktion
- Haben gesehen, dass obiges Modell, wenn F einem Extremwert-Verteilungsmodell folgt, äquivalent zum Log-Log-Modell für $P_{\mathbf{x}_i}(Y_i = r)$ ist
- Das sequentielle Log-Log-Modell entspricht also dem (normalen) Log-Log-Modell

Sequentielle Modelle (Hazard-Modelle)

- Wir können aber auch andere Responsefunktionen F verwenden, z.B. die Verteilungsfunktionen der
  - logistischen Verteilung $\longrightarrow$ *sequentielles logistisches Modell*
  - Standard-Normalverteilung ($F = \Phi$) $\longrightarrow$ *sequentielles Probit-Modell*
- Ein sequentielles Modell (diskretes Hazard-Modell) ist dann sinnvoll, wenn sich die Zustände $1, \ldots, c$ auf unbeachtete Weise zeitlich nacheinander ergeben (auf 1 folgt 2, darauf folgt 3 etc.)
- Wir beobachten nur den End- bzw. Istzustand
- Beispiel: Verschiedene Stadien einer Erkrankung (z.B. MS), die sich zeitlich nacheinander entwickeln

Sequentielle Modelle (Hazard-Modelle)

- Es gilt auch allgemein in einem sequentiellen Modell

$$P_{\mathbf{x}_i}(Y_i = r) = F(\Delta_r + \mathbf{x}_i\beta)\prod_{s=1}^{r-1}\{1 - F(\Delta_s + \mathbf{x}_i\beta)\}$$

- Wieder muss entweder Δ_1 oder β_1 (auf 0) festgelegt werden

Likelihood-Inferenz

Likelihood und Log-Likelihood-Kern

- Haben für alle vorgestellten Modelle, dass $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iq})^T \sim MN(m_i, \pi_i)$, wobei π_i auf verschiedene Weisen von $\mathbf{x}_i$ abhängen kann
- Die Likelihood dieser Daten ist

$$L(\beta) = \prod_{i=1}^{n} f(\mathbf{Y}_i|\boldsymbol{\pi}_i) \quad \text{mit} \quad f(\mathbf{Y}_i|\boldsymbol{\pi}_i) = \frac{n!}{Y_{i1}! \ldots Y_{iq}! Y_{ic}!} \cdot \pi_{i1}^{Y_{i1}} \cdots \pi_{iq}^{Y_{iq}} \cdot \pi_{ic}^{Y_{ic}},$$

wobei $Y_{ic} = n - Y_{i1} - \ldots - Y_{iq}$ und $\pi_{ic} = 1 - \pi_{i1} - \ldots - \pi_{iq}$

- Der Log-Likelihood-Kern ist damit

$$l(\beta) = \sum_{i=1}^{n} (Y_{i1} \log \pi_{i1} + \ldots + Y_{iq} \log \pi_{iq} + Y_{ic} \log \pi_{ic})$$

Score, Fisher-Information und MLE

- Der MLE $\hat{\beta}$ ist die Lösung von $\mathbf{s}(\beta) = \partial l(\beta)/\partial\beta = \mathbf{0}$
- Man kann zeigen, dass der Score $\mathbf{s}(\beta)$ die asymptotische Kovarianz

$$\mathbf{F}(\beta) = E\{\mathbf{s}(\beta)\mathbf{s}(\beta)^T\}$$

  hat
- $\mathbf{F}(\beta)$ ist die Fisher-Matrix
- Unter geeigneten Regularitätsannahmen gilt für den MLE von β

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

Allgemeine Darstellung

- Die verschiedenen Modelle unterscheiden sich in der Weise, wie die π_{ir} $(r = 1, \ldots, q)$ mit $\mathbf{x}_i$ verknüpft werden

- Eine für alle Modelle gültige, allgemeine Darstellung ist

$$\pi_{ir} = h_r(\eta_{i1}, \ldots, \eta_{iq}), \qquad r = 1, \ldots, q, \qquad i = 1, \ldots, n$$

- **Beispiele**
  (a) Im mehrkategorialen Logit-Modell gilt mit $\eta_i = \mathbf{x}_i \beta_r$ bzw. $\eta_{ir} = \mathbf{x}_i \beta_r + (\mathbf{w}_{ir} - \mathbf{w}_{ic})\gamma$ im erweiterten Modell

$$\pi_{ir} = \exp(\eta_{ir}) / \left(1 + \sum_{s=1}^{q} \exp(\eta_{is})\right) = h_r(\eta_{i1}, \ldots, \eta_{iq})$$

  (b) Im ordinalen kumulativen Modell haben wir mit $\eta_{ir} = \theta_r + \mathbf{x}_i \beta$

$$\pi_{ir} = F(\eta_{ir}) - F(\eta_{ir-1}) = h_r(\eta_{i1}, \ldots, \eta_{iq}), \qquad r = 1, \ldots, q$$

