

Technische Universität München

ZENTRUM MATHEMATIK

**Regressionsdiagnostiken
für Zähldaten**

Diplomarbeit

von

Benjamin Weiderer

Themensteller: Prof. Dr. Czado

Betreuer: Prof. Dr. Czado, Susanne Gschlöbl

Abgabetermin: 11.08.2004

Hiermit erkläre ich, daß ich die Diplomarbeit selbständig angefertigt und nur die angegebenen Quellen verwendet habe.

München, 11.08.2004

Inhaltsverzeichnis

1	Einleitung	1
2	Lineare Modelle (LM)	3
2.1	Modellbeschreibung und Annahmen	3
2.2	Parameterschätzung in LM's	4
2.3	Statistische Hypothesentests in LM's	7
2.4	Gewichtete Regression	10
3	Regressionsdiagnostiken für LM's	12
3.1	y-Ausreißer und x-Ausreißer	12
3.2	Einflußreiche Beobachtungen	13
3.3	Residuen und Hatmatrix	13
3.4	Ausreißer-Test (Mean-Shift-Outlier-Test)	21
3.5	Bestimmung des Einflusses einer Beobachtung	23
3.5.1	DFFITS und DFBETAS	23
3.5.2	Cook's Distanz	25
3.5.3	COVRATIO und FVRATIO	26
3.6	Regressionsdiagnostiken bei Gewichtung einer Beobachtung	28
3.6.1	Gewichtung	28
3.6.2	Influenzfunktion und Influenzkurven	31
4	Generalisierte Lineare Modelle (GLM)	35
4.1	Modellbeschreibung und Annahmen	35
4.1.1	Die exponentielle Familie	36
4.1.2	Die Komponenten eines GLM's	38
4.2	Parameterschätzung in GLM's	39
4.3	Bestimmung der Anpassungsgüte und Modellwahl	48
5	Regressionsdiagnostiken für GLM's	51
5.1	Der Patentdatensatz	51
5.1.1	Beschreibung des Datensatzes	51
5.1.2	Explorative Datenanalyse	52
5.1.3	Poissonregression	53
5.2	Residuen und Hatmatrix	54
5.3	Ein-Schritt-Approximation von $\hat{\beta}_{-i}$ bei GLM's	59
5.4	Bestimmung des Einflusses einer Beobachtung	62
5.4.1	DFFITS und DFBETAS	63
5.4.2	Cook's Distanz	68

5.4.3	COVRATIO	71
5.4.4	Reduktion der Devianz	74
5.5	Regressionsdiagnostiken bei Gewichtung einer Beobachtung	79
5.5.1	Gewichtung	79
5.5.2	Influenzkurven	80
5.5.3	Reduktion der Devianz	81
6	Negativ-Binomial-Regression mit geschätztem α	85
6.1	Profiling	85
6.2	Regressionsdiagnostiken	91
7	Regressionsdiagnostik für Negativ-Binomial-Modelle	94
7.1	Negativ-Binomial-Regression mit bekanntem α für den Patentdatensatz	94
7.2	Negativ-Binomial-Regression mit geschätztem α für den Patentdatensatz	102
8	Fallstudie: Aggregierter KFZ-Haftpflichtdatensatz	110
8.1	Offset in der Regression	110
8.2	Ein aggregierter KFZ-Haftpflichtdatensatz	111
8.2.1	Beschreibung des aggregierten KFZ-Haftpflichtdatensatzes	111
8.2.2	Explorative Datenanalyse	112
8.2.3	Poissonregression	113
8.3	Negativ-Binomial-Regression mit bekanntem α	123
8.4	Negativ-Binomial-Regression mit geschätztem α	129
9	Zusammenfassung	139
A	Die Poissonverteilung	141
B	Die Negativ-Binomial-Verteilung	143
C	Maximum-Likelihood-Schätzung	144
D	Newton-Raphson- und Fisher-Scoring-Methode	145
E	Benutzerdefinierte SPlus-Funktionen	147
F	Aggregierter KFZ-Haftpflichtdatensatz	158
	Literaturverzeichnis	160
	Abbildungsverzeichnis	162
	Tabellenverzeichnis	164

Kapitel 1

Einleitung

Wie das Thema bereits verrät, geht es in dieser Arbeit hauptsächlich um Zähldaten. Um eine Anwendung in der Praxis zu verdeutlichen, kann man das Beispiel eines Versicherungsunternehmens anführen. Insbesondere in der KFZ-Versicherung interessiert die durchschnittliche Anzahl der Schäden, die für das nächste Jahr zu erwarten sind. Für die Tarifbildung ist nun entscheidend, welche Gruppe von Fahrzeughaltern deutlich weniger Schäden verursachen als andere Gruppen. Zuerst muss ein statistisches Modell erstellt werden, das die Daten in ihrer wesentlichen Struktur wiedergibt. Die interessierende Größe Schadenzahl wird also in Abhängigkeit von den Größen wie das Alter des Fahrzeughalters, Geschlecht des Fahrzeughalters, PS-Zahl des Autos, usw. versucht zu erklären. Da die zu untersuchende Größe Schadenzahl eine Zählgröße ist, werden die Daten aufgrund eines Zählmodells beschrieben. Das bekannteste ist das Poissonmodell, welches die Eigenschaft hat, dass Erwartungswert und Varianz gleich sind. Nun kann es aber vorkommen, dass die Daten durch dieses Modell nicht ausreichend angepasst werden, was z.B. durch eine deutlich höhere Devianz als die Anzahl der Freiheitsgrade ersichtlich ist. Es muss mittels einer Residuenanalyse überprüft werden, ob die systematische Komponente des Modells richtig spezifiziert wurde und gegebenenfalls müssen noch weitere relevante Größen, die die Schadenzahl beeinflussen, in den linearen Prädiktor aufgenommen werden. Falls ein Fehler in der systematischen Komponente ausgeschlossen ist, kann man davon ausgehen, dass die Ursache für die schlechte Datenanpassung in der Modellannahme zu finden ist. Die häufigste auftretende Modellverletzung beim Poissonmodell ist die sogenannte Überdispersion, welche auftritt, wenn die Zählmodellen in einem größerem Maße um den Erwartungswert streuen, als durch das Poissonmodell erwartet wird. Deshalb geht man häufig zum Negativ-Binomial-Modell über, welches zwei Parameter μ und α besitzt, wovon α fest gewählt oder ebenso geschätzt werden kann.

Der interessierende Sachverhalt in dieser Arbeit ist nun, welchen Einfluß einzelne Beobachtungen auf die geschätzten Werte in den entsprechenden Modellen haben. Statistiker sind immer von ihren Datensätzen und deren Richtigkeit abhängig. Einzelne Beobachtungen in einem Datensatz können die geschätzten Werte deutlich beeinflussen. Deshalb ist es wichtig, das Ausmaß eines solchen Einflusses auf den Fit, die geschätzten Regressionskoeffizienten und das Modell zu quantifizieren. Dieser Einfluß kann mit geeigneten Diagnostiken, welche in dieser Arbeit entwickelt werden sollen, beschrieben werden. Hat eine einzelne Beobachtung einen großen Einfluß auf die geschätzten Werte, bleibt es zu entscheiden, wie mit der Beobachtung verfahren werden soll. Die Richtigkeit des Datenpunktes kann nochmals nachgeprüft werden, d.h. man korrigiert den möglichen Fehler oder man löscht den Datenpunkt aus dem Datensatz. Wenn diese einflußreichen Beobachtungen allerdings keine Meßfehler aufweisen, können diese Punkte somit wichtige Informationen enthalten, die auf einen bestimmten Sachverhalt aufmerksam machen können.

In Kapitel 2 und 3 werden die Linearen Modelle und darauf basierend die Regressionsdiagno-

stiken beschrieben. Die Definition von Ausreißern und einflußreichen Beobachtungen ist hier ein grundlegender Punkt. Bei Linearen Modellen geht man von einer Normalverteilungsannahme der Zielvariable aus. Um den Einfluß einer Beobachtung festzustellen, betrachtet man den gesamten Datensatz und beschreibt eine Änderung in den geschätzten Werten durch Weglassen dieser Beobachtung. Um weitere Regressionen aufgrund der Löschung jeder Beobachtung zu vermeiden, wird eine Formel zur Berechnung der Regressionskoeffizienten nach Löschung der i -ten Beobachtung hergeleitet, welche abhängig von den basierend auf dem kompletten Datensatz ausgerechneten Regressionskoeffizienten und Residuen ist. Mit Hilfe dieser Formel lassen sich die Regressionsdiagnostiken herleiten und wiederum in einer vereinfachten Formel darstellen.

In Kapitel 4 und 5 werden hingegen die Generalisierten Linearen Modelle und darauf basierend die Regressionsdiagnostiken erklärt. Die wesentliche Basis der Generalisierten Modelle ist die Lineare Exponentielle Familie, die am Anfang von Kapitel 4 definiert wird. Man benutzt Ein-Schritt-Approximationen für die geschätzten Regressionskoeffizienten nach Löschung einer Beobachtung, welche es wiederum vermeiden, dass man zahlreiche Regressionen durchführen muss. Die Diagnostiken, welche mit Hilfe dieser Ein-Schritt-Approximation hergeleitet werden, werden in Kapitel 5 durch einen Datensatz, der anhand des Poissonmodells modelliert wird, veranschaulicht.

In Kapitel 6 wird die Theorie der Negativ-Binomial-Regression mit geschätztem α besprochen. Hier werden aufgrund von Ungenauigkeiten keine Ein-Schritt-Approximationen verwendet, sondern die exakten Werte aus der Schätzung. Dadurch müssen zuerst mehrere Regressionen durchgeführt werden, um die Diagnostiken zu berechnen. Es werden in Kapitel 7 für den Datensatz aus Kapitel 5 die Regressionsdiagnostiken für das Negativ-Binomial-Modell mit festem α und geschätztem α betrachtet. Abschließend wird innerhalb einer Fallstudie ein weiterer Datensatz mit Hilfe des Poissonmodells und des Negativ-Binomial-Modells mit festem und geschätztem α untersucht.

Kapitel 2

Lineare Modelle (LM)

Um einen ersten Einblick zu bekommen und den Einstieg verständlicher zu machen, bedient man sich zuerst der Linearen Modelle (LM). Eine übersichtliche Einführung für LM's wird z.B. in Myers (1990) gegeben. Es werden vereinfachte Annahmen getroffen, um die abhängige Variable, die sogenannte *Zielvariable*, durch unabhängige Variablen, die sogenannten *Kovariablen*, darzustellen. Entscheidend ist hierbei die statistische Modellbildung, welche zuerst die explorative Datenanalyse vorsieht. Diese kann u.a. ein geeignetes statistisches Modell vorschlagen. Danach wird das Modell angepasst und geeignete Hypothesen getestet. Als nächstes müssen die Modellannahmen mit Hilfe der Residualanalyse überprüft werden. Dies wird wesentlicher Bestandteil der Diagnostik sein. Es kann zur Annahme oder Veränderung des Modells kommen. Zuletzt werden die Modelle auf ihre Vorhersagefähigkeit überprüft und entsprechende Vorhersagen getroffen. Hier interessiert man sich besonders für die Überprüfung der Modellannahmen. Die Literatur von Myers (1990) und Draper und Smith (1998) bildet die Basis für dieses Kapitel.

2.1 Modellbeschreibung und Annahmen

Um eine Zielvariable in Abhängigkeit von gewissen Kovariablen anzugeben, ist ein geeignetes statistisches Modell zu bestimmen. Das klassische Lineare Modell lautet wie folgt:

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

wobei y_i die zu untersuchende Zielvariable ist, die $\mathbf{x}_i = (1, x_{i1}, \dots, x_{ik})^T$ die gegebenen Kovariablen, $\boldsymbol{\beta} = (\beta_0, \dots, \beta_k)$ die sogenannten *Regressionskoeffizienten* und ϵ_i die zufälligen Fehler sind. Das Modell ist linear in den Regressionskoeffizienten.

In Matrixschreibweise schreibt man das Modell wie folgt:

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_n^T \end{pmatrix} \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (2.1)$$

mit $\mathbf{Y} \in \mathbb{R}^n$, $\mathbf{X} \in \mathbb{R}^{n \times p}$, $\boldsymbol{\beta} \in \mathbb{R}^p$, $\boldsymbol{\epsilon} \in \mathbb{R}^n$, $p = k + 1$. Die Vektoren $\mathbf{x}_i = (1, x_{i1}, \dots, x_{ik})^T$ sind also Zeilenvektoren von \mathbf{X} , die die k Kovariablen angeben. Die Matrix \mathbf{X} wird als *Design-Matrix* bezeichnet.

Folgende Annahmen werden getroffen. Die Kovariablen \mathbf{x}_i sind fest gegeben und nichtzufällig, wohingegen die ϵ_i Zufallsvariablen sind, deren Erwartungswert Null und Varianz konstant σ^2

ist. Die Eigenschaft der konstanten Varianz wird als *Homoskedastizität* bezeichnet. Desweiteren nimmt man die ϵ_i als normalverteilt und unkorreliert an. Eine häufige Annahme bei Linearen Modellen stellt die Normalverteilungsannahme der Y_i dar, die aus der Normalverteilungsannahme der ϵ_i resultiert.

Da man hier Zufallsvektoren betrachtet, benötigt man die multivariate Normalverteilung.

Definition 2.1 (Multivariate Normalverteilung) Falls $\mathbf{Z} \in \mathbb{R}^m$ ein Zufallsvektor mit Erwartungswert $\boldsymbol{\mu}$, positiv definiten Kovarianzmatrix Σ und Dichte

$$f(\mathbf{z}) = f(z_1, \dots, z_m) = \frac{1}{(2\pi)^{m/2} \det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{z} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{z} - \boldsymbol{\mu})\right),$$

dann gilt:

$$\mathbf{Z} \sim N_m(\boldsymbol{\mu}, \Sigma).$$

Also kann man zusammenfassend das Lineare Modell als folgendes statistisches Modell aufstellen:

$$\mathbf{Y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2 I)$$

unter den folgenden Modell-Voraussetzungen:

1. \mathbf{x}_i , $i = 1, \dots, n$, fest und nichtzufällig,
2. $\epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$, $i = 1, \dots, n$.

Bei Generalisierten Linearen Modellen (GLM) wird die Normalverteilungsannahme der Y_i ausgeweitet, wie man noch in Kapitel 4 sehen wird.

2.2 Parameterschätzung in LM's

Man ist nun interessiert in der Schätzung der Parameter $\boldsymbol{\beta}$ und der Varianz σ^2 . In Linearen Modellen schätzt man entweder durch Maximieren des Likelihoods oder durch die Kleinste-Quadrate (KQ)-Methode. Die Kleinste-Quadrate-Methode zur Schätzung von $\boldsymbol{\beta}$ besteht darin die Summe der Fehlerquadrate zu minimieren, die durch

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

gegeben ist. Um einen Schätzer $\hat{\boldsymbol{\beta}}$ für $\boldsymbol{\beta}$ zu bekommen, wird dieser Ausdruck nach $\boldsymbol{\beta}$ differenziert und Null gesetzt:

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}} Q(\boldsymbol{\beta}) &= \frac{\partial}{\partial \boldsymbol{\beta}} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \stackrel{!}{=} \mathbf{0} \\ (\mathbf{X}^T \mathbf{X}) \boldsymbol{\beta} &\stackrel{!}{=} \mathbf{X}^T \mathbf{Y}. \end{aligned}$$

Also erfüllt der KQ-Schätzer $\hat{\boldsymbol{\beta}}$ die Normalengleichung

$$(\mathbf{X}^T \mathbf{X}) \hat{\boldsymbol{\beta}} = \mathbf{X}^T \mathbf{Y}. \quad (2.2)$$

Falls \mathbf{X} vollen Rang hat, gilt damit

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (2.3)$$

Hat \mathbf{X} nicht vollen Rang, so kann mit der generalisierten Inversen gerechnet werden oder man verwendet die Theorie der schätzbaren Funktionen (siehe Bickel und Doksum (1977)). Eine weitere Möglichkeit einen Schätzer $\hat{\boldsymbol{\beta}}$ für $\boldsymbol{\beta}$ zu bestimmen, ist die Maximum-Likelihood (ML)-Methode, mit welcher man dieselben Ergebnisse erhält. Auf diese wird in Kapitel 4 näher eingegangen, da bei Generalisierten Linearen Modellen nur die Maximum-Likelihood-Methode zur Anwendung kommt.

Die Abweichungen zwischen beobachteten y -Wert y_i und gefittetem y -Wert $\hat{y}_i := \mathbf{x}_i^T \hat{\boldsymbol{\beta}}$ bezeichnet man als *Residuum* e_i , d.h. $e_i := y_i - \hat{y}_i$. Damit lässt sich für jeden Datenpunkt überprüfen, wie gut er aufgrund des Modells vorhergesagt worden wäre. Um nun auch die Güte des Modells insgesamt beurteilen zu können, benötigt man die sogenannte *Streuungszerlegung*. Diese ist gegeben durch

$$\underbrace{\sum_{i=1}^n (y_i - \bar{y})^2}_{=:SQT} = \underbrace{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}_{=:SQE} + \underbrace{\sum_{i=1}^n (y_i - \hat{y}_i)^2}_{=:SQR}, \quad (2.4)$$

wobei $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$. *SQT* (*Sum of Squares Total*) bezeichnet dabei die gesamte Streuung, *SQR* (*Sum of Squares Residual*) die Residualstreuung und *SQE* (*Sum of Squares Explained*) die erklärte Streuung. Die obige Gleichung gilt, wie man durch einfache Rechenschritte nachvollziehen kann. Mit Hilfe von Gleichung (2.3) bekommt man $\hat{\mathbf{Y}} = \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$. Es gilt:

$$\begin{aligned} \hat{\mathbf{Y}}^T \hat{\mathbf{Y}} &= \mathbf{Y}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \\ &= \mathbf{Y}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} = \mathbf{Y}^T \hat{\mathbf{Y}}. \end{aligned} \quad (2.5)$$

Nach Differentiation von $Q(\boldsymbol{\beta})$ nach β_0 erhält man eine der Normalgleichungen für $\hat{\boldsymbol{\beta}}$:

$$\sum_{i=1}^n (y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) = 0$$

oder

$$\sum_{i=1}^n (y_i - \hat{y}_i) = 0. \quad (2.6)$$

Es gilt ferner

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

wobei der dritte Term der letzten Gleichung Null ist, da

$$\begin{aligned} \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) &\stackrel{(2.6)}{=} \sum_{i=1}^n (y_i - \hat{y}_i)\hat{y}_i \\ &= (\mathbf{Y} - \hat{\mathbf{Y}})^T \hat{\mathbf{Y}} \stackrel{(2.5)}{=} \mathbf{0}. \end{aligned}$$

Damit ist die Streuungszerlegung bewiesen. Also ist die Residuenquadratsumme $\sum_{i=1}^n e_i^2$, welche für $\beta = \hat{\beta}$ minimal ist, ein Maß wie gut die Variabilität der y_i durch die Regression erklärt ist. Als Maßzahl für die Güte des angepassten Modells verwendet man nun das sogenannte *Bestimmtheitsmaß*. Dies berechnet sich folgendermaßen:

$$R^2 := 1 - \frac{SQR}{SQT} = \frac{SQE}{SQT}.$$

Das Bestimmtheitsmaß R^2 , das zwischen 0 und 1 liegt, gibt den Anteil der Variation an, der durch die Regression erklärt wird. Da R^2 bei Hinzunahme weiterer Kovariablen wächst, lassen sich nur Modelle mit gleicher Anzahl von Parametern vergleichen. Deshalb verwendet man das *angepasste Bestimmtheitsmaß* R_{ang}^2 zum Vergleich von Modellen mit unterschiedlicher Anzahl an Parametern. Dabei wird SQR und SQT durch die jeweilige Anzahl der Freiheitsgrade geteilt, da mit jeder Hinzunahme einer Kovariable ein Freiheitsgrad verbraucht wird:

$$R_{ang}^2 := 1 - \frac{SQR/(n-p)}{SQT/(n-1)}.$$

Einen Schätzer für die Varianz σ^2 erhält man, wenn man die Residuenquadratsumme durch die Anzahl der Freiheitsgrade $n-p$ teilt. Somit lautet der Schätzer wie folgt:

$$s^2 := \frac{(\mathbf{Y} - \hat{\mathbf{Y}})^T (\mathbf{Y} - \hat{\mathbf{Y}})}{n-p} = \frac{SQR}{n-p}. \quad (2.7)$$

Dieser Schätzer ist für σ^2 *unverzerrt* bzw. *erwartungstreu*, d.h. der Erwartungswert der Schätzfunktion s^2 ist identisch mit dem zu schätzenden Parameter σ^2 . Diese und weitere Eigenschaften werden im Gauß-Markov-Theorem für Lineare Modelle nachgewiesen.

Satz 2.1 (Gauß-Markov) *Es gelte:*

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon, \quad \mathbf{Y}, \epsilon \in \mathbb{R}^n, \quad \mathbf{X} \in \mathbb{R}^{n \times p}, \quad \text{rang}(\mathbf{X}) = p, \quad E(\epsilon) = 0, \quad \text{Cov}(\epsilon) = \sigma^2 I$$

Dann folgt:

1. $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$
ist *bester linearer unverzerrter Schätzer (best linear unbiased estimator, BLUE)*.
2. $\text{Cov}(\hat{\beta}) = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}$
3. $s^2 = \frac{SQR}{n-p}$ ist *unverzerrter Schätzer für σ^2* .

Beweis: Ein Beweis dieses Satzes findet sich in Fahrmeir et al. (1996) (Kap.1, Seite 99). □

Somit ist der Erwartungswert und die Varianz des Schätzers $\hat{\beta}$ gegeben durch

$$E(\hat{\beta}) = E((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \underbrace{E(\mathbf{Y})}_{=\mathbf{x}\beta} = \beta,$$

$$Var(\hat{\beta}) = Var((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \underbrace{Var(\mathbf{Y})}_{=\sigma^2} ((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T)^T = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}.$$

Also ist der Schätzer $\hat{\beta}$ unverzerrt, der Erwartungswert der Schätzfunktion folglich wieder der zu schätzende Parameter selbst. Außerdem ist der Schätzer $\hat{\beta}$ effizient, d.h. unter allen Schätzern hat $\hat{\beta}$ die minimale Varianz. Wenn nun die Schätzfunktion linear in den beobachteten Variablen, effizient und unverzerrt ist, dann gilt die in Satz 2.1 aufgeführte *BLUE*-Eigenschaft. Für die Verteilung des Schätzers $\hat{\beta}$ gilt:

$$\hat{\beta} \sim N_p(\beta, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1})$$

Dies folgt aus der linearen Transformation von \mathbf{Y} und aus dem Gauß-Markov-Theorem.

2.3 Statistische Hypothesentests in LM's

In der Statistik werden sogenannte *Hypothesentests* durchgeführt. Die Grundlage hierzu ist, eine Entscheidung darüber zu treffen, ob eine aufgestellte Nullhypothese wahr oder falsch ist. Es werden immer zwei Entscheidungen zugelassen, die sich aus der aufgestellten Nullhypothese und der dazugehörigen Alternative ergeben. Dazu werden geeignete Teststatistiken T aufgestellt und danach entschieden, ob die Nullhypothese verworfen oder angenommen werden kann. Die Entscheidung basiert nun auf dem Wert der Zufallsvariable T mit Verteilung F_ψ , wobei F_ψ zu der Klasse $\mathcal{F} = \{F_\psi, \psi \in \Psi\}$ von Verteilungen gehört. Wenn ψ bekannt ist, wird angenommen, dass man weiß, ob die Nullhypothese wahr oder falsch ist. Falls ψ unbekannt ist, können die Verteilungen von Ψ in zwei Gruppen eingeteilt und mit H_0 und H_1 bezeichnet werden. Der Parameterraum Ψ wird ebenso unterteilt in Ψ_{H_0} und Ψ_{H_1} . Somit lautet das allgemeine Testproblem:

$$H_0 : \psi \in \Psi_{H_0} \quad \text{und} \quad H_1 : \psi \in \Psi_{H_1}.$$

Nun gilt es bei den Problemen, eine Entscheidung zu treffen. Eine Entscheidungsvorschrift wird durch eine Funktion δ gegeben, die jedem möglichen Wert t der Zufallsvariable T einen Wert $d = \delta(t)$ zuordnet. Die Funktion $d = \delta(t)$ wird nun so interpretiert, dass zu jedem Wert t von T eine Entscheidungsregel $\delta(t)$ angewendet wird, die zur tatsächlichen Entscheidung führt. Der Zustandsraum von T wird somit in zwei disjunkte Bereiche, den Annahmehbereich A und dem Ablehnungsbereich B , unterteilt. Fällt t in A , so wird die Hypothese akzeptiert, ansonsten wird sie abgelehnt.

Zum besseren Verständnis werden hier kurz die im folgenden benötigten Stichprobenverteilungen und Bezeichnungen vorgestellt. Eine detaillierte Beschreibung findet man in Fahrmeir et al. (2000) (Seite 298ff). Falls X_1, \dots, X_n unabhängige und identisch $N(0, 1)$ -verteilte Zufallsvariablen sind, so ist die Zufallsvariable $Z := X_1^2 + \dots + X_n^2$ χ^2 -verteilt mit n Freiheitsgraden, kurz $Z \sim \chi_n^2$. Falls die Zufallsvariablen $X \sim N(0, 1)$, $Z \sim \chi_n^2$ sowie X und Z unabhängig, dann ist

die Zufallsvariable $T := \frac{X}{\sqrt{Z/n}}$ t_n -verteilt, kurz $T \sim t_n$. Falls die Zufallsvariablen X und Y χ_m^2 - bzw. χ_n^2 -verteilt und unabhängig sind, so ist die Zufallsvariable $Z := \frac{X/m}{Y/n}$ F -verteilt mit Freiheitsgraden m und n , kurz $Z \sim F_{m,n}$.

Man definiert den Likelihood für Lineare Modelle mit $\psi = (\boldsymbol{\beta}, \sigma^2)$ als

$$l(\boldsymbol{\beta}, \sigma^2, \mathbf{y}) := \prod_{i=1}^n f(y_i; \boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - \mu_i)^2}{2\sigma^2}\right).$$

Wald-Test im Linearen Modell

Beim Wald-Test ist man daran interessiert, ob ein bestimmter Koeffizient gleich Null gesetzt werden kann. Dazu betrachtet man für festes j folgende Hypothese H_0 und Alternative H_1 :

$$H_0 : \beta_j = 0 \quad \text{gegen} \quad H_1 : \beta_j \neq 0.$$

Da $\hat{\beta}_j \sim N(\beta_j, \sigma^2(\mathbf{X}^T \mathbf{X})_{jj}^{-1})$ gilt, gilt dann unter H_0 :

$$t := \frac{\hat{\beta}_j}{\sqrt{s^2(\mathbf{X}^T \mathbf{X})_{jj}^{-1}}} = \frac{\hat{\beta}_j / \sigma \sqrt{(\mathbf{X}^T \mathbf{X})_{jj}^{-1}}}{\sqrt{\frac{(n-p)s^2}{\sigma^2} / (n-p)}} \sim t_{n-p}.$$

Da hier der Zähler $N(0, 1)$ -verteilt unter der Hypothese H_0 und der Zähler des Radikanten im Nenner χ_{n-p}^2 -verteilt und unabhängig von $\hat{\beta}_j$ ist, gilt obiges. Die verwendete Verteilungstheorie und entsprechende Beweise sind in Seber (1977) (Seite 54ff) angegeben. Somit lautet die Testregel wie folgt:

$$\text{Verwerfe } H_0 : \beta_j = 0 \text{ gegen } H_1 : \beta_j \neq 0 \text{ zum Niveau } \alpha \Leftrightarrow |t| > t_{n-p, 1-\alpha/2}.$$

Der Wert α bezeichnet hier das Signifikanzniveau und $t_{n-p, 1-\alpha/2}$ ist das $(1 - \alpha/2)$ -Quantil der t -Verteilung. Der Test ist exakt unter der Normalverteilungsannahme. Falls man nur $E(\epsilon_i) = 0 \quad \forall i$ und $Cov(\epsilon_i, \epsilon_j) = 0 \quad \forall i \neq j$ voraussetzen kann, dann gilt der Test nur asymptotisch, also wenn n groß ist. Der Grund liegt darin, dass der Nenner nur mehr asymptotisch χ_{n-p}^2 -verteilt ist.

Likelihood-Ratio-Test im Linearen Modell

Mit dem Likelihood-Ratio-Test versucht man nun mehrere Koeffizienten zu testen. Diese fasst man in einem Vektor $\boldsymbol{\beta}_2$ zusammen. Man startet also mit dem ursprünglichen Modell $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 I_n)$. Dann unterteilt man den Koeffizientenvektor $\boldsymbol{\beta}$ in $\boldsymbol{\beta}_1$ und $\boldsymbol{\beta}_2$, also $\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}$, $\boldsymbol{\beta}_1 \in \mathbb{R}^{p_1}$, $\boldsymbol{\beta}_2 \in \mathbb{R}^{p_2}$, $p_1 + p_2 = p$. Nun lässt sich wiederum die Hypothese und die Alternative aufstellen:

$$H_0 : \boldsymbol{\beta}_2 = \mathbf{0} \quad \text{gegen} \quad H_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}.$$

Nun stellt man die Likelihood-Ratio-Statistik auf, die wie folgt lautet:

$$L(\mathbf{y}) := \frac{\max_{\boldsymbol{\beta} \in H_0} l(\boldsymbol{\beta}, \sigma^2, \mathbf{y})}{\max_{\boldsymbol{\beta} \in H_1} l(\boldsymbol{\beta}, \sigma^2, \mathbf{y})},$$

wobei $l(\boldsymbol{\beta}, \sigma^2, \mathbf{y})$ der Likelihood des Linearen Modells $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ ist.
Die Entscheidungsregel lautet nun:

$$\text{Verwerfe } H_0 \text{ gegen } H_1 \Leftrightarrow L(\mathbf{y}) \leq k.$$

Für den Likelihood-Ratio $L(\mathbf{y})$ ergibt sich:

$$L(\mathbf{y}) = \frac{\prod_{i=1}^n \frac{1}{\sqrt{2\pi\hat{\sigma}(X_1)^2}} \exp\left(-\frac{1}{2\hat{\sigma}(X_1)^2} \|\mathbf{y} - \mathbf{X}_1\hat{\boldsymbol{\beta}}_1\|^2\right)}{\prod_{i=1}^n \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left(-\frac{1}{2\hat{\sigma}^2} \|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2\right)},$$

wobei $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$, $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$, $\hat{\sigma}^2(\mathbf{X}_1) = \frac{SQR(\mathbf{X}_1)}{n}$, $\hat{\sigma}^2 = \frac{SQR}{n}$, $SQR(\mathbf{X}_1) = \sum_{i=1}^n (Y_i - \mathbf{x}_{i1}^T \hat{\boldsymbol{\beta}}_1)^2 = \|\mathbf{y} - \mathbf{X}_1\hat{\boldsymbol{\beta}}_1\|^2$ und $SQR = \sum_{i=1}^n (Y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}})^2 = \|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2$.
Also gilt weiter:

$$L(\mathbf{y}) = \frac{\prod_{i=1}^n \frac{1}{\sqrt{2\pi\hat{\sigma}(X_1)^2}} \exp\left(-\frac{1}{2\hat{\sigma}(X_1)^2} \|\mathbf{y} - \mathbf{X}_1\hat{\boldsymbol{\beta}}_1\|^2\right)}{\prod_{i=1}^n \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left(-\frac{1}{2\hat{\sigma}^2} \|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2\right)} = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}(\mathbf{X}_1)^2}\right)^{n/2} = \left(\frac{SQR}{SQR(\mathbf{X}_1)}\right)^{n/2}.$$

Nun gilt also:

$$\begin{aligned} \text{Verwerfe } H_0 &\Leftrightarrow \frac{SQR}{SQR(\mathbf{X}_1)} \leq k^{2/n} \\ &\Leftrightarrow \frac{SQR(\mathbf{X}_1) - SQR}{SQR} \geq \underbrace{\frac{1}{k^{2/n}} - 1}_{=:k'} \\ &\Leftrightarrow \frac{n-p}{p-p_1} \frac{SQR(\mathbf{X}_1) - SQR}{SQR} \geq \underbrace{\frac{n-p}{p-p_1} k'}_{=:K} \\ &\Leftrightarrow F := \frac{(SQR(\mathbf{X}_1) - SQR)/p_2}{SQR/(n-p)} \geq K. \end{aligned}$$

Folgende Aussagen über die Verteilungstheorie werden in Seber (1977) zu Beginn des Kapitel 4 beschrieben. Da $\frac{SQR(\mathbf{X}_1) - SQR}{\sigma^2} \sim \chi_{p_2}^2$ und $\frac{SQR}{\sigma^2} \sim \chi_{n-p}^2$ unter H_0 und unabhängig sind, gilt:

$$F = \frac{\frac{(SQR(\mathbf{X}_1) - SQR)}{\sigma^2} / p_2}{\frac{SQR}{\sigma^2} / (n-p)} \sim F_{p_2, n-p} \quad \text{unter } H_0. \quad (2.8)$$

Die Testregel lautet also wie folgt:

$$\text{Verwerfe } H_0 : \boldsymbol{\beta}_2 = 0 \text{ gegen } H_1 : \boldsymbol{\beta}_2 \neq 0 \text{ zum Niveau } \alpha \Leftrightarrow F > F_{p_2, n-p, 1-\alpha}.$$

Der Zähler der Teststatistik in (2.8) misst die Verringerung der SQR , falls man die erklärenden Variablen in \mathbf{X}_2 dazunimmt, pro verwendeten Freiheitsgrad. Der Nenner hingegen ist ein Maß

für die Fehlerstreuung im Modell. Ein F -Wert von 1 bedeutet, dass die Variablen in \mathbf{X}_2 nur Fehlerstreuung erzeugen. Ein Wert größer 1 zeigt hingegen das Vorhandensein eines Signals, was bei signifikanter Größe des F -Wertes eine Zunahme der Kovariablen $\mathbf{x}_{p_1+1}, \dots, \mathbf{x}_{p_2}$ zur Folge hat.

2.4 Gewichtete Regression

Ein gewichtetes Regressionsmodell wird angewandt, wenn man von vornherein schon weiß, dass nicht alle Beobachtungen den gleichen Beitrag liefern oder wenn man verschiedenen Messstellen verschiedenen Informationswert zukommen lassen will. Letzteres spielt z.B. in der Zeitreihenanalyse unter dem Begriff *Exponentielles Glätten* eine wichtige Rolle. Das gewichtete Regressionsmodell ist gegeben durch (2.1), also

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (2.9)$$

wobei alles definiert ist wie in (2.1), außer dass $\text{Var}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{W}^{-1}$ mit $\mathbf{W} = \text{diag}(w_{ii})$, $i = 1, \dots, n$, bekannte Diagonalmatrix, $\mathbf{W} \in \mathbb{R}^{n \times n}$, $w_{ii} > 0$. Die Matrix \mathbf{W} ist also positiv definit. Man bezeichnet die nichtkonstante Varianzannahme als *Heteroskedastizität*, im Gegensatz zur Homoskedastizität in Kapitel 2.1.

Man geht also davon aus, dass Heteroskedastizität vorliegt und $\text{Var}(Y_i) = \frac{1}{w_{ii}}\sigma^2$ mit w_{ii} bekannt ist. Obwohl gewichtete Schätzer direkt ausgerechnet werden können, ist es üblich, in ein ungewichtetes Problem zu transformieren. Beide Seiten von (2.9) werden von links mit $\mathbf{W}^{1/2}$ multipliziert, d.h.

$$\mathbf{W}^{1/2}\mathbf{Y} = \mathbf{W}^{1/2}\mathbf{X}\boldsymbol{\beta} + \mathbf{W}^{1/2}\boldsymbol{\epsilon}. \quad (2.10)$$

Da $\text{Var}(\mathbf{W}^{1/2}\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}$, folgt nach Kapitel 2.2 für die Fehlerquadratsumme

$$\begin{aligned} Q(\boldsymbol{\beta}) &= (\mathbf{W}^{1/2}\mathbf{Y} - \mathbf{W}^{1/2}\mathbf{X}\boldsymbol{\beta})^T (\mathbf{W}^{1/2}\mathbf{Y} - \mathbf{W}^{1/2}\mathbf{X}\boldsymbol{\beta}) \\ &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{W} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \sum_{i=1}^n w_{ii} (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 \\ &= \sum_{i=1}^n w_{ii} \epsilon_i^2. \end{aligned} \quad (2.11)$$

Durch Ableiten nach $\boldsymbol{\beta}$ erhält man ein Gleichungssystem, das in Matrixnotation folgende Form hat:

$$\mathbf{X}^T \mathbf{W} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0}.$$

Löst man nun nach $\boldsymbol{\beta}$ auf, so erhält man den Kleinste-Quadrate-Schätzer $\hat{\boldsymbol{\beta}}(\mathbf{w})$ für $\boldsymbol{\beta}$:

$$\begin{aligned} \mathbf{X}^T \mathbf{W} \mathbf{Y} - \mathbf{X}^T \mathbf{W} \mathbf{X} \boldsymbol{\beta} &\stackrel{!}{=} \mathbf{0} \\ \mathbf{X}^T \mathbf{W} \mathbf{X} \boldsymbol{\beta} &\stackrel{!}{=} \mathbf{X}^T \mathbf{W} \mathbf{Y}. \end{aligned}$$

Der Parameter $\hat{\boldsymbol{\beta}}$ löst die Normalgleichung

$$\mathbf{X}^T \mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X}^T \mathbf{W} \mathbf{Y} \quad (2.12)$$

und der Schätzer lautet somit wie folgt:

$$\hat{\boldsymbol{\beta}}(\mathbf{w}) = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{Y}. \quad (2.13)$$

Mit $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{W}^{-1})$ ist der Erwartungswert und die Varianz des Schätzers $\hat{\boldsymbol{\beta}}(\mathbf{w})$ gegeben durch

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}(\mathbf{w})) &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \underbrace{E(\mathbf{Y})}_{\mathbf{X}\boldsymbol{\beta}} = \boldsymbol{\beta}, \\ \text{Var}(\hat{\boldsymbol{\beta}}(\mathbf{w})) &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \underbrace{\text{Var}(\mathbf{Y})}_{=\sigma^2 \mathbf{W}^{-1}} ((\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W})^T = \sigma^2 (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}. \end{aligned}$$

Der Schätzer $\hat{\boldsymbol{\beta}}(\mathbf{w})$ ist somit unverzerrt. Ebenso hat der Schätzer $\hat{\boldsymbol{\beta}}(\mathbf{w})$ minimale Varianz unter allen unverzerrten Schätzern, was hier kurz gezeigt wird. Sei $\tilde{\boldsymbol{\beta}}(\mathbf{w})$ ein unverzerrter Schätzer der Form $\tilde{\boldsymbol{\beta}}(\mathbf{w}) = \mathbf{A}\mathbf{Y}$ mit $E(\tilde{\boldsymbol{\beta}}(\mathbf{w})) = \boldsymbol{\beta}$ und $\mathbf{A} \in \mathbb{R}^{p \times n}$. Damit gilt

$$E(\tilde{\boldsymbol{\beta}}(\mathbf{w})) = \mathbf{A}E(\mathbf{Y}) = \mathbf{A}\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}.$$

Also muss $\mathbf{A}\mathbf{X} = I_p$ gelten. Weiter gilt

$$\text{Var}(\tilde{\boldsymbol{\beta}}(\mathbf{w})) = \text{Var}(\mathbf{A}\mathbf{Y}) = \text{Var}(\mathbf{A}\boldsymbol{\epsilon}) = \sigma^2 \mathbf{A}\mathbf{W}^{-1} \mathbf{A}^T.$$

Da $\text{Var}(\hat{\boldsymbol{\beta}}(\mathbf{w})) = \sigma^2 (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}$, genügt es zu zeigen, dass $\mathbf{A}\mathbf{W}^{-1} \mathbf{A}^T - (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} =: B \in \mathbb{R}^{p \times p}$ positiv definit ist. Eine Matrix ist nach Rade und Westergren (1997) (Seite 100) positiv definit, falls alle Unterdeterminanten positive Werte besitzen. Um dies zu zeigen, wird die Matrix B umgeschrieben in

$$B = \underbrace{(\mathbf{A}\mathbf{W}^{-1/2} - (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2})}_{=: C} (\mathbf{A}\mathbf{W}^{-1/2} - (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2})^T,$$

was aufgrund \mathbf{W} Diagonalmatrix möglich ist. Nach Rade und Westergren (1997) (Seite 90) gelten für zwei beliebige Matrizen $M, N \in \mathbb{R}^{k \times l}$ folgende Determinantenoperationen:

$$\det(MN) = \det(M)\det(N), \quad \det(M) = \det(M^T)$$

Also gilt:

$$\det(B) = \det(CC^T) = \det(C)\det(C^T) = \det(C)\det(C) = (\det(C))^2 > 0$$

Letzterer Schritt lässt sich für jede Unterdeterminante durchführen und somit ist die Effizienz des Schätzers gezeigt.

Kapitel 3

Regressionsdiagnostiken für LM's

In diesem Kapitel sollen die wichtigsten Regressionsdiagnostiken für Lineare Modelle erläutert werden, die die Grundlage für die auf im Kapitel 4 ausgeweiteten Generalisierten Modelle sind. Die Resultate statistischer Beobachtungen können in besonderem Maße von einzelnen Beobachtungen abhängen. Hierbei kann es zur Verzerrung der Analyseergebnisse kommen. Deshalb sind Beobachtungen, die nicht dem Trend folgen oder weit weg von den restlichen Beobachtungen liegen, geeignet zu identifizieren und in ihrem Einfluß auf Regressionskoeffizienten und vorhergesagte Werte zu untersuchen. Die Diagnostiken helfen, sogenannte *Ausreißer* oder *einflußreiche Beobachtungen* zu entdecken. Die Regression kann also mit und ohne auffallende Beobachtung durchgeführt werden. Sind die Unterschiede in den Ergebnissen 'groß', gilt es, diesen Einfluß zu erklären und gegebenenfalls einzuschränken. In Belsley et al. (1980) sind einige Diagnostiken für LM's angegeben.

Diese Punkte, welche nicht dem üblichen Trend folgen, können aus einer Vielzahl von Gründen entstehen. Zumeist sind Verletzung der Modellannahmen die Ursache. Daneben sind noch weitere Gründe denkbar, wie z.B. falsche Erhebung oder Falscheintragung der Daten. Hierbei spielen die Residuen, welche in diesem Kapitel vorgestellt werden sollen, eine wichtige Rolle. Diese sind ein erstes Hilfsmittel, um auffällige Beobachtungen aufzuspüren. Durch Gewichtung oder Weglassen von entsprechenden Punkten ändern sich die Ergebnisse zumeist wesentlich. Gerade deshalb ist hierauf genaues Augenmerk zu legen und es sollten geeignete Diagnostiken entwickelt werden, um die Identifizierung zu gewährleisten. Die Kapitel 4 und 5 aus dem Buch von Myers (1990) geben Auskunft über die Residuenanalyse und geeignete Diagnostiken.

3.1 y-Ausreißer und x-Ausreißer

Unter diesen auffallenden Punkten treten Beobachtungen auf, die einen extrem hohen bzw. niedrigen y-Wert haben können im Gegensatz zu den meisten anderen Punkten für einen bestimmten x-Wertebereich. Diese werden als *y-Ausreißer* bezeichnet. Ein sehr markantes Kennzeichen für einen y-Ausreißer ist die weite Entfernung von der Regressionsgerade. Im Gegensatz dazu gibt es Punkte mit einem extrem hohen bzw. niedrigen x-Wert. Folglich bezeichnet man diese Beobachtungen als *x-Ausreißer*. Die Bezeichnung *high leverage* Punkt wird ebenso gebraucht. Solche Punkte müssen nicht unbedingt weit von der Regressionsgerade entfernt sein, jedoch sind sie meist weit weg vom Datenzentrum. Damit können Ausreißer sowohl im Response als auch in den erklärenden Variablen auftreten. Die Ursachen dafür können in einer Verletzung der Modellannahmen liegen. Es kann ein Fehler im Modell auftreten, der die Positionsverschiebung hervorruft, d.h. $E(\epsilon_i) = \Delta_i \neq 0$. Dies fällt unter den Bereich des *Mean Shift Outlier Model*, wel-

ches in Kapitel 3.4 beschrieben werden soll. Es kann aber auch sein, dass die Varianz des i -ten Punkts $Var(\epsilon_i)$ die Varianz anderer Datenpunkte überschreitet. Eine falsche Messung oder eine Falscheintragung der Daten ist darüberhinaus ebenso denkbar. Die y - und x -Ausreißer können starken Einfluß auf die Regressionsresultate haben, müssen aber nicht.

Jedoch können Ausreißer auch wichtige neue Informationen enthalten. Es kann der Fall sein, dass eine wichtige Kovariable nicht in das Modell aufgenommen wurde, mit Hilfe derer die Daten besser angepasst worden wären. Ausreißer können aber auch ein Indiz für ein falsch gewähltes Modell sein. Eine Möglichkeit zur Verbesserung der Datenanpassung ist, die Skala des Response zu transformieren. Deshalb ist bei der Untersuchung und Diagnostik von Ausreißern mit Vorsicht vorzugehen.

3.2 Einflußreiche Beobachtungen

Der mögliche Einfluß von Ausreißern auf Regressionsresultate wurde bereits angesprochen. Entscheidend ist nun, *einflußreiche Beobachtungen* herauszufiltern. Eine einflußreiche Beobachtung hat größere Auswirkungen auf die geschätzten Werte, die Regressionskoeffizienten und die Anpassungsgüte als die übrigen Datenpunkte. Zumeist sind diese Beobachtungen weiter entfernt von den restlichen Daten. Darüberhinaus müssen einflußreiche Beobachtungen nicht weit entfernt von der Regressionsgerade sein. Eine Veranschaulichung liefert die Abbildung 3.1. Die obere Grafik in Abbildung 3.1 zeigt einen y -Ausreißer, der die Regressionsgerade bis auf eine geringe y -Achsenverschiebung nicht wesentlich verändert. Die beiden Geraden zeigen denselben Verlauf. Die mittlere Grafik zeigt einen einzelnen isolierten Punkt, den man sowohl als x - als auch als y -Ausreißer bezeichnen könnte. Dieser Punkt bestätigt aber hier den Trend der Daten und ist deshalb nicht als einflußreich betrachtbar. Die beiden Geraden zeigen wiederum denselben Verlauf. Hingegen zeigt die untere Grafik einen x -Ausreißer, der die Regressionsgerade sehr wohl beeinflusst. Wie man aus der Grafik sieht, ist der Abstand zur Regressionsgerade nicht auffallend hoch.

3.3 Residuen und Hatmatrix

Die Aufdeckung von Missspezifikationen des Modells und Annahmeverletzungen kann durch die Residuen gewährleistet werden. Die Residuen messen wie gut \mathbf{Y} durch $\hat{\mathbf{Y}}$ angepasst wird. Wie in Kapitel 2.2 definiert lautet der Residuenvektor:

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{Y} - \underbrace{\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T}_{=:H}\mathbf{Y} =: (\mathbf{I} - H)\mathbf{Y}. \quad (3.1)$$

mit $H = (h_{ij})$, $i = 1, \dots, n$, $j = 1, \dots, n$. Die Matrix H bezeichnet die *Hatmatrix*. Diese besitzt einige hilfreiche Eigenschaften, die im folgenden Lemma aufgelistet werden.