Allgemeine Matrixschreibweise

- Mit dieser Darstellung können wir Score und Fisher-Matrix in eine allgemeine Matrix-Darstellung bringen
- Diese wird zeigen, wie die Maximum-Likleihood-Schätzer bestimmt und entsprechende lineare Hypothesen getestet werden können
- Im Folgenden sei wieder $\boldsymbol{\eta}_i = (\eta_{i1}, \ldots, \eta_{iq})$
- In allen, in diesem Kapitel betrachteten Modellen gilt, dass mit geeignetem β und geeigneter Designmatrix $\mathbf{X}_i$

$$\boldsymbol{\eta}_i = \mathbf{X}_i \beta$$

Allgemeine Matrixschreibweise – Beispiele

(a) Im mehrkategoriellen Logit-Modell ist $\beta = (\beta_1, \ldots, \beta_q, \gamma)^T \in \mathbb{R}^{k \cdot q+1}$ und

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{x}_i & 0 & \ldots & 0 & w_{i1} - w_{ic} \\ 0 & \mathbf{x}_i & & 0 & w_{i2} - w_{ic} \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & \ldots & 0 & \mathbf{x}_i & w_{iq} - w_{ic} \end{pmatrix} \in \mathbb{R}^{q \times (k \cdot q+1)}$$

(b) Für ordinale kumulative Modelle gilt $\beta = (\theta_1, \ldots, \theta_q, \beta_1, \ldots, \beta_k)^T \in \mathbb{R}^{q+k}$ und

$$\mathbf{X}_i = \begin{pmatrix} 1 & 0 & \cdots & 0 & \mathbf{x}_i \\ \vdots & \ddots & & \vdots & \vdots \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & \mathbf{x}_i \end{pmatrix} \in \mathbb{R}^{q \times (q+k)}$$

Allgemeine Matrixschreibweise – Score

- Man kann zeigen, dass der Score dann gegeben ist als

$$\mathbf{s}(\beta) = \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{D}_i \Sigma_i^{-1} (\mathbf{Y}_i - m_i \boldsymbol{\pi}_i)$$

- Dabei ist

$$\mathbf{D}_i = \frac{\partial \mathbf{h}}{\partial \boldsymbol{\eta}_i}(\boldsymbol{\eta}_i) = \begin{pmatrix} \frac{\partial h_1}{\partial \eta_{i1}}(\boldsymbol{\eta}_i) & \dots & \frac{\partial h_q}{\partial \eta_{i1}}(\boldsymbol{\eta}_i) \\ \vdots & & \vdots \\ \frac{\partial h_1}{\partial \eta_{iq}}(\boldsymbol{\eta}_i) & \dots & \frac{\partial h_q}{\partial \eta_{iq}}(\boldsymbol{\eta}_i) \end{pmatrix} \in \mathbb{R}^{q \times q}$$

- und Σ_i die Kovarianzmatrix einer multinomialverteilten Einzelbeobachtung

$$\Sigma_i = \begin{pmatrix} \pi_{i1}(1-\pi_{i1}) & -\pi_{i1}\pi_{i2} & \dots & -\pi_{i1}\pi_{iq} \\ -\pi_{i2}\pi_{i1} & \pi_{i2}(1-\pi_{i2}) & & \\ \vdots & & \ddots & \vdots \\ -\pi_{iq}\pi_{i1} & \dots & \dots & \pi_{iq}(1-\pi_{iq}) \end{pmatrix}$$

Allgemeine Matrixschreibweise – Fisher-Matrix

- Aus dieser Darstellung des Scores ergibt sich

$$\mathbf{F}(\beta) = E\{\mathbf{s}(\beta)\mathbf{s}(\beta)^T\} = \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{W}_i \mathbf{X}_i$$

mit $\mathbf{W}_i = \mathbf{D}\Sigma_i^{-1}\Sigma_i\Sigma_i^{-1}\mathbf{D}_i^T = \mathbf{D}_i\Sigma_i^{-1}\mathbf{D}_i^T$

- Mit den Bezeichnungen $\mathbf{Y} = (\mathbf{Y}_1^T, \ldots, \mathbf{Y}_n^T)^T$, $\boldsymbol{\mu} = (m_1\boldsymbol{\pi}_1^T, \ldots, m_n\boldsymbol{\pi}_n^T)^T$ und $\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_n \end{pmatrix}$, sowie $\boldsymbol{\Sigma} = \text{diag}(\Sigma_1, \ldots, \Sigma_n)$, $\mathbf{D} = \text{diag}(\mathbf{D}_1, \ldots, \mathbf{D}_n)$ und $\mathbf{W} = \text{diag}(\mathbf{W}_1, \ldots, \mathbf{W}_n)$ gilt schliesslich

$$\mathbf{s}(\beta) = \mathbf{X}^T\mathbf{D}\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \boldsymbol{\mu}) \quad \text{und} \quad \mathbf{F}(\beta) = \mathbf{X}^T\mathbf{W}\mathbf{X}.$$