Lemma 3.1 (Eigenschaften von H) Sei die Hatmatrix $H \in \mathbb{R}^{n \times n}$ definiert durch $H = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$. Dann ist H symmetrisch ($H^T = H$) und idempotent ($H^2 = H$). Außerdem gilt:

$$\sum_{i=1}^n h_{ii} = p, \quad 1/n \leq h_{ii} \leq 1.$$

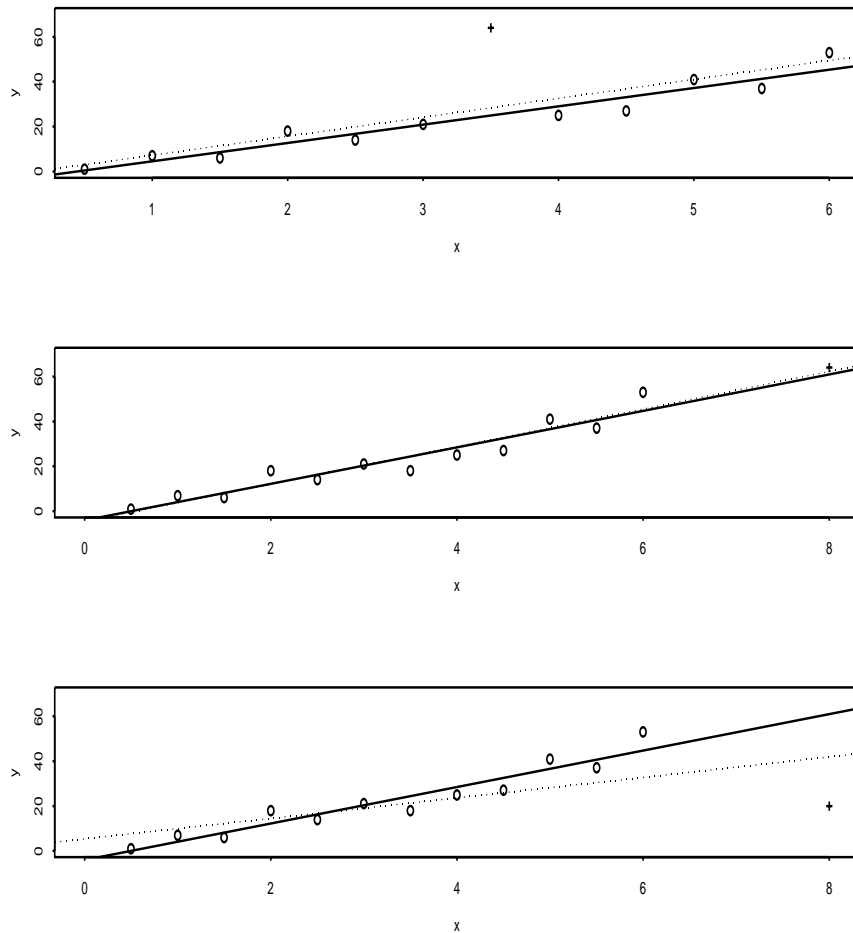


Abbildung 3.1: Der Punkt + ist ein y-Ausreißer (oben), x- bzw. y-Ausreißer (mitte) und x-Ausreißer (unten). Die gestrichelte Linie ist Regressionsgerade basierend auf den gesamten Daten, die durchgezogene Linie bei Löschung des Ausreißers

Beweis:

$$H^T = (\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)^T = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = H,$$

$$H^2 = (\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)^T = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = H.$$

Um $\sum_{i=1}^n h_{ii} = p$ zu beweisen, wird die Spur einer Matrix benutzt. Die Definition der Spur einer Matrix und die wichtigsten Rechenregeln sind in Rade und Westergren (1997) (Seite 89) nachzulesen. Es gilt:

$$Spur(H) = Spur(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T) = Spur((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}) = Spur(I_p) = p,$$

wobei I_p die $p \times p$ Einheitsmatrix ist. Da für eine quadratische Matrix die Spur der Summe der Diagonalelemente entspricht, gilt:

$$Spur(H) = \sum_{i=1}^n h_{ii} = p.$$

Nun soll \mathbf{X} in zwei Komponenten zerlegt werden, in den Einsvektor der Länge n und die Matrix $\mathbf{Z} \in \mathbb{R}^{n \times (p-1)}$, wobei $\mathbf{z}_i^T = (\mathbf{x}_i - \bar{\mathbf{x}})^T = (x_{1i} - \bar{x}_1, \dots, x_{ki} - \bar{x}_k)$ den i -ten Zeilenvektor darstellt. Die Matrix \mathbf{Z} wird also durch Zentrieren der Kovariablen erhalten. Dann folgt

$$\hat{\mathbf{y}} - \bar{\mathbf{y}} = H\mathbf{y} - \bar{\mathbf{y}} =: H^z\mathbf{y}$$

und deswegen folgt für die Elemente der zentrierten Hatmatrix H^z

$$h_{ij}^z = h_{ij} - \frac{1}{n}.$$

Es folgt also ebenso:

$$h_{ii} = 1/n + \underbrace{\mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T}_{=H^z}.$$

Für die einfache lineare Regression $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ gilt daher $h_{ii} = 1/n + \frac{(x_i - \bar{x})^2}{\sum_{j=1}^n (x_j - \bar{x})^2}$. Es folgt also aus letzterer Gleichung

$$h_{ii} \geq 1/n.$$

Da die Hatmatrix idempotent und symmetrisch ist, gilt auch

$$h_{ij} = h_{ji}, \tag{3.2}$$

$$h_{ij} = \sum_{k=1}^n h_{ik}h_{kj}, \tag{3.3}$$

$$h_{ii} = \sum_{j=1}^n h_{ij}h_{ji} = \sum_{j=1}^n h_{ij}h_{ij} = \sum_{j=1}^n h_{ij}^2 = h_{ii}^2 + \sum_{j=1, j \neq i}^n h_{ij}^2 \geq h_{ii}^2. \tag{3.4}$$

Also gilt nach (3.4) auch

$$h_{ii} \leq 1.$$

□

Wichtige Eigenschaften von H werden in Myers (1990) und Fox (1984) aufgelistet. Mit H idempotent, ist auch $(I - H)$ idempotent. Die Matrix H wird auch als Projektionsmatrix bezeichnet. Dies ist eine Matrix, die idempotent und symmetrisch ist. Sie projiziert \mathbf{Y} auf den Raum der Regressionsebene. Das Diagonalelement h_{ii} von H ist ein Maß für den Abstand zwischen \mathbf{x}_i und $\bar{\mathbf{x}}$. Da $\sum_{i=1}^n h_{ii} = p$, ist der Durchschnitt p/n . Der i -te Punkt wird als *high leverage* Punkt bezeichnet, falls $h_{ii} > 2p/n$ ist. In Belsley et al. (1980) wird diese Schranke angeführt, welche auf einem F-Wert basiert, der in zuletzt genanntem Buch begründet wird. Dazu betrachtet man Wilks' Λ Statistik (Rao (1973) (Seite 570)) für zwei Gruppen, wobei eine aus einem einzigen Punkt besteht. Diese Statistik lautet

$$\Lambda(\mathbf{z}_i) = \frac{\det(\mathbf{Z}^T\mathbf{Z} - (n-1)\bar{\mathbf{z}}_{-i}^T\bar{\mathbf{z}}_{-i} - \mathbf{z}_i^T\mathbf{z}_i)}{\det(\mathbf{Z}^T\mathbf{Z})}.$$

Der Zähler kann als

$$\det\left(\mathbf{Z}^T\mathbf{Z} - \frac{n^2}{n-1}\left(\bar{\mathbf{z}} - \frac{\mathbf{z}_i}{n}\right)^T\left(\bar{\mathbf{z}} - \frac{\mathbf{z}_i}{n}\right) - \mathbf{z}_i^T\mathbf{z}_i\right) = \det\left(\mathbf{Z}^T\mathbf{Z} - \frac{n}{n-1}\mathbf{z}_i^T\mathbf{z}_i\right)$$

geschrieben werden, da \mathbf{Z} zentriert wurde. Zur weiteren Vorgehensweise wird das folgende Lemma benötigt.

Lemma 3.2 Sei $\mathbf{X} \in \mathbb{R}^{n \times p}$, $\mathbf{X}_{-i} \in \mathbb{R}^{(n-1) \times p}$ und $H = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$. Dann gilt:

$$\det(\mathbf{X}_{-i}^T \mathbf{X}_{-i}) = (1 - h_{ii}) \det(\mathbf{X}^T \mathbf{X}).$$

Beweis: Zuerst wird gezeigt, dass

$$\det(I - \mathbf{u} \mathbf{v}^T) = 1 - \mathbf{v}^T \mathbf{u},$$

wobei \mathbf{u} und \mathbf{v} entsprechende Zeilenvektoren sind. Sei \mathbf{Q} eine orthonormale Matrix, so dass

$$\mathbf{Q} \mathbf{u} = \|\mathbf{u}\| \boldsymbol{\xi}_1, \quad (3.5)$$

wobei $\boldsymbol{\xi}_1 = (1, 0, \dots, 0)$ der erste Einheitsvektor ist. Dann folgt:

$$\begin{aligned} \det(I - \mathbf{u} \mathbf{v}^T) &= \det(\mathbf{Q}(I - \mathbf{u} \mathbf{v}^T) \mathbf{Q}^T) = \det(I - \|\mathbf{u}\| \boldsymbol{\xi}_1 \mathbf{v}^T \mathbf{Q}^T) \\ &= 1 - \mathbf{v}^T \mathbf{Q}^T \boldsymbol{\xi}_1 \|\mathbf{u}\| \stackrel{(3.5)}{=} 1 - \mathbf{v}^T \mathbf{u}. \end{aligned}$$

Mit $\mathbf{u} = \mathbf{x}_i^T$ und $\mathbf{v} = \mathbf{x}_i (\mathbf{X}^T \mathbf{X})^{-1}$ folgt die Behauptung. □

Wendet man nun das Lemma 3.2 an, so lässt sich letzterer Ausdruck wie folgt darstellen:

$$\det\left(\mathbf{Z}^T \mathbf{Z} - \frac{n}{n-1} \mathbf{z}_i^T \mathbf{z}_i\right) = \left(1 - \frac{n}{n-1} \mathbf{z}_i (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{z}_i^T\right) \det(\mathbf{Z}^T \mathbf{Z}).$$

Somit gilt

$$\begin{aligned} \Lambda(\mathbf{z}_i) &= \frac{\det(\mathbf{Z}^T \mathbf{Z} - (n-1) \bar{\mathbf{z}}_{-i}^T \bar{\mathbf{z}}_{-i} - \mathbf{z}_i^T \mathbf{z}_i)}{\det(\mathbf{Z}^T \mathbf{Z})} \\ &= \frac{\left(1 - \frac{n}{n-1} \mathbf{z}_i (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{z}_i^T\right) \det(\mathbf{Z}^T \mathbf{Z})}{\det(\mathbf{Z}^T \mathbf{Z})} \\ &= 1 - \frac{n}{n-1} h_{ii}^z \\ &= \frac{n}{n-1} (1 - h_{ii}) \end{aligned}$$

Falls nun angenommen wird, dass die Reihen von \mathbf{Z} nach einer $(p-1)$ -dimensionalen Normalverteilung unabhängig identisch verteilt sind, dann gilt nach Rao (1973) (Seite 570)

$$\frac{n-p}{p-1} \left(\frac{1 - \Lambda(\mathbf{z}_i)}{\Lambda(\mathbf{z}_i)} \right) \sim F_{p-1, n-p}$$

und es folgt damit

$$\frac{n-p}{p-1} \left(\frac{h_{ii} - 1/n}{1 - h_{ii}} \right) \sim F_{p-1, n-p}.$$

Für $p > 10$ und $n - p > 50$ ist das 95%-Quantil für die F-Verteilung kleiner als 2 und deshalb ist $2p/n$ eine angemessene Schranke. Falls $p/n > 0.4$, gibt es sehr wenige Freiheitsgrade, so dass alle Beobachtungen verdächtig werden. Für kleine p zeigt der Wert $2p/n$ eventuell zu viele Beobachtungen an. Deswegen sollten bei einer Analyse auch die Werte von p und n berücksichtigt werden.

Falls $h_{ii} = 1$, gilt $\hat{y}_i = y_i$, was leicht durch Betrachten von $e_i = y_i - \hat{y}_i = (1 - h_{ii})y_i$ gezeigt werden kann. Letztere Eigenschaft aus Lemma 3.1 impliziert

$$1/n \leq \frac{\text{Var}(\hat{y}_i)}{\sigma^2} \leq 1,$$

da

$$\text{Var}(\hat{y}_i) = \text{Var}(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}) = \mathbf{x}_i^T \text{Var}(\hat{\boldsymbol{\beta}}) \mathbf{x}_i = \sigma^2 \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i = \sigma^2 h_{ii}.$$

Zum einen ist also die Genauigkeit einer Vorhersage eines Datenpunkts nicht schlechter als die Varianz in einer Beobachtung, d.h. $\text{Var}(\hat{y}_i) = \sigma^2$. Zweitens kann die Genauigkeit einer Vorhersage nicht besser sein als die Präzision des Durchschnitts σ^2/n .

Da die gewöhnlichen Residuen i.a. unterschiedliche Varianz haben und vom Diagonalelement h_{ii} der Hatmatrix abhängig sind, werden die Residuen standardisiert. Dies wird für das Definieren späterer Diagnostiken von Vorteil sein. Außerdem werden Schwierigkeiten besser durch Grafiken von standardisierten Residuen angezeigt. Um die Standardisierung der Residuen durchzuführen, wird der Erwartungswert und die Varianz benötigt, die im folgenden berechnet werden:

$$E(\mathbf{e}) = E((I - H)\mathbf{Y}) = (I - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T) \underbrace{E(\mathbf{Y})}_{=\mathbf{X}\boldsymbol{\beta}} = 0, \quad (3.6)$$

$$\text{Var}(\mathbf{e}) = \text{Var}((I - H)\mathbf{Y}) = (I - H) \underbrace{\text{Var}(\mathbf{Y})}_{=\sigma^2} (I - H)^T = (I - H)\sigma^2. \quad (3.7)$$

Standardisierte Residuen $\frac{e_i}{\sigma\sqrt{1-h_{ii}}}$ haben daher einen Erwartungswert von 0 und eine Varianz von 1. Zunächst werden die *intern standardisierten* Residuen definiert. Der Begriff 'intern' bedeutet, dass die Statistiken e_i und s auf Basis derselben Daten hergeleitet wurden und abhängig sind. Die intern standardisierten Residuen sind gegeben durch

$$r_i := \frac{e_i}{s\sqrt{1-h_{ii}}}, \quad i = 1, \dots, n, \quad (3.8)$$

wobei s^2 den Schätzer der Varianz σ^2 bezeichnet, der in (2.7) definiert wurde. Dadurch wird der Effekt der Lage des Datenpunkts im Regressorraum eliminiert. Da sich r_i als

$$r_i = \frac{e_i/\sigma\sqrt{1-h_{ii}}}{\sqrt{\frac{(n-p)s^2}{(n-p)\sigma^2}}}$$

schreiben lässt, gilt nach der Verteilungstheorie von Seber (1977), dass der Zähler standardnormalverteilt und der Nenner χ_{n-p}^2 -verteilt ist. Somit besitzt das standardisierte Residuum eine t_{n-p} -Verteilung und ein kritischer Wert kann wegen des 95%-Niveaus durch 2 angegeben werden, falls $n - p$ groß ist und sich damit die t -Verteilung der Normalverteilung nähert.

Da die Residuen in die Schätzung von σ^2 eingehen, kann man die Schätzung ohne die i -te Beobachtung durchführen. Man möchte daraus schließen, ob die i -te Beobachtung großen Einfluß auf die geschätzten Werte hat. Man sieht sich also ein Modell an, bei dem die i -te Beobachtung herausgenommen wurde. Dies definiert man wie folgt:

$$y_{k,-i} = \mathbf{x}_{\mathbf{k},-i}^T \boldsymbol{\beta}_{-i} + \epsilon_{-i}, \quad k = 1, \dots, n, \quad k \neq i. \quad (3.9)$$

Der Parameterschätzer $\hat{\beta}_{-i}$ wird ohne i -te Beobachtung und wie in Kapitel 2.2 beschrieben berechnet, d.h.

$$\hat{\beta}_{-i} = (\mathbf{X}_{-i}^T \mathbf{X}_{-i})^{-1} \mathbf{X}_{-i}^T \mathbf{Y}_{-i}, \quad (3.10)$$

wobei $\mathbf{X}_{-i} \in \mathbb{R}^{(n-1) \times p}$ und $\mathbf{Y}_{-i} \in \mathbb{R}^{n-1}$ die Matrix bzw. der Vektor mit entfernter i -ten Zeile bzw. Komponente. Also gilt für die Vorhersage der i -ten Beobachtung folgendes:

$$\hat{y}_{i,-i} := \mathbf{x}_i^T \hat{\beta}_{-i}.$$

Da im Modell ohne i -te Beobachtung die geschätzte Varianz nicht gleich s^2 ist, werden neue Residuen definiert, die sogenannten *Jack-knifed* oder *extern standardisierten* Residuen. Hier ist die Statistik e_i und der Schätzer für σ unabhängig, was den Begriff 'extern' erklärt. Die extern standardisierten Residuen sind gegeben durch

$$t_i = \frac{e_i}{s_{-i} \sqrt{1 - h_{ii}}}, \quad i = 1, \dots, n, \quad (3.11)$$

wobei s_{-i}^2 der Schätzer für σ^2 im Modell ohne i -te Beobachtung ist. Hier kann ebenso die t -Verteilung als Maßstab herangezogen werden, jedoch nun mit $n - p - 1$ Freiheitsgraden statt $n - p$ wegen Löschung des i -ten Punkts. Eine signifikante Schranke bleibt der Wert 2. Der Schätzer s_{-i}^2 ist gegeben durch

$$s_{-i}^2 = \frac{(n - p)s^2 - \frac{e_i^2}{1 - h_{ii}}}{n - p - 1} \quad (3.12)$$

$$= s^2 \left(\frac{n - p - r_i^2}{n - p - 1} \right). \quad (3.13)$$

Um (3.12) zu zeigen, wird folgendes Lemma benötigt:

Lemma 3.3 *Seien \mathbf{A} eine nicht-singuläre Matrix und \mathbf{u} und \mathbf{v} zwei Spaltenvektoren. Dann gilt:*

$$(\mathbf{A} - \mathbf{u}\mathbf{v}^T)^{-1} = \mathbf{A}^{-1} + \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1 - \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}.$$

Beweis: Ein Beweis dieses Lemmas findet sich in Myers (1990) (Appendix A, Seite 459). □

Speziell hier wählt man $\mathbf{A} := \mathbf{X}^T \mathbf{X}$ und $\mathbf{u} = \mathbf{v} := \mathbf{x}_i$. Man erhält somit:

$$(\mathbf{X}^T \mathbf{X} - \mathbf{x}_i \mathbf{x}_i^T)^{-1} = (\mathbf{X}^T \mathbf{X})^{-1} + \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1}}{1 - \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i} \quad (3.14)$$

$$(\mathbf{X}_{-i}^T \mathbf{X}_{-i})^{-1} = (\mathbf{X}^T \mathbf{X})^{-1} + \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1}}{1 - h_{ii}}. \quad (3.15)$$

Die Gleichheit in (3.14) lässt sich durch einfaches Einsetzen und Ausrechnen leicht herleiten. Da $\mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \neq 1$, existiert die Inverse der linken Seite:

$$(\mathbf{X}^T \mathbf{X} - \mathbf{x}_i \mathbf{x}_i^T)(\mathbf{X}^T \mathbf{X} - \mathbf{x}_i \mathbf{x}_i^T)^{-1} = (\mathbf{X}^T \mathbf{X} - \mathbf{x}_i \mathbf{x}_i^T) \left((\mathbf{X}^T \mathbf{X})^{-1} + \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1}}{1 - \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i} \right)$$

$$\begin{aligned}
I &= (\mathbf{X}^T \mathbf{X})(\mathbf{X}^T \mathbf{X})^{-1} - \mathbf{x}_i \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} + \frac{(\mathbf{X}^T \mathbf{X})(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1}}{1 - \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i} \\
&\quad - \frac{\mathbf{x}_i \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1}}{1 - \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i} \\
I &= I - \mathbf{x}_i \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} + \frac{\mathbf{x}_i \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} - \mathbf{x}_i \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1}}{1 - \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i} \\
I &= I - \mathbf{x}_i \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} + \mathbf{x}_i \left(\frac{1 - \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i}{1 - \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i} \right) \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \\
I &= I - \mathbf{x}_i \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} + \mathbf{x}_i \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \\
I &= I.
\end{aligned}$$

Für ein weiteres Vorgehen wird gezeigt, dass $\mathbf{X}^T \mathbf{Y} - \mathbf{x}_i y_i = \mathbf{X}_{-i}^T \mathbf{Y}_{-i}$:

$$\mathbf{X}^T \mathbf{Y} - \mathbf{x}_i y_i = \begin{pmatrix} \sum_{j=1}^n y_j \\ \sum_{j=1}^n x_{j1} y_j \\ \vdots \\ \sum_{j=1}^n x_{jk} y_j \end{pmatrix} - \begin{pmatrix} y_i \\ x_{i1} y_i \\ \vdots \\ x_{ik} y_i \end{pmatrix} = \begin{pmatrix} \sum_{j=1, j \neq i}^n y_j \\ \sum_{j=1, j \neq i}^n x_{j1} y_j \\ \vdots \\ \sum_{j=1, j \neq i}^n x_{jk} y_j \end{pmatrix} = \mathbf{X}_{-i}^T \mathbf{Y}_{-i}. \quad (3.16)$$

Multipliziert man nun beide Seiten von (3.15) mit $\mathbf{X}^T \mathbf{Y} - \mathbf{x}_i y_i = \mathbf{X}_{-i}^T \mathbf{Y}_{-i}$, so erhält man folgendes:

$$\begin{aligned}
\underbrace{(\mathbf{X}_{-i}^T \mathbf{X}_{-i})^{-1} \mathbf{X}_{-i}^T \mathbf{Y}_{-i}}_{=\hat{\beta}_{-i}} &= \underbrace{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}}_{=\hat{\beta}} - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i y_i + \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^T \overbrace{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}}^{=\hat{\beta}}}{1 - h_{ii}} - \\
&\quad - \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^T \overbrace{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i y_i}^{=h_{ii}}}{1 - h_{ii}} \\
\Rightarrow \hat{\beta} - \hat{\beta}_{-i} &= \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i y_i - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i y_i h_{ii}}{1 - h_{ii}} - \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^T \hat{\beta}}{1 - h_{ii}} + \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i y_i h_{ii}}{1 - h_{ii}} \\
&= \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i y_i - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^T \hat{\beta}}{1 - h_{ii}} \\
&= \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \overbrace{(y_i - \mathbf{x}_i^T \hat{\beta})}^{=e_i}}{1 - h_{ii}}.
\end{aligned}$$

Somit erhält man folgendes zur Herleitung nachfolgender Diagnostiken wichtiges Resultat:

$$\hat{\beta} - \hat{\beta}_{-i} = \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i e_i}{1 - h_{ii}}. \quad (3.17)$$

Dies spiegelt nun die Veränderung des Schätzers $\hat{\beta}$, falls die i -te Beobachtung gelöscht wird. Daraus lässt sich schließen, welchen Einfluß die i -te Beobachtung auf die Regressionskoeffizienten

hat. Wie man sieht, ist die Änderung von den Residuen und den Diagonalelementen von H abhängig.

Für den Schätzer s_{-i}^2 für σ^2 gilt nun nach Kapitel 2.2 folgendes:

$$(n-p-1)s_{-i}^2 = \sum_{j=1, j \neq i}^n (y_j - \mathbf{x}_j^T \hat{\boldsymbol{\beta}}_{-i})^2. \quad (3.18)$$

Dies ist die Fehlerquadratsumme im Modell (3.9) geteilt durch $n-p-1$ statt $n-p$, da man die Schätzung ohne den i -ten Datenpunkt durchführt.

Löst man nun in (3.17) nach $\hat{\boldsymbol{\beta}}_{-i}$ auf und setzt in (3.18) ein, so erhält man nach einigen Umformungen (3.12). Insbesondere gilt:

$$\begin{aligned} (n-p-1)s_{-i}^2 &= \sum_{j=1, j \neq i}^n \left(y_j - \mathbf{x}_j^T \left(\hat{\boldsymbol{\beta}} - \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i e_i}{1 - h_{ii}} \right) \right)^2 \\ &= \sum_{j=1}^n \left(y_j - \mathbf{x}_j^T \hat{\boldsymbol{\beta}} + \frac{\mathbf{x}_j^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i e_i}{1 - h_{ii}} \right)^2 - \left(y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}} + \frac{\mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i e_i}{1 - h_{ii}} \right)^2 \\ &= \sum_{j=1}^n \left(e_j + \frac{h_{ij} e_i}{1 - h_{ii}} \right)^2 - \frac{e_i^2}{(1 - h_{ii})^2} \\ &= \sum_{j=1}^n e_j^2 + \frac{2e_i}{1 - h_{ii}} \underbrace{\sum_{j=1}^n e_j h_{ij}}_{=\sum_{j=1}^n (y_j - \sum_{k=1}^n h_{jk} y_k) h_{ij} \stackrel{(3.3)}{=} 0} + \frac{e_i^2}{(1 - h_{ii})^2} \underbrace{\sum_{j=1}^n h_{ij}^2}_{\stackrel{(3.4)}{=} h_{ii}} - \frac{e_i^2}{(1 - h_{ii})^2} \\ &= \sum_{j=1}^n e_j^2 + \frac{h_{ii} e_i^2}{(1 - h_{ii})^2} - \frac{e_i^2}{(1 - h_{ii})^2} \\ &= \sum_{j=1}^n e_j^2 - \frac{e_i^2}{1 - h_{ii}} \\ &= (n-p)s^2 - \frac{e_i^2}{1 - h_{ii}}. \end{aligned}$$

Damit ist (3.12) gezeigt.

Zwischen t_i , die t_{n-p-1} -verteilt sind, und r_i besteht eine Beziehung. Diese erhält man durch Einsetzen von (3.13) in (3.11), also ergibt sich

$$t_i = r_i \left(\frac{n-p-1}{n-p-r_i^2} \right)^{1/2}, \quad (3.19)$$

was zeigt, dass t_i eine monotone Transformation von r_i ist und falls $r_i^2 \rightarrow n-p$, dann $t_i^2 \rightarrow \infty$. Deshalb zeigt t_i große Abweichungen stärker als r_i . Außerdem ist s_{-i}^2 robust gegenüber Problemen mit großem Fehler in der i -ten Beobachtung. Ein weiterer Grund, warum zumeist t_i vorgezogen wird, ist, dass t_i die t -Statistik für das *Mean-Shift-Outlier-Model* ist. Dieses wird in Kapitel 3.4 besprochen.

Desweiteren lassen sich die vorhergesagten Residuen, die *PRESS*-Residuen, ausrechnen, die wie folgt definiert sind:

$$\begin{aligned}
 e_{i,-i} &:= y_i - \hat{y}_{i,-i} \\
 &= y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{-i} \\
 &= y_i - \hat{y}_i + \hat{y}_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{-i} \\
 &= \underbrace{y_i - \hat{y}_i}_{=e_i} + \underbrace{\mathbf{x}_i^T (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{-i})}_{\stackrel{(3.17)}{=} \frac{h_{ii} e_i}{1-h_{ii}}} \\
 &= \frac{e_i}{1-h_{ii}}.
 \end{aligned} \tag{3.20}$$

Wenn man diese standardisiert, erhält man dasselbe Ergebnis wie für die standardisierten Residuen, da

$$\text{Var}(e_{i,-i}) = \text{Var}\left(\frac{e_i}{1-h_{ii}}\right) = \frac{1}{(1-h_{ii})^2} \text{Var}(e_i) \stackrel{(3.7)}{=} \frac{\sigma^2}{1-h_{ii}}. \tag{3.21}$$

Also ergibt eine Standardisierung folgendes:

$$\frac{e_i}{1-h_{ii}} \Big/ \frac{s}{\sqrt{1-h_{ii}}} = \frac{e_i \sqrt{1-h_{ii}}}{s(1-h_{ii})} = \frac{e_i}{s\sqrt{1-h_{ii}}} = r_i.$$

Darüberhinaus lässt sich die *PRESS*-Statistik berechnen. Diese ist die Summe der quadrierten vorhergesagten Residuen, also

$$\text{PRESS} := \sum_{i=1}^n e_{i,-i}^2.$$

Mit Hilfe dieser Statistik lässt sich ein Maß für die Vorhersagefähigkeit des Modells aufstellen, welches folgendermaßen definiert wird:

$$R_{pred}^2 := 1 - \frac{\text{PRESS}}{\text{SQT}}.$$

Wie beim Bestimmtheitsmaß strebt man einen hohen Wert an, welcher eine gute Vorhersagefähigkeit prognostiziert.

3.4 Ausreißer-Test (Mean-Shift-Outlier-Test)

Nun wird ein Ausreißertest konstruiert, um etwaige Ausreißer zu identifizieren. Insbesondere soll getestet werden, ob eine einzelne Beobachtung als Ausreißer bewertet werden kann. Da

$$\begin{aligned}
 \frac{y_i - \hat{y}_{i,-i}}{\sqrt{\widehat{\text{Var}}(y_i - \hat{y}_{i,-i})}} &\stackrel{(3.20),(3.21)}{=} \frac{\frac{e_i}{1-h_{ii}}}{s_{-i}/\sqrt{1-h_{ii}}} \\
 &= \frac{e_i}{s_{-i}\sqrt{1-h_{ii}}} \\
 &= t_i
 \end{aligned}$$

gilt, kann man diesen Test konstruieren. Das Modell, das sogenannte *Mean-Shift-Outlier-Model*, lautet folgendermaßen:

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \gamma d_{li} + \epsilon_i, \quad i = 1, \dots, n, \quad d_{li} = \begin{cases} 1 & \text{für } l = i \\ 0 & \text{sonst} \end{cases}.$$

Die addierte Variable d_{li} wird als Indikatorvariable bezeichnet. Damit hat die l -te Beobachtung einen Erwartungswert von $\mathbf{x}_l^T \boldsymbol{\beta} + \gamma$. Es wird folgendes definiert:

$$\begin{aligned} \Delta_l &:= y_l - \mathbf{x}_l^T \boldsymbol{\beta}, \\ \hat{\Delta}_l &:= y_l - \mathbf{x}_l^T \hat{\boldsymbol{\beta}}_{-l} = y_l - \hat{y}_{l,-l}, \end{aligned}$$

wobei $\hat{\boldsymbol{\beta}}_{-l}$ der Schätzer für $\boldsymbol{\beta}$ im Modell ohne die l -te Beobachtung und $\hat{y}_{l,-l}$ der Schätzer für y_l basierend auf dem Modell ohne l -te Beobachtung ist. Die Nullhypothese und die Alternative lauten dann wie folgt:

$$\begin{aligned} H_0 : \Delta_l = 0 & \quad \text{gegen} \quad H_1 : \Delta_l \neq 0 \\ \text{bzw. } H_0 : \gamma = 0 & \quad \text{gegen} \quad H_1 : \gamma \neq 0. \end{aligned}$$

Die Hypothese besagt also, dass die l -te Beobachtung kein Ausreißer ist. Als Teststatistik verwendet man

$$\begin{aligned} t_i &= \frac{e_i}{s_{-i} \sqrt{1 - h_{ii}}} \\ \stackrel{(3.12)}{=} & \frac{\frac{e_i}{\sqrt{1 - h_{ii}}}}{\left(\left((n - p) s^2 - \frac{e_i^2}{1 - h_{ii}} \right) / (n - p - 1) \right)^{1/2}}. \end{aligned} \tag{3.22}$$

Unter der Normalitätsannahme ist die Verteilung von t_i unter der Nullhypothese eine t -Verteilung mit $n - p - 1$ Freiheitsgraden. Es gilt also bei Betrachtung einer einzelnen verdächtigen Beobachtung l folgende Testregel:

$$\text{Verwerfe } H_0 : \Delta_l = 0 \text{ gegen } H_1 : \Delta_l \neq 0 \text{ zum Niveau } \alpha \Leftrightarrow |t_l| = \frac{|\hat{\Delta}_l|}{\sqrt{\widehat{\text{Var}}(\hat{\Delta}_l)}} > t_{n-p-1, 1-\alpha/2}.$$

Der Wert α bezeichnet hier das Signifikanzniveau und $t_{n-p-1, 1-\alpha/2}$ ist das $(1 - \alpha/2)$ -Quantil der t -Verteilung. Ist der Betrag der Teststatistik größer als der entsprechende t -Wert, wird die Nullhypothese zum Signifikanzniveau α verworfen und der l -te Punkt somit als Ausreißer angesehen. Der t -Test ist also im Fall einer einzigen Beobachtung l , die verdächtig ist, geeignet. Jedoch kann ein Problem auftreten, wenn man testet ohne den Index l zu kennen. Intuitiv verwendet man beim multiplen Testen das Maximum von t_i , d.h.

$$t_{max} = \max_{1 \leq i \leq n} |t_i|.$$

Allerdings bedeutet das Testen der Beobachtung, welche den größten t -Wert hat, dass n t -Tests gemacht werden. Diese Tests bei einem α -Niveau von 0.05 zu machen, ist nicht richtig. Für eine Stichprobe von $n = 65$ und Parameterzahl $p = 5$ ist die Wahrscheinlichkeit, dass eine t -Statistik mit 60 Freiheitsgraden den Wert 2.00 überschreitet, 0.05. Hingegen ist die Wahrscheinlichkeit,

dass der größte Wert t_{max} der 65 unabhängigen t -Tests den Wert 2.00 überschreitet, 0.964. Deswegen sind multiple Tests, die wie einzelne Tests gehandhabt werden, nicht geeignet. Da der Wert t_{max} das Maximum von n t -Tests ist, ist seine Verteilung unter der Nullhypothese nicht mehr die t -Verteilung. Der exakte p -Wert für t_{max} ist schwierig zu berechnen. Brauchbare kritische Werte für die Teststatistik können über die *Bonferroni Ungleichung* ausgerechnet werden. Genaueres dazu ist im Artikel von Cook und Prescott (1981) zu finden, in dem eine Schranke hergeleitet wird, um kritische Werte zu bestimmen. Man berechnet eine Schranke auf Basis des p -Wertes:

$$p(t_{max}) \leq 2nP(t_{n-p-1} > t_{max}) = np(t_{n-p-1}), \quad (3.23)$$

wobei $p(t_{max})$ der p -Wert von t_{max} und $p(t_{n-p-1})$ der p -Wert von t_{n-p-1} ist. Diese Ungleichung besagt also, dass der p -Wert von t_{max} nicht größer als n -mal der p -Wert von t_{n-p-1} ist. Nimmt man ein Signifikanzniveau von α für n Tests an, dann ergibt sich aus dem (α/n) 100%-Punkt der t_{n-p-1} -Verteilung ein Signifikanzniveau, dass nicht größer als α ist. Der Test garantiert, dass keine auffälligen Punkte vorher bekannt sind. Im Buch von Myers (1990) ist im Anhang eine Tabelle mit kritischen Werten für entsprechende Tests aufgeführt.

Darüberhinaus kann man den Wert γ schätzen. Näheres dazu findet man in Cook und Weisberg (1982).

3.5 Bestimmung des Einflusses einer Beobachtung

Nun ist es möglich, den Einfluß einer Beobachtung auf die vorhergesagten Werte und die Regressionskoeffizienten zu untersuchen. Danach kann man geeignete Interpretationen über den Grad des Einflusses der einzelnen Beobachtungen zulassen.

3.5.1 DFFITS und DFBETAS

Eine geeignete Diagnostik, um den Einfluß der i -ten Beobachtung auf die vorhergesagten Werte zu untersuchen, wird gegeben durch

$$\begin{aligned} DFFITS_i & := \frac{\hat{y}_i - \hat{y}_{i,-i}}{s_{-i}\sqrt{h_{ii}}} \\ & = \frac{\mathbf{x}_i^T \hat{\boldsymbol{\beta}} - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{-i}}{s_{-i}\sqrt{h_{ii}}} \\ & = \frac{\mathbf{x}_i^T (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{-i})}{s_{-i}\sqrt{h_{ii}}} \\ (3.17) \quad & \stackrel{=}{=} \frac{h_{ii}e_i/(1 - h_{ii})}{s_{-i}\sqrt{h_{ii}}} \\ & = \left(\frac{e_i}{s_{-i}\sqrt{1 - h_{ii}}} \right) \left(\frac{h_{ii}}{1 - h_{ii}} \right)^{1/2} \\ (3.11) \quad & \stackrel{=}{=} t_i \left(\frac{h_{ii}}{1 - h_{ii}} \right)^{1/2}. \end{aligned} \quad (3.24)$$

Die Bezeichnung DF steht für die Differenz des Ergebnis mit und ohne der i -ten Beobachtung. Hier werden die vorhergesagten Werte, welche einerseits auf Basis aller Beobachtungen und andererseits basierend auf allen außer der i -ten Beobachtung errechnet wurden, subtrahiert.

Der Nenner in der ersten Gleichung ist eine Standardisierung, falls $Var(\hat{y}_i) = \sigma^2 h_{ii}$. Der Wert $DFFITs_i$ gibt die Zahl der geschätzten Standardfehler für den i -ten Punkt an, für die der geschätzte Wert geändert wird, wenn der i -te Punkt gelöscht wird. Wie man feststellt, ist also $DFFITs_i$ von den extern standardisierten Residuen und vom Diagonalelement h_{ii} der Hatmatrix H abhängig. Falls der Wert für h_{ii} nahe 1 und das Residuum groß ist, so ergibt sich ein großer Wert für $DFFITs_i$. Hingegen lässt ein Wert von h_{ii} nahe 0 einen mäßigen Einfluß des Residuums zu. Wichtig ist jedenfalls, dass sowohl das Residuum als auch das Diagonalelement von H einzeln und zusammen in $DFFITs_i$ betrachtet wird. Ab wann betrachtet man nun aber die i -te Beobachtung als einflußreich. Hierauf wird im Anschluß an die folgende Diagnostik zwecks ähnlicher Problematik eingegangen.

Eine Statistik, um die Zahl der Standardfehler, für die der Koeffizient geändert wird, anzugeben, wird nachfolgend definiert:

$$\begin{aligned}
 DFBETAS_{j,i} &:= \frac{\hat{\beta}_j - \hat{\beta}_{j,-i}}{s_{-i} \sqrt{c_{jj}}} \\
 (3.17) \quad &= \frac{r_{j,i} e_i / (1 - h_{ii})}{s_{-i} \sqrt{c_{jj}}} \\
 &= \frac{r_{j,i} e_i / (1 - h_{ii})}{s_{-i} \sqrt{((\mathbf{X}^T \mathbf{X})^{-1})_{jj} ((\mathbf{X}^T \mathbf{X})_{jj} ((\mathbf{X}^T \mathbf{X})^{-1})_{jj})}} \\
 &= \frac{\frac{r_{j,i}}{\sqrt{\mathbf{r}_j^T \mathbf{r}_j}} \frac{e_i}{s_{-i} (1 - h_{ii})}}{1} \\
 (3.11) \quad &= \frac{r_{j,i}}{\sqrt{\mathbf{r}_j^T \mathbf{r}_j}} \frac{1}{\sqrt{1 - h_{ii}}} t_i, \tag{3.25}
 \end{aligned}$$

wobei $c_{jj} = ((\mathbf{X}^T \mathbf{X})^{-1})_{jj}$, $R = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ und \mathbf{r}_j^T ist die j -te Zeile von R . Der Nenner in der ersten Gleichung ist ebenso eine Standardisierung, falls $Var(\hat{\beta}_j) = \sigma^2 \sqrt{c_{jj}}$.

Die Diagnostik ist wiederum von den extern standardisierten Residuen und dem i -ten Diagonalelement der Hatmatrix H abhängig. Der Vorfaktor $\frac{r_{j,i}}{\sqrt{\mathbf{r}_j^T \mathbf{r}_j}}$ ist ein normalisiertes Maß des

Einflusses der i -ten Beobachtung auf den j -ten Regressionskoeffizienten. Dieser Wert wird größer mit anwachsendem h_{ii} . Es sollten bei der Betrachtung und Interpretation der $DFBETAS_{j,i}$ genau wie bei voriger Diagnostik auch die einzelnen Komponenten, sprich extern standardisiertes Residuum und Diagonalelement von H , gesondert und zusammen bezüglich der Diagnostik angesehen werden.

Nun fragt man sich, ab welchem Wert die $DFFITs_i$ und $DFBETAS_{j,i}$ einen signifikanten Einfluß bescheinigen. Belsley et al. (1980) schlagen hierzu Schranken vor. Intuitiv kann zuerst eine erste Schranke von 2 aufgrund des 95%-Niveaus der Standardnormalverteilung angegeben werden, da die Diagnostiken skaliert wurden. Um aber genauere Aussagen zu treffen und extreme Werte zu bestimmen, sollte auch die Stichprobengröße n betrachtet werden. Somit wird für $DFFITs_i$ ein Wert von $2\sqrt{p/n}$ und für $DFBETAS_{j,i}$ ein Wert von $2/\sqrt{n}$ vorgeschlagen. Hier fließt zusätzlich der Wert p ein, da bei Ansteigen von p auch $DFFITs_i$ wächst. Natürlich ist klar, dass bei großen Stichproben die Löschung einer Beobachtung große Veränderungen bewirkt. Deshalb bleibt immer noch die absolute Schranke von 2 zu beachten. Ausserdem kann man auf Werte achten, die sich sehr von den übrigen unterscheiden. Hier ist nicht ein Vergleichswert entscheidend, sondern die Lücke zwischen dem auffälligen Wert und dem Rest.

3.5.2 Cook's Distanz

Darüberhinaus lässt sich ein reellwertiges standardisiertes Maß herleiten, das den Grad der Veränderung der gesamten Regressionskoeffizienten zeigt. Dieses Maß wird als *Cook's Distanz* bezeichnet und ist gegeben durch

$$\begin{aligned}
 D_i &:= \frac{(\hat{\beta} - \hat{\beta}_{-i})^T (\mathbf{X}^T \mathbf{X}) (\hat{\beta} - \hat{\beta}_{-i})}{ps^2} \\
 &\stackrel{(3.17)}{=} \frac{\mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{X}) (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i e_i^2}{(1 - h_{ii})^2 ps^2} \\
 &= \left(\frac{e_i^2}{(1 - h_{ii})^2} \right) \left(\frac{h_{ii}}{s^2 p} \right) \\
 &\stackrel{(3.8)}{=} \left(\frac{r_i^2}{p} \right) \left(\frac{h_{ii}}{1 - h_{ii}} \right). \tag{3.26}
 \end{aligned}$$

Die Cook's Distanz ist ein positives Maß, falls $\mathbf{X}^T \mathbf{X}$ eine positiv definite Matrix ist. Ein großer Wert für D_i entspricht einer großen Veränderung der Regressionskoeffizienten und somit signifikantem Einfluß. Man betrachtet eine Beobachtung als einflußreich, wenn D_i einen Wert größer als 1 annimmt. Um diesen Wert zu erklären, ist die Verteilungstheorie von Seber (1977) (Seite 54) nötig. Es gilt neben der Unabhängigkeit

$$\begin{aligned}
 \frac{(\hat{\beta} - \beta)^T (\mathbf{X}^T \mathbf{X}) (\hat{\beta} - \beta)}{\sigma^2} &\sim \chi_p^2 \quad \text{und} \\
 \frac{(n-p)s^2}{\sigma^2} &\sim \chi_{n-p}^2.
 \end{aligned}$$

Deshalb ist

$$\frac{(\hat{\beta} - \beta)^T (\mathbf{X}^T \mathbf{X}) (\hat{\beta} - \beta)}{ps^2} = \frac{\frac{(\hat{\beta} - \beta)^T (\mathbf{X}^T \mathbf{X}) (\hat{\beta} - \beta)}{ps^2}}{\frac{(n-p)s^2}{(n-p)\sigma^2}} \sim F_{p, n-p} \tag{3.27}$$

verteilt. Der erste Ausdruck in (3.27) entspricht einem Konfidenzbereich für β basierend auf $\hat{\beta}$. Der Konfidenzbereich ist zentriert um $\hat{\beta}$. Um die Ähnlichkeit von (3.26) und (3.27) auszunutzen, kann man die Cook's Distanz D_i als eine F -Statistik interpretieren, wobei anzumerken ist, dass D_i nicht F -verteilt ist. Aber wenn D_i approximativ dem 50%-Quantil der entsprechenden F -Verteilung gleicht, dann verschiebt eine Löschung der i -ten Beobachtung den Schätzer für β auf den Rand eines 50%-Konfidenzbereichs relativ zu $\hat{\beta}$. Einflußreiche Beobachtungen haben ein großes D_i zur Folge. Die genaue Definition von 'groß' hängt vom jeweiligen Problem ab, aber ein D_i , das größer als 1 ist, entspricht Veränderungen zwischen $\hat{\beta}$ und $\hat{\beta}_{-i}$ jenseits eines 50%-Konfidenzbereichs und dient als angemessener Vergleich. Um genauer zu sehen, welche Koeffizienten beeinflusst werden, müssen die $DFBETAS_{j,i}$ angesehen werden.

Die Cook's Distanz D_i lässt sich ebenso über die Konfidenzbereiche herleiten. Darauf soll erst genauer bei den GLM's eingegangen werden.

Hierbei sollte auch die enge Verbindung der bisher dargestellten drei Diagnostiken klar geworden sein. Jedoch zeigen sie nicht auf, ob nun das Modell mit oder ohne i -te Beobachtung am besten angepasst wird. Der Wert $DFBETAS_{j,i}$ zeigt nicht, ob die Betrachtung mit oder ohne die i -te Beobachtung die Schätzung der Regressionskoeffizienten verbessert. Im nächsten Kapitel wird diesem Sachverhalt nachgegangen.

3.5.3 COVRATIO und FVRATIO

Man konzentriert sich nun auf die *allgemeine Varianz* AV der Regressionskoeffizienten, die aus $Var(\hat{\beta}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$ und somit aus dem Gauß-Markov-Satz hervorgeht:

$$AV := det(Var(\hat{\beta})) = det(\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}).$$

Um den Einfluß der i -ten Beobachtung zu erfassen, definiert man einen *Kovarianz-Ratio* wie folgt:

$$\begin{aligned} COVRATIO_i &:= \frac{det(\widehat{Var}(\hat{\beta}_{-i}))}{det(\widehat{Var}(\hat{\beta}))} \\ &= \frac{det(s_{-i}^2(\mathbf{X}_{-i}^T\mathbf{X}_{-i})^{-1})}{det(s^2(\mathbf{X}^T\mathbf{X})^{-1})} \\ &= \frac{s_{-i}^{2p} det((\mathbf{X}_{-i}^T\mathbf{X}_{-i})^{-1})}{s^{2p} det((\mathbf{X}^T\mathbf{X})^{-1})}. \end{aligned}$$

Die Bezeichnung *COV* steht für die Koeffizientenvarianz. Mit Hilfe des Lemmas 3.2 kann der $COVRATIO_i$ noch vereinfacht dargestellt werden. Es folgt dann:

$$COVRATIO_i = \frac{s_{-i}^{2p}}{s^{2p}} \left(\frac{1}{1 - h_{ii}} \right). \quad (3.28)$$

Hier wird also die allgemeine Varianz der Regressionskoeffizienten ohne i -te Beobachtung durch die allgemeine Varianz der Regressionskoeffizienten mit i -ter Beobachtung geteilt, wobei σ^2 jeweils durch die entsprechenden Schätzer ersetzt wird. Der Wertebereich von $COVRATIO_i$ liegt in \mathbb{R}_0^+ . Der $COVRATIO_i$ ist ein Produkt aus zwei Faktoren, nämlich dem Quotienten aus jeweils der p -ten Potenz des Schätzers der Varianz ohne i -te Beobachtung und des Schätzers der Varianz mit i -ter Beobachtung und einem Ausdruck basierend auf dem Diagonalelement h_{ii} . Hat nun der i -te Punkt ein großes Residuum, d.h ein y -Ausreißer liegt vor, dann ist der erste Faktor kleiner als 1. Ein Punkt mit geringem Residuum hat für den ersten Faktor einen Wert um 1 zur Folge. Der zweite Faktor dagegen ist größer als 1, wenn ein high leverage Punkt auftritt. Bei einem sehr geringen Wert von h_{ii} ist der zweite Faktor auch ungefähr 1. Wie man feststellt, können sich beide Faktoren aufheben, was es nötig macht zuvor die einzelnen Komponenten des Produkts zu betrachten. Hier wird nun der Ausdruck (3.28) in einer anderen Form geschrieben und danach versucht geeignete Schranken zu finden, um das Ausmaß des Einflusses festzustellen. Belsley, Kuh, und Welsch (1980) schlagen Schranken vor, die für große Stichproben vernünftigerweise erscheinen.