Numerische Bestimmung des MLE

- Der Maximum-Likelihood-Schätzer für β kann wieder mit dem Fisher-Scoring-Verfahren numerisch-iterativ berechnet werden:

$$\hat{\beta}^{(k+1)} = \hat{\beta}^{(k)} + \mathbf{F}^{-1}(\hat{\beta}^{(k)})\,\mathbf{s}(\hat{\beta}^{(k)})$$

- Der Scoring-Algorithmus kann alternativ wieder als eine *iterative Kleinste-Quadratschätzung* dargestellt werden:

$$\hat{\beta}^{(k+1)} = (\mathbf{X}^T\mathbf{W}^{(k)}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{(k)}\,\tilde{\mathbf{Y}}^{(k)}$$

mit $\tilde{\mathbf{Y}}^{(k)} = \mathbf{X}\hat{\beta}^{(k)} + (\mathbf{D}^{(k)})^{-T}\{\mathbf{Y} - \boldsymbol{\mu}^{(k)}\}$ wobei $\mathbf{D}^{-T}$ die Matrixinverse von $\mathbf{D}^T$ ist und $\mathbf{W}^{(k)}$, $\mathbf{D}^{(k)}$ sowie $\boldsymbol{\mu}^{(k)}$ durch Einsetzen von $\hat{\beta}^{(k)}$ gebildet werden

Asymptotische Eigenschaften und lineare Hypothesentests

- Unter relativ schwachen Regularitätseigenschaften gilt $\hat{\beta} \overset{a}{\sim} N(\beta, \mathbf{F}^{-1}(\hat{\beta}))$
- Damit können wieder lineare Hypothesen der Form

$$H_0 : \mathbf{C}\beta = \mathbf{d} \quad \text{vs.} \quad H_1 : \mathbf{C}\beta \neq \mathbf{d}$$

  durch den Likelihood-Quotienten-, Wald- oder Score-Test getestet werden
- Alle drei Statistiken sind unter H_0 approximativ χ^2_r-verteilt mit $r = \text{Rang}(\mathbf{C})$, sodass wir H_0 verwerfen können, falls die entsprechende Statistik größer oder gleich dem Quantil $Q_r^{\chi^2}(1-\alpha)$ ist

Teststatistiken

Die entsprechenden Teststatistiken sind wieder wie folgt:

- Likelihood-Quotienten-Statistik:

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

  wobei $\tilde{\beta}$ der MLE von β unter der Restriktion $\mathbf{C}\tilde{\beta} = \mathbf{d}$ ist

- Wald-Statistik

$$W = (\mathbf{C}\hat{\beta} - \mathbf{d})^T [\mathbf{C}\,\mathbf{F}^{-1}(\hat{\beta})\,\mathbf{C}^T]^{-1}(\mathbf{C}\hat{\beta} - \mathbf{d})$$

  wobei $\hat{\beta}$ der unrestringierte MLE von β ist

- Score-Statistik

$$U = \mathbf{s}(\tilde{\beta})^T \mathbf{F}^{-1}(\tilde{\beta})\mathbf{s}(\tilde{\beta})$$

  mit $\tilde{\beta}$ wie bei der Likelihood-Quitienten-Statistik

Beispiel: Infektionen nach Kaiserschnitt-Geburten

- Wollen testen, ob die Einflüsse von NPLAN und NRISK für die zwei Typen von Infektionen identisch sind.

- Dazu betrachten wir die Hypothesen

$$H_0 : \beta_{1N} = \beta_{2N} \text{ und } \beta_{1R} = \beta_{2R} \qquad \text{vs.} \qquad H_1 : \beta_{1N} \neq \beta_{2N} \text{ oder } \beta_{1R} \neq \beta_{2R}$$

- Sie sind von der Form $H_0 : \mathbf{C}\beta = \mathbf{d}$ mit

$$\mathbf{C} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \end{pmatrix} \quad \text{und} \quad \mathbf{d} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Beispiel: Infektionen nach Kaiserschnitt-Geburten

- Da Rang($\mathbf{C}$) $= 2$ gilt $lq \sim \chi_2^2$ unter H_0
- Aus den Daten ergibt sich z.B. für die Likelihood-Quotienten-Teststatistik $lq = 0.8467$
- Mit $Q_2^{\chi^2}(0.95) = 5.99$ folgt, dass wir H_0 nicht verwerfen können
- Es gibt damit keinen klaren Hinweis darauf, dass die Kovariablen NPLAN und NRISK die beiden Typen von Infektionen unterschiedlich beeinflussen