Der $COVRATIO_i$ lässt sich mit Hilfe von (3.12) und (3.11) schreiben als

$$\begin{aligned} COVRATIO_i &= \frac{s_{-i}^{2p}}{s^{2p}} \left(\frac{1}{1 - h_{ii}} \right) \\ &\stackrel{(3.12)}{=} \frac{s_{-i}^{2p}}{\left(\frac{(n-p-1)s_{-i}^2 + e_i^2/(1-h_{ii})}{n-p} \right)^p} \left(\frac{1}{1 - h_{ii}} \right) \\ &\stackrel{(3.11)}{=} \left(\frac{(n-p-1)s_{-i}^2}{(n-p)s_{-i}^2} + \frac{s_{-i}^2(1-h_{ii})t_i^2}{(n-p)s_{-i}^2} \right)^{-p} \left(\frac{1}{1 - h_{ii}} \right) \end{aligned}$$

$$= \frac{1}{\left(\frac{n-p-1}{n-p} + \frac{t_i^2}{n-p}\right)^p (1 - h_{ii})}. \quad (3.29)$$

Man ist nun interessiert an Werten die nicht in der Nähe von 1 liegen, denn eben diese sind als einflußreich zu deklarieren. Eine obere bzw. untere Schranke für $COVRATIO_i$ lässt sich wie folgt herleiten. Um einen signifikanten Bereich zu konstruieren, nimmt man zwei Extremfälle an, wobei bei beiden n groß angenommen wird. Zuerst sieht man sich den Fall mit $|t_i| \geq 2$ mit Minimum der $h_{ii} = 1/n$ an:

$$\begin{aligned} COVRATIO_i &\approx \frac{1}{\left(1 + \frac{t_i^2 - 1}{n-p}\right)^p} \leq \frac{1}{\left(1 + \frac{3}{n-p}\right)^p} \\ &\approx \left(1 + \frac{3p}{n}\right)^{-1} \approx 1 - \frac{3p}{n}. \end{aligned}$$

Für die obige Approximation wurde der Einfachheit halber $n - p$ durch n ersetzt. Für die letzte Approximation wurde eine Taylor-Entwicklung erster Ordnung angewandt. Die beiden Approximationen zuletzt sind natürlich nur gegeben, wenn $n \geq 3p$.

Der zweite Extremfall ergibt sich aus $h_{ii} \geq 2p/n$ mit $t_i = 0$:

$$\begin{aligned} COVRATIO_i &= \frac{1}{\left(1 - \frac{1}{n-p}\right)^p} \frac{1}{1 - h_{ii}} \geq \frac{1}{\left(1 - \frac{1}{n-p}\right)^p \left(1 - \frac{2p}{n}\right)} \\ &\approx \frac{1}{\left(1 - \frac{p}{n}\right) \left(1 - \frac{2p}{n}\right)} \\ &\approx \left(1 - \frac{3p}{n}\right)^{-1} \approx 1 + \frac{3p}{n}. \end{aligned}$$

Der Einfachheit halber wird wiederum $n - p$ durch n ersetzt und $n \geq 3p$ angenommen. Bei der vorletzten Approximation werden Terme quadratischer Ordnung, also $1/n^2$ weggelassen. Letztere Approximation resultiert aus einer Anwendung von Taylor erster Ordnung. Für große Stichproben hat man also einen signifikanten Bereich, der einflußreiche Beobachtungen identifiziert. Falls $COVRATIO_i$ grösser bzw. kleiner als $1 + 3p/n$ bzw. $1 - 3p/n$ ist, kann man ein ungewöhnliches Ausmaß des Einflusses auf die allgemeine Varianz feststellen.

Die Formel in (3.29) ist abhängig von den extern standardisierten Residuen und von dem Wert h_{ii} . $COVRATIO_i$ ist umso höher, je höher h_{ii} ist und umso niedriger, je höher t_i ist. Wie vorher erwähnt, ist es darüberhinaus möglich, dass sich die beiden Einflußfaktoren aufheben. Deshalb ist es wichtig, die Faktoren getrennt voneinander und gemeinsam in (3.29) zu betrachten. Danach lassen sich erst Rückschlüsse für das Modell ziehen.

Darüberhinaus interessiert die Varianz der gefitteten y -Werte \hat{y}_i mit und ohne Betrachtung der i -ten Beobachtung. Wir betrachten also die Varianzen:

$$\begin{aligned} Var(\hat{y}_i) &= Var(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}) = \mathbf{x}_i^T Var(\hat{\boldsymbol{\beta}}) \mathbf{x}_i \\ &= \sigma^2 \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i = \sigma^2 h_{ii}, \\ Var(\hat{y}_{i,-i}) &= Var(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{-i}) = \mathbf{x}_i^T Var(\hat{\boldsymbol{\beta}}_{-i}) \mathbf{x}_i \\ &= \mathbf{x}_i^T Var((\mathbf{X}_{-i}^T \mathbf{X}_{-i})^{-1} \mathbf{X}_{-i}^T \mathbf{Y}_{-i}) \mathbf{x}_i \end{aligned}$$

$$\begin{aligned}
&= \mathbf{x}_i^T (\mathbf{X}_{-i}^T \mathbf{X}_{-i})^{-1} \mathbf{X}_{-i}^T \text{Var}(\mathbf{Y}_{-i}) \mathbf{X}_{-i} (\mathbf{X}_{-i}^T \mathbf{X}_{-i})^{-1} \mathbf{x}_i \\
&= \sigma^2 \mathbf{x}_i^T (\mathbf{X}_{-i}^T \mathbf{X}_{-i})^{-1} \mathbf{x}_i \\
(3.15) \quad &\stackrel{=}{=} \sigma^2 \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i + \sigma^2 \mathbf{x}_i^T \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1}}{1 - h_{ii}} \mathbf{x}_i \\
&= \frac{\sigma^2 h_{ii} (1 - h_{ii}) + \sigma^2 h_{ii}^2}{1 - h_{ii}} \\
&= \sigma^2 \frac{h_{ii}}{1 - h_{ii}}.
\end{aligned}$$

Wenn man nun die geschätzten Varianzen

$$\begin{aligned}
\widehat{\text{Var}}(\hat{y}_i) &= s^2 h_{ii}, \\
\widehat{\text{Var}}(\hat{y}_{i,-i}) &= s_{-i}^2 \frac{h_{ii}}{1 - h_{ii}}
\end{aligned}$$

betrachtet, ergibt sich daraus der *FVRATIO*_{*i*}:

$$\begin{aligned}
FVRATIO_i &= \frac{\text{Var}(\hat{y}_{i,-i})}{\text{Var}(\hat{y}_i)} = \frac{s_{-i}^2}{s^2 (1 - h_{ii})} \\
&= \frac{1}{\left(\frac{n-p-1}{n-p} + \frac{t_i^2}{n-p} \right) (1 - h_{ii})}.
\end{aligned} \tag{3.30}$$

Dabei ist die Abkürzung *FV* auf den Fit und die Varianz zurückzuführen. Dieser Ausdruck ist dem *COVRATIO*_{*i*} somit ähnlich, es geht nur keine *p*-te Potenz in den Ratio ein. Als Diagnostikmaß zeigt dieser Ratio aber dieselben Verhaltensmuster. Als kritische Werte kann man also $1 + 3/n$ und $1 - 3/n$ annehmen.

In Tabelle 3.1 werden nochmals die wichtigsten vorgestellten Regressionsdiagnostiken für LM's zur Übersicht aufgelistet.

3.6 Regressionsdiagnostiken bei Gewichtung einer Beobachtung

3.6.1 Gewichtung

Nun kann man aber auch eine Gewichtung einer einzelnen Beobachtung durchführen. Es wird die *i*-te Beobachtung nicht gelöscht, sondern geeignet gewichtet. Man ersetzt also im Modell $\text{Var}(\epsilon_i) = \sigma^2$ durch $\text{Var}(\epsilon_i) = \sigma^2/w_i$. Hier nimmt man $w := w_i$ an, da man sich auf die Gewichtung einer Beobachtung beschränkt. Dies entspricht dann der in Kapitel 2.4 angesprochenen gewichteten Regression, der Schätzer für $\hat{\beta}$ ist derselbe wie in (2.13). Es gilt also

$$\hat{\beta}_i(\mathbf{w}) = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{Y}. \tag{3.31}$$

Dabei soll der Index *i* für die *i*-te Beobachtung stehen, welche gewichtet wird. Für $w = 1$ erhält man den vertrauten Schätzer in (2.3). Also bezeichnet im folgenden $\hat{\beta}_i(1) = \hat{\beta}$ den Schätzer mit *i*-ter Beobachtung und $\hat{\beta}_i(0) = \hat{\beta}_{-i}$ den Schätzer ohne *i*-te Beobachtung.

Nun kann wiederum die Veränderung des Schätzers $\hat{\beta}_i(1)$ zum Schätzer mit Gewichtung der *i*-ten Beobachtung $\hat{\beta}_i(\mathbf{w})$ angesehen werden. Diese spiegelt sich in der folgenden Formel

Zusammenfassung: Regressionsdiagnostiken für LM's				
<i>Bezeichnung</i>	<i>Formel</i>	<i>(Nr.)</i>	$ \cdot \leq \text{Schranke}$ $ \cdot \geq \text{Schranke}$	<i>Zweck</i>
h_{ii}	$\mathbf{x}_i^T (\mathbf{X}^T \mathbf{X}) \mathbf{x}_i$	(3.1)	$\geq 2p/n$	Hatmatrix-Diagonalelemente zur Identifizierung von x-Ausreißern
e_i	$(1 - h_{ii})y_i$	(3.1)	-	Gewöhnliche Residuen zur Entdeckung von y-Ausreißern
r_i	$\frac{e_i}{s\sqrt{1-h_{ii}}}$	(3.8)	≥ 2	Standardisierte Residuen zur Entdeckung von y-Ausreißern
t_i	$\frac{e_i}{s_{-i}\sqrt{1-h_{ii}}}$	(3.11)	≥ 2	Extern standardisierte Residuen zur Aufdeckung des Einflusses der i-ten Beobachtung
$e_{i,-i}$	$\frac{e_i}{1-h_{ii}}$	(3.20)	-	Vorhergesagte Residuen zur Aufdeckung des Einflusses der i-ten Beobachtung
$DFFITs_i$	$t_i \left(\frac{h_{ii}}{1-h_{ii}} \right)^{1/2}$	(3.24)	$\geq 2\sqrt{p/n}$	Standardisierte Veränderung der geschätzten Werte (Fit) bei Löschung der i-ten Beobachtung
$DFBETAS_{j,i}$	$\frac{r_{j,i}}{\sqrt{\mathbf{r}_j^T \mathbf{r}_j}} \frac{1}{\sqrt{1-h_{ii}}} t_i$	(3.25)	$\geq 2/\sqrt{n}$	Standardisierte Veränderung von β_j bei Löschung der i-ten Beobachtung
D_i	$\left(\frac{r_i^2}{p} \right) \frac{h_{ii}}{1-h_{ii}}$	(3.26)	≥ 1	Cook's Distanz zur Aufdeckung des Einflusses der i-ten Beobachtung auf die Regressionskoeffizienten
$COVRATIO_i$	$\frac{1}{\left(\frac{n-p-1}{n-p} + \frac{t_i^2}{n-p} \right)^p (1-h_{ii})}$	(3.29)	$\leq 1 - 3p/n$ $\geq 1 + 3p/n$	Untersuchung des Einflusses der i-ten Beobachtung auf das Modell
$FVRATIO_i$	$\frac{1}{\left(\frac{n-p-1}{n-p} + \frac{t_i^2}{n-p} \right) (1-h_{ii})}$	(3.30)	$\leq 1 - 3/n$ $\geq 1 + 3/n$	Untersuchung des Einflusses der i-ten Beobachtung auf das Modell

Tabelle 3.1: Die wichtigsten Diagnostiken für LM's sind hier aufgeführt. In der vierten Spalte basieren einige Schranken auf signifikanten Werten von bestimmten Verteilungen. Das intern standardisierte Residuum r_i und das extern standardisierte Residuum t_i verhält sich ähnlich zur t -Verteilung und die Cook's Distanz D_i ähnlich zur F -Verteilung.

$$\hat{\beta}_i(1) - \hat{\beta}_i(\mathbf{w}) = \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i (1-w) e_i}{1 - (1-w) h_{ii}}, \quad (3.32)$$

welche sich für $w = 0$ zu (3.17) vereinfacht. Um diese Formel aufzustellen, benutzt man wiederum Lemma 3.3 mit $\mathbf{A} := \mathbf{X}^T \mathbf{X}$ und $\mathbf{u} := (1-w) \mathbf{x}_i$, $\mathbf{v} := \mathbf{x}_i$:

$$(\mathbf{X}^T \mathbf{X})^{-1} - (1-w) \mathbf{x}_i \mathbf{x}_i^T = (\mathbf{X}^T \mathbf{X})^{-1} + \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} (1-w)}{1 - \mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i (1-w)}.$$

Und multipliziert danach beide Seiten mit $\mathbf{X}^T \mathbf{Y} - \mathbf{x}_i y_i (1-w) = \mathbf{X}^T \mathbf{W} \mathbf{Y}$. Letztere Gleichheit ergibt sich analog zu (3.16):

$$\begin{aligned} \underbrace{(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{Y}}_{=\hat{\beta}_i(\mathbf{w})} &= \underbrace{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}}_{=\hat{\beta}} - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i y_i (1-w) + \\ &\quad + \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^T \overbrace{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}}^{=\hat{\beta}} (1-w)}{1 - (1-w) h_{ii}} \\ &\quad - \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^T \overbrace{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i}^{=h_{ii}} y_i (1-w)^2}{1 - (1-w) h_{ii}} \\ \Rightarrow \hat{\beta} - \hat{\beta}_i(\mathbf{w}) &= \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i y_i (1-w) - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i y_i (1-w)^2 h_{ii}}{1 - (1-w) h_{ii}} - \\ &\quad - \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^T \hat{\beta} (1-w)}{1 - (1-w) h_{ii}} + \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i y_i (1-w)^2 h_{ii}}{1 - (1-w) h_{ii}} \\ &= \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i y_i (1-w) - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^T \hat{\beta} (1-w)}{1 - (1-w) h_{ii}} \\ &= \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i \overbrace{(y_i - \mathbf{x}_i^T \hat{\beta})}^{=e_i} (1-w)}{1 - (1-w) h_{ii}}. \end{aligned}$$

Für den Schätzer $s_i^2(w)$ für σ^2 ergibt sich analog zur Rechnung in Kapitel 3.3

$$(n-p) s_i^2(w) = (n-p) s^2 - \frac{(1-w) e_i^2}{1 - (1-w) h_{ii}}.$$

Analog zu den Berechnungen der gewöhnlichen und standardisierten Residuen in Kapitel 3.3 lassen sich auch hier die studentisierten Residuen berechnen:

$$t_i(w) := \frac{e_i}{s_i(w) \sqrt{1 - h_{ii}}}. \quad (3.33)$$

Ebenso lassen sich die in Kapitel 3.5 angegebenen Diagnostiken zur Bestimmung des Einflusses berechnen:

$$\begin{aligned}
DFFITs_i(w) &:= \frac{\hat{y}_i - \hat{y}_i(w)}{s(w)\sqrt{h_{ii}}} \\
&= \frac{\mathbf{x}_i^T \hat{\boldsymbol{\beta}} - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}(w)}{s(w)\sqrt{h_{ii}}} \\
&= \frac{\mathbf{x}_i^T (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}(w))}{s(w)\sqrt{h_{ii}}} \\
(3.32) \quad &\stackrel{=}{=} \frac{h_{ii}(1-w)e_i / (1 - (1-w)h_{ii})}{s(w)\sqrt{h_{ii}}} \\
&= \left(\frac{e_i}{s(w)\sqrt{1-h_{ii}}} \right) \left(\frac{(1-w)^2 h_{ii}(1-h_{ii})}{(1 - (1-w)h_{ii})^2} \right)^{1/2} \\
(3.33) \quad &\stackrel{=}{=} t_i(w) \left(\frac{(1-w)^2 h_{ii}(1-h_{ii})}{(1 - (1-w)h_{ii})^2} \right)^{1/2}, \tag{3.34}
\end{aligned}$$

$$\begin{aligned}
DFBETAS_{j,i}(w) &:= \frac{\hat{\beta}_j - \hat{\beta}_j(w)}{s(w)\sqrt{c_{jj}}} \\
(3.32) \quad &\stackrel{=}{=} \frac{r_{j,i}(1-w)e_i / (1 - (1-w)h_{ii})}{s(w)\sqrt{c_{jj}}} \\
&= \frac{r_{j,i}(1-w)e_i / (1 - (1-w)h_{ii})}{s(w)\sqrt{((\mathbf{X}^T \mathbf{X})^{-1})_{jj}((\mathbf{X}^T \mathbf{X})_{jj}((\mathbf{X}^T \mathbf{X})^{-1})_{jj})}} \\
&= \frac{r_{j,i}}{\sqrt{\mathbf{r}_j^T \mathbf{r}_j}} \frac{e_i}{s(w)\sqrt{1-h_{ii}}} \frac{(1-w)\sqrt{1-h_{ii}}}{1 - (1-w)h_{ii}} \\
(3.33) \quad &\stackrel{=}{=} \frac{r_{j,i}}{\sqrt{\mathbf{r}_j^T \mathbf{r}_j}} t_i(w) \frac{(1-w)\sqrt{1-h_{ii}}}{1 - (1-w)h_{ii}}, \tag{3.35}
\end{aligned}$$

$$\begin{aligned}
D_i(w) &:= \frac{(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}(w))^T (\mathbf{X}^T \mathbf{X}) (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}(w))}{ps^2} \\
(3.32) \quad &\stackrel{=}{=} \frac{\mathbf{x}_i^T (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{X}) (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i (1-w)^2 e_i^2}{(1 - (1-w)h_{ii})^2 ps^2} \\
&= \left(\frac{(1-w)^2 e_i^2}{(1 - (1-w)h_{ii})^2} \right) \left(\frac{h_{ii}}{s^2 p} \right) \\
(3.8) \quad &\stackrel{=}{=} \left(\frac{r_i^2}{p} \right) \left(\frac{(1-w)^2 h_{ii}(1-h_{ii})}{(1 - (1-w)h_{ii})^2} \right). \tag{3.36}
\end{aligned}$$

3.6.2 Influenzfunktion und Influenzkurven

Die *Influenz-Funktion* ist eine alternative Klasse von Maßen, um die i -te Beobachtung zu messen. Die Influenz-Funktion wird nach Cook und Weisberg (1982) wie folgt definiert:

$$IF_i(\mathbf{x}_i, y_i, F, T) = \lim_{\epsilon \rightarrow 0} \frac{T[(1-\epsilon)F + \epsilon \delta_{(\mathbf{x}_i, y_i)}] - T[F]}{\epsilon}, \tag{3.37}$$

wobei $T(\cdot)$ eine vektorwertige Statistik ist, basierend auf einer Zufallsstichprobe mit Verteilungsfunktion F und $\delta_{(\mathbf{x}_i, y_i)} = 1$ für (\mathbf{x}_i, y_i) und Null sonst. Nach Hinkley (1977) nimmt man an, dass der $(p+1)$ -Vektor (\mathbf{x}_i^T, y_i) , $\forall i$, eine gemeinsame Verteilung F hat mit

$$E \left[\begin{pmatrix} \mathbf{x}_i \\ y_i \end{pmatrix} (\mathbf{x}_i^T, y_i) \right] = \begin{pmatrix} \Sigma(F) & \gamma(F) \\ \gamma(F)^T & \tau(F) \end{pmatrix}.$$

Das entsprechende Funktional des Kleinste-Quadrat-Schätzers ist

$$T(F) = \Sigma(F)^{-1} \gamma(F), \quad (3.38)$$

wobei angenommen wird, dass Σ nicht-singulär ist. Die p -dimensionale Influenzkurve als Funktion von (\mathbf{x}_i^T, y_i) ist durch (3.37) punktweise definiert. Eine explizite Formel erhält man durch

$$\begin{aligned} \Sigma((1-\epsilon)F + \epsilon\delta_{(\mathbf{x}_i, y_i)}) &= (1-\epsilon) \left(\Sigma(F) + \frac{\epsilon}{1-\epsilon} \mathbf{x}_i \mathbf{x}_i^T \right), \\ \gamma((1-\epsilon)F + \epsilon\delta_{(\mathbf{x}_i, y_i)}) &= (1-\epsilon)\gamma(F) + \epsilon y_i \mathbf{x}_i. \end{aligned} \quad (3.39)$$

Ausgehend von Gleichung (3.39) ist die Aktualisierung von $\Sigma(F)^{-1}$ zu $\Sigma((1-\epsilon)F + \epsilon\delta_{(\mathbf{x}_i, y_i)})^{-1}$ äquivalent zur Addition einer neuen Variable zu \mathbf{x}_i mit Gewicht $\epsilon/(1-\epsilon)$. Mit Hilfe von Lemma 3.3 erhält man

$$\begin{aligned} \Sigma((1-\epsilon)F + \epsilon\delta_{(\mathbf{x}_i, y_i)})^{-1} &= \frac{1}{1-\epsilon} \left[\Sigma(F)^{-1} - \right. \\ &\quad \left. - \frac{\epsilon}{1-\epsilon} \Sigma(F)^{-1} \mathbf{x}_i \left(1 + \frac{\epsilon}{1-\epsilon} \mathbf{x}_i^T \Sigma(F)^{-1} \mathbf{x}_i \right)^{-1} \mathbf{x}_i^T \Sigma(F)^{-1} \right]. \end{aligned}$$

Setzt man nun in der Definition der Influenzkurve (3.37) für $T((1-\epsilon)F + \epsilon\delta_{(\mathbf{x}_i, y_i)})$ ein, vereinfacht und nimmt den Grenzwert, dann ergibt dies folgendes Resultat:

$$IF_i(\mathbf{x}_i, y_i, F, \beta) = \Sigma(F)^{-1} \mathbf{x}_i (y_i - \mathbf{x}_i^T T(F)), \quad (3.40)$$

wobei $\Sigma(F)^{-1}$ ein Design-Maß ist. Für das Lineare Modell gilt $\Sigma(F) = \mathbf{X}^T \mathbf{X}$. Die Influenz-Funktion IF_i misst den Einfluß auf die Statistik T , wenn man eine Beobachtung (\mathbf{x}_i, y_i) zu einer großen Stichprobe addiert. Für eine endliche Stichprobe gibt es einige Approximationen für die Influenzkurve IF_i . Die *empirische Influenzkurve* und die *stichprobenbasierte Influenzkurve* sind die Bekanntesten. Sei nun \hat{F} die empirische Verteilungsfunktion basierend auf allen Daten und \hat{F}_i die empirische Verteilungsfunktion, wenn die i -te Beobachtung weggelassen wird. Die empirische Influenzkurve EIK_i erhält man, wenn man \hat{F} für F und $\hat{\beta}$ für $T(\hat{F})$ einsetzt:

$$\begin{aligned} EIK_i &= n(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i^T (y_i - \mathbf{x}_i^T \hat{\beta}) \\ &= n(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i^T e_i. \end{aligned} \quad (3.41)$$

Die empirische Influenzkurve EIK_{-i} mit gelöschter i -ter Beobachtung erhält man, wenn man \hat{F}_{-i} für F und $\hat{\beta}_{-i}$ für $T(\hat{F}_{-i})$ in (3.37) einsetzt. Man erhält:

$$\begin{aligned} EIK_{-i} &= (n-1)(\mathbf{X}_{-i}^T \mathbf{X}_{-i})^{-1} \mathbf{x}_i^T (y_i - \mathbf{x}_i^T \hat{\beta}_{-i}) \\ &= n(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i^T \frac{e_i}{(1-h_{ii})^2}. \end{aligned} \quad (3.42)$$

Die stichprobenbasierte Influenzkurve SIK_i erhält man hingegen, wenn man den Grenzwert in (3.37) weglässt und \hat{F} für F , $\hat{\beta}$ für $T(\hat{F})$ und $-\frac{1}{n-1}$ für ϵ einsetzt:

$$\begin{aligned} SIK_i &= (n-1)(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i^T(y_i - \mathbf{x}_i^T\hat{\beta}_{-i}) \\ &= (n-1)(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i^T \frac{e_i}{1-h_{ii}}. \end{aligned} \quad (3.43)$$

Wie man sieht, ist SIK_i einfacher zu interpretieren, da sie proportional zu der Veränderung zwischen $\hat{\beta}$ und $\hat{\beta}_{-i}$ ist, wie mit Hilfe von Gleichung (3.17) gesehen werden kann. EIK_{-i} ist dagegen empfindlicher gegenüber von extremen Werten von h_{ii} .

Desweiteren ist es möglich, den Effekt einer infinitesimalen Änderung anzusehen, indem man $\hat{\beta}_i(\mathbf{w})$ nach w ableitet:

$$\begin{aligned} \hat{\beta}'_i(\mathbf{w}) = \frac{\partial}{\partial w}\hat{\beta}_i(\mathbf{w}) &= -\frac{-((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i e_i)(1 - (1-w)h_{ii}) - ((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i(1-w)e_i)h_{ii}}{(1 - (1-w)h_{ii})^2} \\ &= \frac{(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i e_i - (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i e_i(1-w)h_{ii} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i e_i(1-w)h_{ii}}{(1 - (1-w)h_{ii})^2} \\ &= \frac{(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i e_i}{(1 - (1-w)h_{ii})^2}. \end{aligned}$$

Insbesondere gilt

$$\hat{\beta}'_i(1) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i e_i, \quad (3.44)$$

$$\hat{\beta}'_i(0) = \frac{(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i e_i}{(1-h_{ii})^2}. \quad (3.45)$$

Eine Auswertung für $w = 1$ beschreibt lokale Veränderungen in $\hat{\beta}_i(\mathbf{w})$ für die eigentliche Kleinste-Quadrate-Lösung und abgesehen von n ist dieser Ausdruck wieder eine Influenzkurve, nämlich die empirische Influenzkurve EIK_i . Für Zwecke der Diagnostik dient diese Funktion nicht besonders, da der gefittete Wert für eine extreme Beobachtung gegen den ursprünglichen Wert geht, falls $w \rightarrow 1$, und infinitesimale Änderungen in diesem Punkt verändern den gefitteten Wert nicht genügend, um signifikante Änderungen in den Koeffizienten zu verursachen.

Eine Auswertung für $w = 0$ hingegen beschreibt die Änderungen in $\hat{\beta}_i(\mathbf{w})$ ohne die i -te Beobachtung. Dieser Ausdruck beschreibt abgesehen von $n-1$ eine der vorher besprochenen Influenzkurve, nämlich die empirische Influenzkurve ohne i -te Beobachtung EIK_{-i} . Diese Funktion ist empfindlicher gegenüber extremen y -Werten und extremen Kovariablenwerten. Dies kann man sich durch Betrachten der Formel nochmal verdeutlichen, denn im Nenner steht $(1-h_{ii})^2$ und große Werte für h_{ii} basieren auf extremen Kovariablenwerten. Für Zwecke der Diagnostik ist diese Form wünschenswerter.

Falls $\hat{\beta}_i(\mathbf{w})$ auf dem Einheitsintervall differenzierbar ist, kann man den Mittelwertsatz anwenden. Daraus ergibt sich

$$\frac{\hat{\beta}_i(1) - \hat{\beta}_i(0)}{1-0} = \frac{(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i e_i}{1-h_{ii}} = \hat{\beta}'_i(\bar{w}) \quad (3.46)$$

für ein \bar{w} , $0 < \bar{w} < 1$. Also ist die Änderung des Kleinste-Quadrate-Schätzers $\hat{\beta}$ für β bezüglich der Schätzung ohne i -te Beobachtung gegeben durch die Ableitung $\hat{\beta}'_i(\mathbf{w})$ ausgewertet an einer

Stelle im Einheitsintervall. Abgesehen von $n - 1$ beschreibt eine Auswertung für \bar{w} die stichprobenbasierte Influenzkurve SIK_i . Aufgrund letzterer Formel kann man einen Kompromiss schließen, so dass ein Gewicht \bar{w} ins Modell einfließt. Außerdem lässt sich $\hat{\beta}'_i(\bar{w})$ als Linearkombination von $\hat{\beta}'_i(1)$ und $\hat{\beta}'_i(0)$ schreiben:

$$\begin{aligned} \alpha \hat{\beta}'_i(1) + (1 - \alpha) \hat{\beta}'_i(0) &= \hat{\beta}'_i(\bar{w}) \\ \alpha (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i e_i + (1 - \alpha) \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i e_i}{(1 - h_{ii})^2} &= \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i e_i}{1 - h_{ii}}. \end{aligned}$$

Es lässt sich nach α auflösen:

$$\begin{aligned} \alpha \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i e_i (1 - h_{ii})^2}{(1 - h_{ii})^2} + (1 - \alpha) \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i e_i}{(1 - h_{ii})^2} &= \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i e_i (1 - h_{ii})}{(1 - h_{ii})^2} \\ \alpha \left(\frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i e_i (1 - h_{ii})^2}{(1 - h_{ii})^2} - \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i e_i}{(1 - h_{ii})^2} \right) &= \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i e_i (1 - h_{ii})}{(1 - h_{ii})^2} - \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i e_i}{(1 - h_{ii})^2} \\ \alpha \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i e_i}{(1 - h_{ii})^2} ((1 - h_{ii})^2 - 1) &= \frac{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i e_i}{(1 - h_{ii})^2} (1 - h_{ii} - 1) \\ \alpha &= -\frac{1}{h_{ii} - 2} \quad \text{für } 1/n \leq h_{ii} \leq 1. \end{aligned}$$

Zusammengefasst unterscheiden sich die drei Versionen der Influenzkurven (3.44), (3.45) und (3.46) im Nennerterm, welcher durch h_{ii} charakterisiert ist, da der Zähler in allen drei Diagnostiken gleich ist. Der Wert von $\hat{\beta}'_i(0)$ ist stark abhängig vom entsprechenden Wert h_{ii} der Hatmatrix und empfindlich gegenüber großen Werten der h_{ii} , wohingegen $\hat{\beta}'_i(1)$ gering empfindlich ist. Die Diagnostik $\hat{\beta}'_i(\bar{w})$ liegt zwischen diesen beiden Versionen und entspricht einem relativen Gewicht.

Kapitel 4

Generalisierte Lineare Modelle (GLM)

In diesem Kapitel werden Diagnostiken für die Generalisierten Linearen Modelle (GLM) entwickelt. Die Annahme des letzten Kapitel, dass die Y_i normalverteilt sind, wird nun verallgemeinert. Es werden jetzt alle Verteilungen für die Zielvariable zugelassen, die die *exponentielle Familie* umfassen. Dies sind z.B. Binomial-, Poisson- oder Negativ-Binomial-Verteilung. Hier wird jedoch besonderer Wert auf Poisson- und Negativ-Binomial-Verteilung gelegt, da das vorrangige Ziel die Darstellung von Regressionsdiagnostiken für Zähldaten ist. Die Definition und einige Eigenschaften der beiden Verteilungen findet man im Anhang. Einen guten Überblick über GLM's geben McCullagh und Nelder (1989). Weitere Details im Bezug auf Zähldaten beschreibt das Buch von Cameron und Trivedi (1998).

Das Problem der Skalierung, was beim klassischen Linearen Modell eine große Rolle spielt, ist hier weitgehend eingeschränkt. Normalität und Konstanz der Varianz werden nicht mehr verlangt, es muss nur die Abhängigkeitsstruktur der Varianz vom Erwartungswert bekannt sein. Additivität der Effekte kann auf einer transformierten Skala spezifiziert werden, wenn nötig.

Desweiteren ist es möglich, Transformationen des Erwartungswertes durchzuführen, was mit der sogenannten *Linkfunktion* erreicht wird. Damit muss die Beziehung zwischen Zielvariable und erklärender Variable nicht unbedingt linear sein.

Zunächst wird das Modell beschrieben und geeignete Annahmen getroffen. Es sollen die Komponenten der GLM's vorgestellt werden, die im folgenden eine wichtige Rolle spielen. Das Problem der Parameterschätzung wird betrachtet, was hier durch den *Iterative Weighted Least Squares Algorithmus* geschieht. Dieser wird in Kapitel 4.2 hergeleitet und stellt die Basis für spätere Diagnostikformen dar.

4.1 Modellbeschreibung und Annahmen

Zugrunde liegt das in Gleichung (2.1) beschriebene Modell. Es wird nunmehr nicht wie in den Linearen Modellen die Zielvariable als normalverteilt angenommen, sondern die Zielvariable kann jetzt eine Verteilung aus der Klasse der exponentiellen Familie besitzen. Jedoch werden einige der Eigenschaften der Normalverteilung mit der Verteilung der exponentiellen Familie geteilt.

4.1.1 Die exponentielle Familie

Die Wahrscheinlichkeitsfunktion bzw. Dichte in einem GLM gehört zur exponentiellen Familie, die nachfolgend definiert wird.

Definition 4.1 (Lineare exponentielle Familie) Eine Dichte $f(y; \theta, \phi)$ von Y gehört zur linearen exponentiellen Familie mit natürlichem oder kanonischem Parameter θ und Skalen- oder Dispersionsparameter ϕ , falls sie sich in folgender Form schreiben lässt:

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

Die Funktion $c(y, \phi)$ beschreibt eine normalisierende Konstante und die Funktionen $a(\cdot)$ und $b(\cdot)$ spezifizieren die jeweilige Verteilung.

Mit Hilfe des folgenden Lemmas lassen sich Erwartungswert und Varianz einer Verteilung der exponentiellen Familie berechnen.

Lemma 4.1 Es bezeichnet $\mathcal{L}(y; \theta, \phi) = \log f(y; \theta, \phi)$ die Log-Likelihood-Funktion. Dann gilt unter den Regularitätsbedingungen

$$E\left(\frac{\partial \mathcal{L}}{\partial \theta}\right) = 0 \quad , \quad E\left(\frac{\partial^2 \mathcal{L}}{\partial \theta^2}\right) + E\left(\left(\frac{\partial \mathcal{L}}{\partial \theta}\right)^2\right) = 0.$$

Beweis: Es gilt $\frac{\partial \mathcal{L}}{\partial \theta} = \frac{\partial \log f}{\partial \theta} = \frac{\partial f}{\partial \theta} / f$, daraus folgt unter Regularitätsbedingungen

$$E\left(\frac{\partial \mathcal{L}}{\partial \theta}\right) = E\left(\frac{\partial f}{\partial \theta} / f\right) = \int \frac{\partial f}{\partial \theta} \frac{1}{f} f \, dy = \frac{\partial}{\partial \theta} \underbrace{\int f \, dy}_{=1} = 0.$$

Ferner gilt

$$\int \frac{\partial^2 f}{\partial \theta^2} \, dy = \frac{\partial^2}{\partial \theta^2} \underbrace{\int f \, dy}_{=1} = 0, \tag{4.1}$$

$$\begin{aligned} E\left(\frac{\partial^2 \mathcal{L}}{\partial \theta^2}\right) &= E\left(\frac{\partial}{\partial \theta} \left(\frac{\partial f}{\partial \theta} / f\right)\right) = \int \left(\frac{\partial^2 f}{\partial \theta^2} f - \left(\frac{\partial f}{\partial \theta}\right)^2\right) \frac{1}{f^2} f \, dy \\ &\stackrel{(4.1)}{=} - \int \left(\frac{\partial f}{\partial \theta}\right)^2 \frac{1}{f^2} f \, dy = -E\left(\left(\frac{\partial f}{\partial \theta} / f\right)^2\right) = -E\left(\left(\frac{\partial \mathcal{L}}{\partial \theta}\right)^2\right). \end{aligned}$$

□

Es ergeben sich für den Erwartungswert und die Varianz folgende Formeln:

$$E(Y|x) = \mu = b'(\theta), \tag{4.2}$$

$$Var(Y|x) = b''(\theta)a(\phi). \tag{4.3}$$

Diese erhält man, wenn man die Log-Likelihood-Funktion der Dichte der exponentiellen Familie betrachtet:

$$\mathcal{L}(y; \theta, \phi) := \log f(y; \theta, \phi) = \frac{\theta y - b(\theta)}{a(\phi)} + c(y, \phi).$$

Es ergeben sich aus erster und zweiter Ableitung folgende Gleichungen:

$$\frac{\partial \mathcal{L}}{\partial \theta} = \frac{y - b'(\theta)}{a(\phi)} \quad \text{und} \quad \frac{\partial^2 \mathcal{L}}{\partial \theta^2} = -\frac{b''(\theta)}{a(\phi)}.$$

Mit Hilfe von Lemma 4.1 und durch einfaches Einsetzen und Ausrechnen ergeben sich die Formeln (4.2) und (4.3). Dabei wird $b''(\theta)$ als *Varianzfunktion* bezeichnet.

Beispiel 4.1 (Poissonverteilung) Die Poissonverteilung gehört zur Klasse der exponentiellen Familie, da sich die Dichte wie folgt schreiben lässt:

$$f(y; \theta, \phi) = \exp(-\mu + y \log(\mu) - \log(y!)).$$

Dabei ist $a(\phi) = 1$, $\theta = \log(\mu)$, $b(\theta) = \mu = e^{\log \mu} = e^\theta$ und $c(y, \phi) = -\log(y!)$. Für den Erwartungswert und die Varianz ergeben sich also folgende Werte:

$$E(Y) = \mu = b'(\theta) = e^\theta = \mu, \quad \text{Var}(Y) = b''(\theta)a(\phi) = e^\theta = \mu.$$

Beispiel 4.2 (Negativ-Binomial-Verteilung mit bekanntem α) Die Negativ-Binomial-Verteilung mit bekanntem α gehört zur Klasse der exponentiellen Familie, da sich die Dichte wie folgt schreiben lässt:

$$f(y; \theta, \phi) = \exp \left\{ y \log \left(\frac{\alpha \mu}{1 + \alpha \mu} \right) - \left(-\alpha^{-1} \log \left(\frac{1}{1 + \alpha \mu} \right) \right) + \log \left(\frac{\Gamma(y + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y + 1)} \right) \right\}.$$

Dabei ist

$$\begin{aligned} a(\phi) &= 1, \\ \theta &= \log \left(\frac{\alpha \mu}{1 + \alpha \mu} \right) \Leftrightarrow \mu = \alpha^{-1} \frac{e^\theta}{1 - e^\theta}, \\ b(\theta) &= -\alpha^{-1} \log \left(\frac{1}{1 + \alpha \mu} \right) = -\alpha^{-1} \log(1 - e^\theta), \\ c(y, \phi) &= \log \left(\frac{\Gamma(y + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y + 1)} \right). \end{aligned}$$

Für den Erwartungswert und die Varianz ergeben sich also folgende Werte:

$$\begin{aligned} E(Y) &= b'(\theta) = -\alpha^{-1} \frac{1}{1 - e^\theta} (-e^\theta) = \mu, \\ \text{Var}(Y) &= b''(\theta)a(\phi) = \alpha^{-1} \frac{e^\theta(1 - e^\theta) - e^\theta(-e^\theta)}{(1 - e^\theta)^2} = \alpha^{-1} \frac{e^\theta}{1 - e^\theta} \left(1 + \frac{e^\theta}{1 - e^\theta} \right) = \mu(1 + \alpha\mu). \end{aligned}$$

4.1.2 Die Komponenten eines GLM's

Es ist ein Vektor $\mathbf{y} = (y_1, \dots, y_n)^T$ von Beobachtungen aus n Komponenten, die Realisierungen der Zufallsvariable Y_i , gegeben. Die Zufallsvariablen Y_i , $i = 1, \dots, n$, seien unabhängig und der Erwartungswert von Y_i ist μ_i . Desweiteren ist der Kovariablenvektor \mathbf{x}_i , $i = 1, \dots, n$, gegeben. Man trifft also für ein GLM folgende Vereinbarungen:

- Das Modell beinhaltet die zufällige Komponente. Die Zufallsvariablen Y_i , $i = 1, \dots, n$, sind unabhängig und besitzen eine Verteilung aus der exponentiellen Familie. Der Erwartungswert ist definiert durch $E(Y_i|\mathbf{x}_i) =: \mu_i$, $i = 1, \dots, n$.
- Im Modell gibt es eine systematische Komponente. Diese wird durch die k Kovariablenvektoren $\mathbf{x}_{\cdot 1}, \dots, \mathbf{x}_{\cdot k} \in \mathbb{R}^n$ definiert und wird als sogenannter *linearer Prädiktor* bezeichnet, der durch

$$\boldsymbol{\eta} := \begin{pmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_n^T \end{pmatrix} \boldsymbol{\beta} = \mathbf{X}\boldsymbol{\beta} \quad \text{mit} \quad \mathbf{X} = (\mathbf{1}, \mathbf{x}_{\cdot 1}, \dots, \mathbf{x}_{\cdot k}) \in \mathbb{R}^{n \times p} \quad \text{und} \quad \boldsymbol{\beta} = (\beta_0, \dots, \beta_k)^T$$

bzw.

$$\eta_i := \mathbf{x}_i^T \boldsymbol{\beta} \quad \text{mit} \quad \mathbf{x}_i = (1, x_{i1}, \dots, x_{ik})^T, \quad i = 1, \dots, n$$

definiert ist.

- Nun werden beide Komponenten miteinander verbunden. Dazu beschreibt man den linearen Prädiktor η_i als Funktion des Erwartungswerts μ_i , d.h. es gilt

$$g(\mu_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta}.$$

Die Funktion g wird als *Linkfunktion* bezeichnet und verbindet die zufällige Komponente mit der systematischen, welche eine Spezifikation für den Vektor $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ aus einer Anzahl von unbekanntem zu schätzenden Parametern β_0, \dots, β_k ist.

Hier zeigen sich nochmals die wesentlichen Unterschiede zum klassischen Linearen Modell. Man kann nun nicht nur die Normalverteilung, sondern eine Verteilung aus der Klasse der exponentiellen Familie wählen. Während im Linearen Modell der Erwartungswert und der lineare Prädiktor identisch ist $\eta_i = \mu_i$, d.h. die Linkfunktion entspricht der Identität, wird in den GLM's eine beliebige monotone und differenzierbare Funktion als Linkfunktion zugelassen.

Darüberhinaus bezeichnet man die Linkfunktion als *kanonisch*, falls der lineare Prädiktor η_i gleich θ_i ist, d.h. $\eta_i = \theta_i$. Folgende Beispiele dienen zur Veranschaulichung des gerade Erwähnten.

Beispiel 4.3 (Poissonregression) *Es wird die Verallgemeinerung auf den Fall der Poissonverteilung angewendet. Die Verteilung der Y_i gegeben die Kovariablen \mathbf{x}_i ist eine Poissonverteilung, also $(Y_i|\mathbf{x}_i) \sim Poi(\mu_i)$. Die Beobachtungen seien unabhängig.*

Die Linkfunktion ist hier die Log-Link-Funktion:

$$\eta_i = \ln(\mu_i) = \log(\exp(\theta_i)) = \theta_i.$$

Der bedingte Erwartungswert der Beobachtungen Y_i gegeben die Kovariablen \mathbf{x}_i ist dann

$$E(Y_i|\mathbf{x}_i) = \exp(\eta_i) = \exp(\mathbf{x}_i^T \boldsymbol{\beta}), \quad i = 1, \dots, n.$$

Verteilung	Poisson	Negativ-Binomial
<i>Notation</i>	$Poi(\mu)$	$NB(\mu, \alpha)$
<i>Erwartungswert</i>	μ	μ
<i>Varianz</i>	μ	$\mu(1 + \alpha\mu)$
$a(\phi)$	1	1
$b(\theta)$	$\exp(\theta)$	$-\alpha^{-1} \log(1 - e^\theta)$
$c(y, \phi)$	$-\log(y!)$	$\log\left(\frac{\Gamma(y+\alpha^{-1})}{\Gamma(y+1)\Gamma(\alpha^{-1})}\right)$
η	$\log(\mu)$	$\log(\mu)$

Tabelle 4.1: Charakteristiken der Poisson- und Negativ-Binomial-Verteilung in der exponentialen Familie

Man erhält nach Zusammenfassung der Annahmen folgende Wahrscheinlichkeitsfunktion:

$$P(Y_i = y | \mathbf{x}_i) = \frac{\exp(-\exp(\mathbf{x}_i^T \boldsymbol{\beta})) \exp(\mathbf{x}_i^T \boldsymbol{\beta} y)}{y!}, \quad y = 0, 1, 2, \dots$$

Beispiel 4.4 (Negativ-Binomial-Regression mit bekanntem α) Es wird die Verallgemeinerung auf den Fall der Negativ-Binomial-Verteilung mit bekanntem α angewendet. Die Verteilung der Y_i gegeben die Kovariablen \mathbf{x}_i ist eine Negativ-Binomial-Verteilung mit bekanntem α , also $(Y_i | \mathbf{x}_i) \sim NB(\mu_i, \alpha)$ mit $\mu_i = e^{\mathbf{x}_i^T \boldsymbol{\beta}}$. Die Beobachtungen seien unabhängig. Die Linkfunktion ist hier die Log-Link-Funktion:

$$\eta_i = \log(\mu_i) = \log\left(\alpha^{-1} \frac{\exp(\theta_i)}{1 - \exp(\theta_i)}\right).$$

Der bedingte Erwartungswert der Beobachtungen Y_i gegeben die Kovariablen \mathbf{x}_i ist dann

$$E(Y_i | \mathbf{x}_i) = \exp(\eta_i) = \exp(\mathbf{x}_i^T \boldsymbol{\beta}), \quad i = 1, \dots, n.$$

Man erhält nach Zusammenfassung der Annahmen folgende Wahrscheinlichkeitsfunktion:

$$P(Y_i = y | \mathbf{x}_i) = \frac{\Gamma(y + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y + 1)} \left(\frac{1}{1 + \alpha \exp(\mathbf{x}_i^T \boldsymbol{\beta})}\right)^{\alpha^{-1}} \left(\frac{\alpha \exp(\mathbf{x}_i^T \boldsymbol{\beta})}{1 + \alpha \exp(\mathbf{x}_i^T \boldsymbol{\beta})}\right)^y$$

für $\alpha \geq 0$ und $y = 0, 1, 2, \dots$

In Tabelle 4.1 werden die wichtigsten Charakteristiken der hier betrachteten GLM's aufgelistet.

4.2 Parameterschätzung in GLM's

Nachdem man ein geeignetes Modell gewählt hat, interessiert man sich für die Schätzung der Parameter. Im Gegensatz zu den LM's, wo man mit der Kleinsten-Quadrate-Methode arbeitet, schätzt man in GLM's mit der Maximum-Likelihood-Methode. Die Kleinsten-Quadrate-Methode in LM entspricht unter der Normalverteilungsannahme genau der Maximum-Likelihood-Methode.

Da man aber nun mit GLM's arbeitet, sind diese Methoden nicht mehr äquivalent. Es wird die Maximum-Likelihood-Methode benutzt, da alle Informationen der Daten im Likelihood enthalten sind. Eine allgemeine Vorgehensweise zur ML-Schätzung findet man in Bickel und Doksum (1977) und im Anhang C. Um einen ML-Schätzer für die unbekannt Parameter zu erhalten, wird vorausgesetzt, dass die zugrundeliegende Verteilung vollständig bekannt ist und dass die Design-Matrix \mathbf{X} der Kovariablen einen vollen Spaltenrang p besitzt.

Im Fall der GLM's ergibt sich ausgehend von einer Dichte der exponentiellen Familie die Likelihood-Funktion als

$$l(\mathbf{y}; \boldsymbol{\theta}, \phi) = \prod_{i=1}^n f(y_i; \theta_i, \phi) = \prod_{i=1}^n \exp\left(\frac{\theta_i y_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)\right)$$

und die Log-Likelihood-Funktion als

$$\mathcal{L}(\mathbf{y}; \boldsymbol{\theta}, \phi) = \log l(\mathbf{y}; \boldsymbol{\theta}, \phi) = \sum_{i=1}^n \left(\frac{\theta_i y_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)\right), \quad (4.4)$$

wobei $\boldsymbol{\beta}$ durch die Beziehung $\theta_i = \theta_i(\mu_i)$, $\mu_i = h(\theta_i)$ und $g(\mu_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$ in die Funktion eingeht.

Um eine Maximalstelle $\hat{\boldsymbol{\beta}}$ für $\boldsymbol{\beta}$ zu erhalten, muss man die Log-Likelihood-Funktion nach $\boldsymbol{\beta}$ differenzieren und Null setzen. Die theoretischen Details zu den folgenden Schritten können genauer in McCullagh und Nelder (1989) nachgelesen werden. Der Einfachheit halber schreibt man $\mathcal{L} := \mathcal{L}(\mathbf{y}; \boldsymbol{\theta}, \phi)$. Man erhält dann die sogenannten *Scoregleichungen*, die zu lösen sind und wie folgt lauten:

$$s_j(\boldsymbol{\beta}) := \frac{\partial \mathcal{L}}{\partial \beta_j} = 0, \quad \forall j = 1, \dots, p. \quad (4.5)$$

Um diese zu lösen, schreibt man

$$\frac{\partial \mathcal{L}}{\partial \beta_j} = \frac{\partial \mathcal{L}}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j}.$$

Somit hat man mehrere Ableitungen zu bestimmen, um die Scoregleichungen zu erhalten. Diese werden im folgenden bestimmt. Es gilt

$$\frac{\partial \eta_i}{\partial \beta_j} = x_{ij}.$$

Um $\frac{\partial \mathcal{L}}{\partial \mu_i}$ zu bekommen, schreibt man folgendes:

$$\frac{\partial \mathcal{L}}{\partial \theta_i} = \frac{\partial \mathcal{L}}{\partial \mu_i} \frac{\partial \mu_i}{\partial \theta_i}.$$

Daraus folgt:

$$\frac{\partial \mathcal{L}}{\partial \mu_i} = \frac{\partial \mathcal{L}}{\partial \theta_i} \bigg/ \frac{\partial \mu_i}{\partial \theta_i} = \frac{y_i - b'(\theta_i)}{a(\phi)} \bigg/ b''(\theta_i) = \frac{y_i - b'(\theta_i)}{a(\phi)} \bigg/ \frac{v_i}{a(\phi)} = \frac{y_i - b'(\theta_i)}{v_i}$$

mit $v_i = \text{Var}(Y_i) = a(\phi)b''(\theta_i)$. Also ergibt sich somit:

$$\frac{\partial \mathcal{L}}{\partial \beta_j} = \frac{\partial \mathcal{L}}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = \left(\frac{y_i - b'(\theta_i)}{v_i}\right) \frac{\partial \mu_i}{\partial \eta_i} x_{ij}.$$

Daher gilt also für n Beobachtungen:

$$\mathcal{L}(\mathbf{y}; \boldsymbol{\theta}, \phi) = \sum_{i=1}^n \mathcal{L}(y_i, \mu_i, \phi) = \sum_{i=1}^n \left(\frac{y_i \theta_i - b(\theta_i)}{a(\phi)} \right) + \text{Konst. unabh. von } \boldsymbol{\beta}$$

mit $\mu_i = b'(\theta_i) = E(Y_i)$, $g(\mu_i) = \eta_i$, $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$.

Die Ableitung nach β_j lautet dann:

$$\frac{\partial \mathcal{L}(\mathbf{y}; \boldsymbol{\theta}, \phi)}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial \mathcal{L}(y_i, \mu_i, \phi)}{\partial \beta_j} = \sum_{i=1}^n \left(\frac{y_i - \mu_i}{v_i} \right) \frac{d\mu_i}{d\eta_i} x_{ij}.$$

Nun sei $w_i = \frac{1}{v_i \left(\frac{d\eta_i}{d\mu_i} \right)^2} = \left(\frac{d\mu_i}{d\eta_i} \right)^2 / v_i$ und zur Vereinfachung $\mathcal{L} := \mathcal{L}(\mathbf{y}; \boldsymbol{\theta}, \phi)$.

Folglich lauten dann die Scoregleichungen wie folgt:

$$s_j(\boldsymbol{\beta}) = \frac{\partial \mathcal{L}}{\partial \beta_j} = \sum_{i=1}^n w_i (y_i - \mu_i) \frac{d\eta_i}{d\mu_i} x_{ij} = 0, \quad \forall j,$$

und in Matrixschreibweise

$$S(\boldsymbol{\beta}) := \begin{pmatrix} s_1(\boldsymbol{\beta}) \\ \vdots \\ s_p(\boldsymbol{\beta}) \end{pmatrix} = \mathbf{X}^T \mathbf{W} \mathbf{D}^{-1} \mathbf{r} = \mathbf{0} \quad (4.6)$$

mit $\mathbf{W} = \text{diag}(w_1, \dots, w_n)$ und $\mathbf{D}^{-1} = \text{diag} \left(\frac{d\eta_1}{d\mu_1}, \dots, \frac{d\eta_n}{d\mu_n} \right)$. Diese Gleichungen sind nichtlineare Gleichungen in $\boldsymbol{\beta}$. Eine Methode zur Lösung der Gleichungen ist die *Newton-Raphson-Methode*. Eine Beschreibung dieser Methode findet man in McCullagh und Nelder (1989) (Seite 40ff) und im Anhang D. Darüberhinaus wird die *Fisher-Scoring-Methode* angewandt, die ebenfalls im Anhang D beschrieben wird. Diese ist ein iterativer Algorithmus, so dass sich eine Folge von $\hat{\boldsymbol{\beta}}^{t+1}$ von Lösungen ergibt. Es sei also $\hat{\boldsymbol{\beta}}^t$ die momentane Schätzung der Lösung zu $S(\boldsymbol{\beta}) = \mathbf{0}$, dann ist die neue Schätzung $\hat{\boldsymbol{\beta}}^{t+1}$ durch folgende Gleichung definiert:

$$\hat{\boldsymbol{\beta}}^{t+1} = \hat{\boldsymbol{\beta}}^t + FI^{-1}(\hat{\boldsymbol{\beta}}^t) S(\hat{\boldsymbol{\beta}}^t),$$

wobei $FI(\boldsymbol{\beta}) := \mathbf{X}^T \mathbf{W} \mathbf{X}$ die Fisher-Informationsmatrix mit $\mathbf{W} = \text{diag}(w_1, \dots, w_n)$. Es gilt nun

$$FI(\hat{\boldsymbol{\beta}}^t) \hat{\boldsymbol{\beta}}^{t+1} = FI(\hat{\boldsymbol{\beta}}^t) \hat{\boldsymbol{\beta}}^t + S(\hat{\boldsymbol{\beta}}^t) \quad (4.7)$$

und in Komponentenschreibweise:

$$\begin{aligned} (FI(\hat{\boldsymbol{\beta}}^t) \hat{\boldsymbol{\beta}}^{t+1})_j &= \sum_{l=1}^p FI_{jl}(\hat{\boldsymbol{\beta}}^t) \hat{\beta}_l^t + s_j(\hat{\boldsymbol{\beta}}^t) \\ &= \sum_{l=1}^p \sum_{i=1}^n w_i^t x_{ij} x_{il} \hat{\beta}_l^t + \sum_{i=1}^n w_i^t (y_i - \mu_i^t) \frac{d\eta_i^t}{d\mu_i^t} x_{ij} \\ &= \sum_{i=1}^n w_i^t x_{ij} \left[\underbrace{\sum_{l=1}^p x_{il} \hat{\beta}_l^t}_{=\mathbf{x}_i^T \hat{\boldsymbol{\beta}}^t = \eta_i^t} + (y_i - \mu_i^t) \frac{d\eta_i^t}{\mu_i^t} \right], \end{aligned}$$

wobei $w_i^t := w_i(\hat{\boldsymbol{\beta}}^t)$, $\mu_i^t := \mu_i(\hat{\boldsymbol{\beta}}^t)$ und $\eta_i^t := \eta_i(\hat{\boldsymbol{\beta}}^t)$. Definiert man nun

$$Z_i^t := \eta_i^t + (y_i - \mu_i^t) \frac{d\eta_i^t}{d\mu_i^t},$$

dann gilt einerseits

$$(FI(\hat{\boldsymbol{\beta}}^t)\hat{\boldsymbol{\beta}}^{t+1})_j = \sum_{i=1}^n w_i^t x_{ij} Z_i^t.$$

In dieser Regression ist Z_i^t die abhängige Variable. Sie wird als *adjusted dependent variable* bezeichnet und hängt ebenso wie die Gewichtungsmatrix \mathbf{W} von der momentanen Schätzung $\hat{\boldsymbol{\beta}}^t$ ab, da in jedem neuen Schritt die Variablen verändert werden. Andererseits gilt aber auch

$$\begin{aligned} (FI(\hat{\boldsymbol{\beta}}^t)\hat{\boldsymbol{\beta}}^{t+1})_j &= \sum_{l=1}^p FI_{jl}(\hat{\boldsymbol{\beta}}^t)\hat{\beta}_l^{t+1} \\ &= \sum_{l=1}^p \left(\sum_{i=1}^n w_i^t x_{il} x_{ij} \right) \hat{\beta}_l^{t+1} \\ &= \sum_{i=1}^n w_i^t x_{ij} \underbrace{\left(\sum_{l=1}^p x_{il} \hat{\beta}_l^{t+1} \right)}_{=\mathbf{x}_i^T \hat{\boldsymbol{\beta}}^{t+1} = \eta_i^{t+1}} \\ &= \sum_{i=1}^n w_i^t x_{ij} \eta_i^{t+1}. \end{aligned}$$

Also gilt nach (4.7)

$$\sum_{i=1}^n w_i^t x_{ij} \eta_i^{t+1} = \sum_{i=1}^n w_i^t x_{ij} Z_i^t, \quad \forall j = 1, \dots, p. \quad (4.8)$$

Die Gleichungen in (4.8) entsprechen den Normalgleichungen einer gewichteten Regression, welche schon für gewichtete LM in Gleichung (2.12) besprochen wurden. Insbesondere sind die Gleichungen in (4.8) die Normalgleichungen einer gewichteten Kleinst-Quadrate-Schätzung mit Zielvariable Z_i^t , Kovariablen $\mathbf{x}_{\cdot 1}, \dots, \mathbf{x}_{\cdot p}$ und Gewichten w_i^t . Eine Schätzung $\hat{\boldsymbol{\beta}}^{t+1}$ lautet dann genau

$$\hat{\boldsymbol{\beta}}^{t+1} = (\mathbf{X}^T \mathbf{W}^t \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^t \mathbf{Z}^t, \quad (4.9)$$

wobei

$$\begin{aligned} \mathbf{W}^t &:= \mathbf{W}(\hat{\boldsymbol{\beta}}^t) := \text{diag}(w_1^t, \dots, w_n^t), \\ \mathbf{Z}^t &= (Z_1^t, \dots, Z_n^t) = \left(\eta_1^t + (y_1 - \mu_1^t) \frac{d\eta_1^t}{d\mu_1^t}, \dots, \eta_n^t + (y_n - \mu_n^t) \frac{d\eta_n^t}{d\mu_n^t} \right), \end{aligned}$$

und mit $(\mathbf{D}^t)^{-1} := \mathbf{D}^{-1}(\hat{\boldsymbol{\beta}}^t) = \text{diag}\left(\frac{d\eta_1^t}{d\mu_1^t}, \dots, \frac{d\eta_n^t}{d\mu_n^t}\right)$ gilt

$$\begin{aligned}\hat{\beta}^{t+1} &= (\mathbf{X}^T \mathbf{W}^t \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^t (\mathbf{X} \hat{\beta}^t + (\mathbf{D}^t)^{-1} (\mathbf{y} - \boldsymbol{\mu}^t)) \\ &= \hat{\beta}^t + (\mathbf{X}^T \mathbf{W}^t \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^t (\mathbf{D}^t)^{-1} (\mathbf{y} - \boldsymbol{\mu}^t).\end{aligned}\quad (4.10)$$

Daher spricht man auch vom *IWLS (Iterative Weighted Least Squares) Algorithmus*, der in Hinkley, Reid, und Snell (1991) (Seite 63) schrittweise beschrieben wird. Dieser wird nachfolgend nochmals kurz erklärt:

IWLS-Algorithmus

Schritt 1: Sei $\hat{\beta}^t$ die momentane Schätzung vom MLE $\hat{\beta}$. Dann bilde

- $\eta_i^t := \mathbf{x}_i^T \hat{\beta}^t$, $i = 1, \dots, n$ (momentaner linearer Prädiktor)
- $\mu_i^t := g^{-1}(\eta_i^t)$, $i = 1, \dots, n$ (momentan gefitteter Mittelwert)
- $Z_i^t := \eta_i^t + (y_i - \mu_i^t) \left(\frac{d\eta_i}{d\mu_i} \Big|_{\mu_i = \mu_i^t} \right)$
- $w_i^t := \left(v_i^t \left(\frac{d\eta_i}{d\mu_i} \Big|_{\mu_i = \mu_i^t} \right)^2 \right)^{-1}$, $v_i^t = \widehat{Var}(Y_i) = a(\phi) b''(\theta_i) |_{\theta_i = \theta_i^t}$, $\theta_i^t = (b')^{-1}(\mu_i^t)$

Schritt 2: Regressiere Z_i^t auf x_{i1}, \dots, x_{ip} mit Gewicht w_i^t , um eine neue Schätzung $\hat{\beta}^{t+1}$ zu finden und gehe zu Schritt 1, falls die Veränderung $\|\hat{\beta}^{t+1} - \hat{\beta}^t\|^1$ nicht klein genug ist.

Bemerkungen:

1. Falls die Veränderung $\|\hat{\beta}^{t+1} - \hat{\beta}^t\|$ klein genug im $(t+1)$ -ten Schritt, so ist $\hat{\beta} := \hat{\beta}^{t+1}$ der Schätzer für β nach dem IWLS. Die Werte für Z_i , $\hat{\eta}_i$, $\hat{\mu}_i$, \hat{v}_i , w_i und somit auch die Matrizen \mathbf{W} und \mathbf{D}^{-1} hängen von dem geschätzten Wert $\hat{\beta}$ ab.
2. Z_i ist eine linearisierte Form der Linkfunktion, da mit Taylor folgendes gilt:

$$\begin{aligned}g(y_i) &\approx g(\hat{\mu}_i) + (y_i - \hat{\mu}_i) \underbrace{\frac{g'(\hat{\mu}_i)}{\frac{d\eta_i}{d\mu_i} \Big|_{\beta = \hat{\beta}}}}_{= \hat{Z}_i} \\ &\implies Z_i \approx g(y_i) \quad \text{in erster Ordnung.}\end{aligned}$$

3. Falls man $\hat{\eta}_i$, $\hat{\mu}_i$ als fest und bekannt annimmt, dann ist

$$\begin{aligned}Var(Z_i) &= Var \left(\hat{\eta}_i + (y_i - \hat{\mu}_i) \frac{d\eta_i}{d\mu_i} \Big|_{\beta = \hat{\beta}} \right) \\ &\approx \left(\frac{d\eta_i}{d\mu_i} \Big|_{\beta = \hat{\beta}} \right)^2 \underbrace{Var(y_i - \hat{\mu}_i)}_{\hat{v}_i} = w_i^{-1}\end{aligned}$$

¹ $\|\cdot\|$ bezeichnet die euklidische Norm.

4. Häufig kann man die Daten als Startwerte benutzen, wobei

$$\mu_i^0 := y_i \implies \eta_i^0 := g(\mu_i^0).$$

Falls aber z.B. $y_i = 0$, muss man die Startwerte verändern, um nicht $\log(0)$ zu bekommen.

5. Es bleibt nach Errechnung eines Schätzers $\hat{\beta}$ der Vollständigkeit halber noch die negative Definitheit der Hessematrix $H(\beta) = \frac{\partial^2 \mathcal{L}}{\partial \beta \partial \beta^T}$ zu überprüfen. Ist diese an der Stelle $\beta = \hat{\beta}$ negativ definit, so ist der Schätzer $\hat{\beta}$ wirklich ein Maximum.

Zur Veranschaulichung der Vorgehensweise des IWLS sollen wieder die Beispiele für die Poisson- und die Negativ-Binomial-Verteilung dienen.

Beispiel 4.5 (Poissonregression) Hier wird der IWLS anhand der Poissonverteilung besprochen. Die Poissonverteilung gehört zur exponentiellen Familie.

$$f(y; \theta, \phi) = \exp(-\mu + y \log(\mu) - \log(y!))$$

mit $\theta = \log(\mu)$, $a(\phi) = 1$, $b(\theta) = \mu$ und $c(y, \phi) = -\log(y!)$. Um $b(\theta)$ in Abhängigkeit von θ zu erhalten, löst man zuerst den Ausdruck für θ nach μ auf und setzt dann in $b(\theta)$ ein. Man erhält sodann folgendes

$$\begin{aligned} \theta = \log(\mu) &\Leftrightarrow \mu = e^\theta \\ b(\theta) &= e^\theta. \end{aligned}$$

Desweiteren wird der Log-Link verwendet. Es soll also gelten:

$$\begin{aligned} \eta &= \log(\mu) = \log(e^\theta) = \theta, \\ \mu &= \exp(\eta). \end{aligned}$$

Da $\eta = \theta$, ist die Linkfunktion kanonisch. Die Log-Likelihood-Funktion für n Beobachtungen ist gegeben durch

$$\mathcal{L}(\beta, \alpha) = \sum_{i=1}^n y_i \log(\mu_i) - \mu_i - \log(y_i!).$$

Die Scoregleichungen lauten

$$\begin{aligned} s_j(\beta) = \frac{\partial \mathcal{L}}{\partial \beta_j} &= \frac{\partial}{\partial \beta_j} \left[\sum_{i=1}^n y_i \log(e^{\mathbf{x}_i^T \beta}) - e^{\mathbf{x}_i^T \beta} - \log(y_i!) \right] \\ &= \sum_{i=1}^n y_i x_{ij} - \underbrace{e^{\mathbf{x}_i^T \beta}}_{=\mu_i} x_{ij} \\ &= \sum_{i=1}^n (y_i - \mu_i) x_{ij}. \end{aligned}$$

Um die Elemente der Hessematrix bzw. der Informationsmatrix zu bekommen, differenziert man letzteren Ausdruck nochmals nach β :

$$\begin{aligned}\frac{\partial^2 \mathcal{L}}{\partial \beta_r \partial \beta_s} &= \frac{\partial}{\partial \beta_r} \left[\sum_{i=1}^n (y_i - e^{\mathbf{x}_i^T \beta}) x_{is} \right] \\ &= \sum_{i=1}^n -e^{\mathbf{x}_i^T \beta} x_{is} x_{ir} \\ &= \sum_{i=1}^n -x_{ir} x_{is} \mu_i.\end{aligned}$$

Hier soll das Fisher-Scoring-Verfahren angewendet werden. Also wird die erwartete Informationsmatrix benötigt. Die Elemente ergeben sich aus

$$E \left[- \sum_{i=1}^n -x_{ir} x_{is} e^{\mathbf{x}_i^T \beta} \right] = \sum_{i=1}^n x_{ir} x_{is} \mu_i \quad (4.11)$$

Es sei $w_i = \left(\frac{d\mu_i}{d\eta_i} \right)^2 / v_i$ mit

$$\begin{aligned}v_i &= \text{Var}(Y_i) = a(\phi) b''(\theta_i) = e^\theta = \mu_i \\ \frac{d\eta_i}{d\mu_i} &= \frac{1}{\mu_i} = \frac{1}{e^{\mathbf{x}_i^T \beta}}.\end{aligned}$$

Also erhält man für w_i :

$$\begin{aligned}w_i &= \left(e^{\mathbf{x}_i^T \beta} \right)^2 / e^{\mathbf{x}_i^T \beta} \\ &= e^{\mathbf{x}_i^T \beta} = \mu_i.\end{aligned}$$

Also ist der Schätzer $\hat{\beta}$ für β gegeben durch

$$\hat{\beta} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{Z},$$

wobei

$$\begin{aligned}\mathbf{W} &:= \mathbf{W}(\hat{\beta}) = \text{diag}(w_1, \dots, w_n) = \text{diag}(\hat{\mu}_1, \dots, \hat{\mu}_n), \\ \mathbf{Z} &:= \mathbf{Z}(\hat{\beta}) = (Z_1, \dots, Z_n)^T = \left(\hat{\eta}_1 + (y_1 - \hat{\mu}_1) \frac{1}{\hat{\mu}_1}, \dots, \hat{\eta}_n + (y_n - \hat{\mu}_n) \frac{1}{\hat{\mu}_n} \right)^T = \hat{\boldsymbol{\eta}} + \mathbf{D}^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}}), \\ \mathbf{D}^{-1} &:= \mathbf{D}^{-1}(\hat{\beta}) = \text{diag} \left(\frac{1}{\hat{\mu}_1}, \dots, \frac{1}{\hat{\mu}_n} \right).\end{aligned}$$

Beispiel 4.6 (Negativ-Binomial-Regression mit bekanntem α) Hier wird der IWLS anhand der Negativ-Binomial-Verteilung besprochen. Die Negativ-Binomial-Verteilung gehört zur exponentiellen Familie, falls α bekannt ist.

$$f(y; \theta, \phi) = \exp \left\{ y \log \left(\frac{\alpha \mu}{1 + \alpha \mu} \right) - \left(-\alpha^{-1} \log \left(\frac{1}{1 + \alpha \mu} \right) \right) + \log \left(\frac{\Gamma(y + \alpha^{-1})}{\Gamma(\alpha^{-1}) \Gamma(y + 1)} \right) \right\}$$

mit $\theta = \log\left(\frac{\alpha\mu}{1+\alpha\mu}\right)$, $a(\phi) = 1$, $b(\theta) = -\alpha^{-1}\log\left(\frac{1}{1+\alpha\mu}\right)$ und $c(y, \phi) = \log\left(\frac{\Gamma(y+\alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y+1)}\right)$. Um $b(\theta)$ in Abhängigkeit von θ zu erhalten, löst man zuerst den Ausdruck für θ nach μ auf und setzt dann in $b(\theta)$ ein. Man erhält sodann folgendes

$$\begin{aligned}\theta = \log\left(\frac{\alpha\mu}{1+\alpha\mu}\right) &\Leftrightarrow e^\theta = \frac{\alpha\mu}{1+\alpha\mu} \\ &\Leftrightarrow \alpha\mu(1-e^\theta) = e^\theta \\ &\Leftrightarrow \mu = \alpha^{-1}\frac{e^\theta}{1-e^\theta}, \\ b(\theta) &= -\alpha^{-1}\log\left(\frac{1}{1+\alpha\alpha^{-1}\frac{e^\theta}{1-e^\theta}}\right) \\ &= -\alpha^{-1}\log\left(\frac{1}{\frac{1-e^\theta+e^\theta}{1-e^\theta}}\right) \\ &= -\alpha^{-1}\log(1-e^\theta).\end{aligned}$$

Desweiteren wird der Log-Link verwendet. Es soll also gelten:

$$\begin{aligned}\eta &= \log(\mu) = \log\left(\alpha^{-1}\frac{e^\theta}{1-e^\theta}\right), \\ \mu &= \exp(\eta).\end{aligned}$$

Die Log-Likelihood-Funktion für n Beobachtungen ist gegeben durch

$$\mathcal{L}(\boldsymbol{\beta}, \alpha) = \sum_{i=1}^n y_i \log\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right) + \alpha^{-1}\log\left(\frac{1}{1+\alpha\mu_i}\right) + \log\left(\frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y_i + 1)}\right).$$

Die Scoregleichungen lauten

$$\begin{aligned}s_j(\boldsymbol{\beta}) = \frac{\partial \mathcal{L}}{\partial \beta_j} &= \frac{\partial}{\partial \beta_j} \left[\sum_{i=1}^n y_i \log\left(\frac{\alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}}\right) + \alpha^{-1} \log\left(1 - \exp\left(\log\left(\frac{\alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}}\right)\right)\right) + \right. \\ &\quad \left. + \log\left(\frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y_i + 1)}\right) \right] \\ &= \sum_{i=1}^n y_i \frac{1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{\alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \frac{\alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}} x_{ij} (1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}) - (\alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}})^2 x_{ij}}{(1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}})^2} + \\ &\quad + \alpha^{-1} (1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}) \frac{-\alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}} x_{ij}}{(1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}})^2} \\ &= \sum_{i=1}^n y_i \frac{1}{1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}} x_{ij} - \frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}} x_{ij}\end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n \frac{1}{1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \left(y_i - \underbrace{e^{\mathbf{x}_i^T \boldsymbol{\beta}}}_{\mu_i} \right) x_{ij} \\
&= \sum_{i=1}^n \frac{1}{1 + \alpha \mu_i} (y_i - \mu_i) x_{ij}.
\end{aligned}$$

Um die Elemente der Hessematrix bzw. der Informationsmatrix zu bekommen, differenziert man letzteren Ausdruck nochmals nach $\boldsymbol{\beta}$:

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}}{\partial \beta_r \partial \beta_s} &= \frac{\partial}{\partial \beta_r} \left[\sum_{i=1}^n \frac{1}{1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}} (y_i - e^{\mathbf{x}_i^T \boldsymbol{\beta}}) x_{is} \right] \\
&= \sum_{i=1}^n \frac{-\alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}} x_{ir}}{(1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}})^2} (y_i - e^{\mathbf{x}_i^T \boldsymbol{\beta}}) x_{is} + \frac{1}{1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}} (-e^{\mathbf{x}_i^T \boldsymbol{\beta}} x_{ir}) x_{is} \\
&= \sum_{i=1}^n \frac{-\alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}} x_{ir} x_{is} (y_i - e^{\mathbf{x}_i^T \boldsymbol{\beta}}) - (1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}) x_{ir} x_{is} e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{(1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}})^2} \\
&= \sum_{i=1}^n -\frac{x_{ir} x_{is} e^{\mathbf{x}_i^T \boldsymbol{\beta}} (\alpha y_i + 1)}{(1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}})^2} \\
&= \sum_{i=1}^n -\frac{x_{ir} x_{is} \mu_i (\alpha y_i + 1)}{(1 + \alpha \mu_i)^2}.
\end{aligned}$$

Hier soll das Fisher-Scoring Verfahren angewendet werden. Also wird die erwartete Informationsmatrix benötigt. Die Elemente ergeben sich aus

$$\begin{aligned}
E \left[-\sum_{i=1}^n -\frac{x_{ir} x_{is} e^{\mathbf{x}_i^T \boldsymbol{\beta}} (\alpha y_i + 1)}{(1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}})^2} \right] &= \sum_{i=1}^n \frac{x_{ir} x_{is} e^{\mathbf{x}_i^T \boldsymbol{\beta}} (\alpha E(y_i) + 1)}{(1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}})^2} \\
&= \sum_{i=1}^n \frac{x_{ir} x_{is} \mu_i (\alpha \mu_i + 1)}{(1 + \alpha \mu_i)^2} \\
&= \sum_{i=1}^n \frac{x_{ir} x_{is} \mu_i}{1 + \alpha \mu_i}.
\end{aligned}$$

Es sei $w_i = \left(\frac{d\mu_i}{d\eta_i} \right)^2 / v_i$ mit

$$v_i = \text{Var}(Y_i) = a(\phi) b''(\theta_i) = \alpha^{-1} \frac{\frac{\alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}}}{\left(1 - \frac{\alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right)^2}$$

$$\frac{d\eta_i}{d\mu_i} = \frac{1}{\mu_i} = \frac{1 - \frac{\alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}}}{\alpha^{-1} \frac{\alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}}}.$$

Also erhält man für w_i :

$$\begin{aligned} w_i &= \left(\frac{\frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}}}{1 - \frac{\alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}}} \right)^2 \frac{\left(1 - \frac{\alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right)^2}{\frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}}} \\ &= \frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + \alpha e^{\mathbf{x}_i^T \boldsymbol{\beta}}} = \frac{\mu_i}{1 + \alpha \mu_i}. \end{aligned}$$

Also ist der Schätzer $\hat{\boldsymbol{\beta}}$ für $\boldsymbol{\beta}$, gegeben durch

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{Z},$$

wobei

$$\begin{aligned} \mathbf{W} &:= \mathbf{W}(\hat{\boldsymbol{\beta}}) = \text{diag}(w_1, \dots, w_n) = \text{diag} \left(\frac{\hat{\mu}_1}{1 + \alpha \hat{\mu}_1}, \dots, \frac{\hat{\mu}_n}{1 + \alpha \hat{\mu}_n} \right), \\ \mathbf{Z} &:= \mathbf{Z}(\hat{\boldsymbol{\beta}}) = (Z_1, \dots, Z_n)^T = \left(\hat{\eta}_1 + (y_1 - \hat{\mu}_1) \frac{1}{\hat{\mu}_1}, \dots, \hat{\eta}_n + (y_n - \hat{\mu}_n) \frac{1}{\hat{\mu}_n} \right)^T = \hat{\boldsymbol{\eta}} + \mathbf{D}^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}}), \\ \mathbf{D}^{-1} &:= \mathbf{D}^{-1}(\hat{\boldsymbol{\beta}}) = \text{diag} \left(\frac{1}{\hat{\mu}_1}, \dots, \frac{1}{\hat{\mu}_n} \right). \end{aligned}$$

4.3 Bestimmung der Anpassungsgüte und Modellwahl

Die Anpassung der Daten kann als Methode angesehen werden, wo man die Daten y_1, \dots, y_n durch geschätzte Werte $\hat{\mu}_1, \dots, \hat{\mu}_n$ ersetzt. Im allgemeinen gilt nicht $\hat{\mu}_i = y_i \forall i$, daher benötigt man ein *Diskrepanzmaß* oder ein *Goodness-of-Fit-Maß*, um die Anpassungsgüte zu beurteilen. Für n Datenpunkte kann man bis zu n Parameter schätzen. Beim *Nullmodell* wird ein Parameter verwendet, der den gemeinsamen Erwartungswert repräsentiert. Dies erscheint aber ein zu einfaches Modell zu sein. Hingegen benutzt das *gesättigte Modell* (*saturated model*) n Parameter, wobei $\hat{\mu}_i = y_i$ gesetzt wird. Hier tritt zwar kein Fehler auf, das Modell ist aber nicht informativ. Jedoch kann das volle Modell als Basis zur Berechnung eines Modells mit p Parametern verwendet werden. Es gilt dann für den Log-Likelihood für n Beobachtungen mit $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta} = g(\mu_i)$, $b'(\theta_i) = \mu_i$, $\theta_i = h(\mu_i)$, h Umkehrfunktion von b' :

$$\begin{aligned} \mathcal{L}(\mathbf{y}; \boldsymbol{\beta}, \phi) &= \sum_{i=1}^n \left(\frac{y_i \theta_i - b(\theta_i)}{a(\phi)} - c(y_i, \phi) \right) \\ &= \sum_{i=1}^n \left(\frac{y_i h(\mu_i) - b(h(\mu_i))}{a(\phi)} - c(y_i, \phi) \right) \\ &= \mathcal{L}(\mathbf{y}; \boldsymbol{\mu}, \phi). \end{aligned}$$

Dies wird als sogenannte *mean parameterization* bezeichnet.

Nun sei

$$\begin{aligned} \mathcal{L}(\mathbf{y}; \hat{\boldsymbol{\mu}}, \phi) &= \text{Log-Likelihood maximiert über } \boldsymbol{\beta}, \\ \mathcal{L}(\mathbf{y}; \mathbf{y}, \phi) &= \text{maximaler Log-Likelihood, d.h. im vollen Modell mit } n \text{ Parametern } (\hat{\mu}_i = y_i \forall i), \end{aligned}$$

wobei $\hat{\mu}_i = g^{-1}(\mathbf{x}_i^T \hat{\boldsymbol{\beta}})$, $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \dots, \hat{\mu}_n)^T$ und ϕ bekannt.

Dann gilt

$$-2[\mathcal{L}(\mathbf{y}; \hat{\boldsymbol{\mu}}, \phi) - \mathcal{L}(\mathbf{y}; \mathbf{y}, \phi)] = 2 \sum_{i=1}^n \frac{y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)}{a(\phi)},$$

wobei $\tilde{\theta}_i = h(y_i)$ (n Parameter), $\hat{\theta}_i = h(\hat{\mu}_i)$ (p Parameter).

Falls $a(\phi) = \frac{\phi}{w}$, dann gilt

$$-2[\mathcal{L}(\mathbf{y}; \hat{\boldsymbol{\mu}}, \phi) - \mathcal{L}(\mathbf{y}; \mathbf{y}, \phi)] = 2w \sum_{i=1}^n \frac{y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)}{\phi} =: \frac{D(\mathbf{y}, \hat{\boldsymbol{\mu}})}{\phi}.$$

Der Ausdruck $D := D(\mathbf{y}, \hat{\boldsymbol{\mu}})$ entspricht der *Devianz* im GLM, welche unabhängig vom Dispersionsparameter ϕ ist. Für die Normalverteilung ist die Devianz gerade die *SQR*, wie in Gleichung (2.4) definiert. Da die Devianz ein wichtiger Bestandteil der Diagnostik ist, wird die Formel hier nochmal explizit angegeben:

$$\begin{aligned} D(\mathbf{y}, \hat{\boldsymbol{\mu}}) &= -2[\mathcal{L}(\mathbf{y}; \hat{\boldsymbol{\mu}}) - \mathcal{L}(\mathbf{y}; \mathbf{y})] \\ &= 2 \sum_{i=1}^n [\mathcal{L}(y_i; y_i) - \mathcal{L}(y_i; \hat{\mu}_i)]. \end{aligned} \quad (4.12)$$

Der Ausdruck

$$D^*(\mathbf{y}, \hat{\boldsymbol{\mu}}) := \frac{D(\mathbf{y}, \hat{\boldsymbol{\mu}})}{\phi}$$

wird als *skalierte Devianz* bezeichnet.

Beispiel 4.7 (Poissonfall) Sei $Y_i \sim Poi(\mu_i)$ mit $\mu_i = e^{\mathbf{x}_i^T \boldsymbol{\beta}}$. Dann gilt

$$\begin{aligned} D(\mathbf{y}, \hat{\boldsymbol{\mu}}) &= -2 \sum_{i=1}^n \mathcal{L}(y_i; \hat{\mu}_i) - \mathcal{L}(y_i; y_i) \\ &= -2 \sum_{i=1}^n [-\hat{\mu}_i + y_i \log(\hat{\mu}_i) - \log(y_i!) - (y_i + y_i \log(y_i) - \log(y_i!))] \\ &= 2 \sum_{i=1}^n \left[y_i \log\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i) \right], \end{aligned}$$

wobei $y \log(y) = 0$, falls $y = 0$.

Beispiel 4.8 (Negativ-Binomial-Fall mit bekanntem α) Sei $Y_i \sim NB(\mu_i, \alpha)$ mit $\mu_i = e^{\mathbf{x}_i^T \boldsymbol{\beta}}$ und α bekannt. Dann gilt

$$\begin{aligned} D(\mathbf{y}, \hat{\boldsymbol{\mu}}) &= -2 \sum_{i=1}^n \mathcal{L}(y_i; \hat{\mu}_i) - \mathcal{L}(y_i; y_i) \\ &= -2 \sum_{i=1}^n \left[y_i \log\left(\frac{\alpha \hat{\mu}_i}{1 + \alpha \hat{\mu}_i}\right) - \left(-\alpha^{-1} \log\left(\frac{1}{1 + \alpha \hat{\mu}_i}\right)\right) + \right. \\ &\quad \left. + \log\left(\frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y_i + 1)}\right) - \left(y_i \log\left(\frac{\alpha y_i}{1 + \alpha y_i}\right) - \right. \right. \end{aligned}$$

$$\begin{aligned}
& - \left(-\alpha^{-1} \log \left(\frac{1}{1 + \alpha y_i} \right) + \log \left(\frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1}) \Gamma(y_i + 1)} \right) \right) \\
= & 2 \sum_{i=1}^n \left[y_i \left(\log \left(\frac{y_i}{\hat{\mu}_i} \right) - \log \left(\frac{1 + \alpha y_i}{1 + \alpha \hat{\mu}_i} \right) \right) + \alpha^{-1} \log \left(\frac{1 + \alpha y_i}{1 + \alpha \hat{\mu}_i} \right) \right] \\
= & 2 \sum_{i=1}^n \left[y_i \log \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i + \alpha^{-1}) \log \left(\frac{1 + \alpha y_i}{1 + \alpha \hat{\mu}_i} \right) \right], \tag{4.13}
\end{aligned}$$

wobei $y \log(y) = 0$, falls $y = 0$. Dabei ist zu beachten, dass die Devianz $D(\mathbf{y}, \hat{\boldsymbol{\mu}})$ von α abhängt.

Ein weiteres wichtiges Diskrepanzmaß ist die generalisierte Pearson χ^2 -Statistik, welche durch

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{v}_i}$$

definiert ist. Der Wert $\hat{v}_i = b''(\hat{\theta}_i)$ ist die geschätzte Varianzfunktion für die entsprechende Verteilung. Da für die Spezialfälle von Poisson- und Negativ-Binomial-Verteilung $a(\phi) = 1$ gilt, ist die geschätzte Varianzfunktion gleich der Varianz der Y_i , d.h. $\hat{v}_i = \widehat{\text{Var}}(Y_i)$.

Beispiel 4.9 (Poisson) Sei $Y_i \sim \text{Poi}(\mu_i)$ mit $\mu_i = e^{\mathbf{x}_i^T \boldsymbol{\beta}}$, dann gilt

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}.$$

Beispiel 4.10 (Negativ-Binomial mit bekanntem α) Sei $Y_i \sim \text{NB}(\mu_i, \alpha)$ mit $\mu_i = e^{\mathbf{x}_i^T \boldsymbol{\beta}}$ und α bekannt, dann folgt

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i (1 + \alpha \hat{\mu}_i)}.$$

Die beiden Statistiken (Devianz und Pearson χ^2) sind für GLM's approximativ χ_{n-p}^2 -verteilt, was in Dobson (2002) angegeben ist. Die Devianz hat den Vorteil der Additivität für genestete Modelle, wenn Maximum-Likelihood Schätzer verwendet werden, d.h die Devianz eines Modells mit p Parametern ist geringer als die Devianz desselben Modells modelliert mit q Parametern für $p > q$. Man kann also ähnlich zu den ANOVA-Tabellen in LM eine Devianztabelle in GLM erstellen. Dies gilt im allgemeinen nicht für die Pearson χ^2 -Statistik. Die Pearson χ^2 -Statistik ist aber gut interpretierbar, da sie eine ähnliche Form wie die *SQR* in Gleichung (2.4) in LM hat.

Kapitel 5

Regressionsdiagnostiken für GLM's

Nachdem man nun ein Modell gewählt und eine Schätzung der Modellparameter vorgenommen hat, ist man an Vorhersagen interessiert. Um möglichst genaue Vorhersagen zu treffen, müssen wiederum Ausreißer, die die Regressionsresultate entscheidend beeinflussen, identifiziert werden. Es sollen also im folgenden analog zur Entwicklung von Diagnostiken in LM's geeignete Diagnostiken für GLM's entwickelt werden. Das Problem ist hier, dass die in Kapitel 3 vorgestellten Diagnostiken nur für die Normalverteilung gelten. Deshalb benötigt man für GLM's eine erweiterte Definition der Residuen, die auf alle möglichen Verteilungen angewandt werden kann. Darüberhinaus ist es nötig eine verallgemeinerte Form der Hatmatrix herzuleiten, mit der die Residuen standardisiert werden können. Da die Link-Funktion in den GLM's eine große Rolle spielt, ist das Aufstellen der Diagnostiken nicht so einfach wie in den LM's. Ein wichtiger Punkt betrifft den Schätzer $\hat{\beta}$, von dem alle weiteren geschätzten Werte abhängen. Somit wird hier nochmals darauf hingewiesen, dass die geschätzten Werte Z_i , $\hat{\eta}_i$, $\hat{\mu}_i$, w_i , \hat{v}_i und somit auch die Matrizen \mathbf{W} und \mathbf{D}^{-1} von dem Schätzer $\hat{\beta}$ abhängen. Wenn im folgenden nicht explizit auf eine andere Abhängigkeit hingewiesen wird, so sind die geschätzten Werte von $\hat{\beta}$ abhängig. Somit gilt also: $Z_i := Z_i(\hat{\beta})$, $\hat{\eta}_i := \eta_i(\hat{\beta})$, $\hat{\mu}_i := \mu_i(\hat{\beta})$, $\hat{v}_i := v_i(\hat{\beta})$, $w_i := w_i(\hat{\beta})$ und $\mathbf{W} := \mathbf{W}(\hat{\beta})$ und $\mathbf{D}^{-1} := \mathbf{D}^{-1}(\hat{\beta})$.

Außerdem werden in den nächsten Kapiteln neben den Beispielen für die Diagnostiken konkrete Beispiele anhand eines Patentdatensatzes für den Poissonfall gegeben. Zu diesem Zweck soll dieser Datensatz hier kurz vorgestellt werden.

5.1 Der Patentdatensatz

5.1.1 Beschreibung des Datensatzes

Die untersuchten Daten sind auf der Internetseite

<http://coe.ubc.ca/users/marty/patent.data>

erhältlich und beschreiben Patent-Daten von amerikanischen High-Tech-Firmen aus dem Jahr 1976. Insgesamt umfasst der Datensatz 70 Firmen, d.h. die Länge der Daten ist auf 70 beschränkt. Die erste Spalte in gegebenen Datensatz gibt die Namen der Firmen an. Die zweite Spalte *patente* enthält die jeweilige Anzahl an Patenten. Diese Spalte soll als Response verwendet werden. Desweiteren gibt die dritte Spalte *rdsales* das Verhältnis der Geldbeträge an, die für Forschung und Entwicklung (Research and Development) investiert wurden im Vergleich zu den Einnahmen, die sich durch die Entwicklung der Patente ergaben. In der vierten Spalte, hier als *log(rd)* angegeben, stehen die mit dem Logarithmus transformierten Geldbeträge für Forschung

und Entwicklung. Da hier im folgenden aber auch die reinen Investitionsgelder für Forschung und Entwicklung benötigt werden, wird die durch die Exponentialfunktion zurücktransformierte Spalte rd betrachtet. Diese Spalte ist in der Einheit 1 Mio. Dollar angegeben. Der Datensatz enthält noch eine fünfte Spalte $(\log(rd))^2$, auf die aber im weiteren Verlauf nicht eingegangen wird. Es liegen also drei brauchbare Variablen vor, um die Daten zu modellieren. Diese sind der Response $patente$ und die beiden Kovariablen $rdsales$ und $\log(rd)$. Die Daten sind nicht gruppiert, sondern basieren auf Einzeldaten, d.h. jede Zeile des Datensatzes kann genau einer Firma zugeordnet werden.

5.1.2 Explorative Datenanalyse

Zuerst soll ein Modell angepasst werden, das die Zählvariable *Anzahl der Patente* erklärt. Es wird zunächst die Poissonverteilung mit der Log-Link-Funktion zur Modellierung der Daten herangezogen. Also erhält man:

$$patente \sim Poi(\mu_i), \quad i = 1, \dots, 70,$$

$$\log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} \quad \text{bzw.} \quad \mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta}).$$

Wenn man nun die jeweilige Kovariable gegen den Logarithmus der Patente anträgt, so müsste ein linearer Zusammenhang zu sehen sein. In der Abbildung 5.1 soll dies nun veranschaulicht werden.

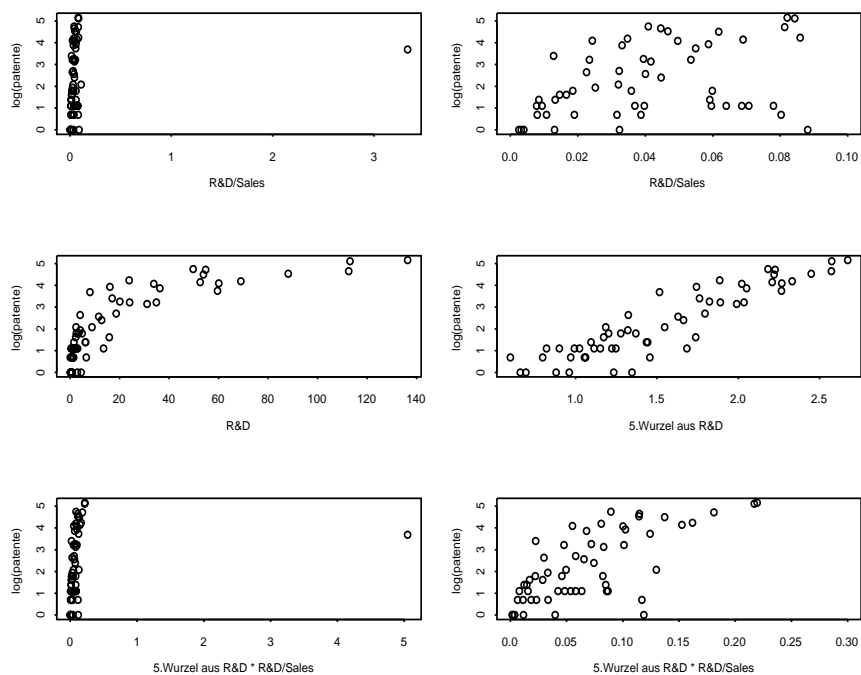


Abbildung 5.1: Explorative Datenanalyse des Patentdatensatzes

Links oben sieht man den vollständigen Datensatz. Da die Beobachtung 6 hier besonders von den anderen Daten abweicht, wird die x-Achse skaliert, so dass in der Grafik rechts oben die

vorherige Punktwolke überschaubarer wird. Die Beobachtung 6 ist nun nicht mehr zu sehen. Nun ist ein leichter linearer Zusammenhang sichtbar. Es stören die Punkte rechts unten, also mit hohem Wert *rdsales* und dazugehörigem niedrigem Wert *log(patente)*. Man lässt die Variable aber untransformiert, da eine geeignete Transformation nicht ersichtlich ist. Die Variable *rd* in der linken mittleren Grafik hingegen zeigt die Notwendigkeit einer Transformation. Es gibt nun die Möglichkeit, die Logarithmusfunktion oder die Wurzelfunktion anzuwenden. Da die Logarithmusfunktion keine adäquate Linearisierung bewirkt, wird im weiteren Verlauf mit der Wurzelfunktion gearbeitet. Wie man in der rechten mittleren Grafik sieht, ist die 5. Wurzel der Variable *rd* eine annehmbare Transformation. Die Variable wird mit *rdw5* bezeichnet. Die letzten beiden Grafiken zeigen die Interaktion zwischen den Variablen *rdsales* und *rdw5*. Die linke Grafik umfasst den vollständigen Datensatz und die rechte ist in der x-Achse skaliert, so dass die Beobachtung 6 wegen des hohen x-Wertes nicht sichtbar ist. Man erkennt eine schwache Linearität und lässt somit auch diese Interaktion in das Modell einfließen, das nun erstellt werden soll.

5.1.3 Poissonregression

Mit den entsprechenden Haupteffekten und Interaktionen kann man nun ein Poisson-Modell in SPlus eingeben, dass wie folgt lautet:

```
patentpoi_glm(patente~rdsales+rdw5+rdsales*rdw5,family=poisson,link=log,x=T)
```

Mit Hilfe der *summary*-Funktion bekommt man Informationen über die geschätzten Regressionskoeffizienten $\hat{\beta}_j$, $j = 1, \dots, 4$, die in der Spalte *value* angegeben sind.

```
Call: glm(formula = patente ~ rdsales + rdw5 + rdsales * rdw5,
family = poisson, x = T, link = log)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-4.81634	-1.719084	-0.8486778	0.1408976	7.279557

Coefficients:

	Value	Std. Error	t value
(Intercept)	-1.165792	0.1850514	-6.299826
rdsales	-5.500628	1.9909157	-2.762863
rdw5	2.299744	0.1035849	22.201529
rdsales:rdw5	3.903443	1.3119114	2.975386

(Dispersion Parameter for Poisson family taken to be 1)

Null Deviance: 3324.765 on 69 degrees of freedom

Residual Deviance: 377.1818 on 66 degrees of freedom

Number of Fisher Scoring Iterations: 4

Tabelle 5.1: *summary* des Poissonmodells für den Patentdatensatz

Aus der Tabelle 5.1 kann man erkennen, dass alle Kovariablen hoch signifikant sind, da der *t value* für jede Kovariable betragsmäßig größer als 2 ist, d.h. die Hypothesen $H : \beta_j = 0, j = 1, \dots, 4$ können zu einem kleinen Signifikanzniveau verworfen werden. Jedoch zeigt der Wert der Devianz mit 377.18 einen deutlich höheren Wert als die Anzahl der Freiheitsgrade mit 66, was auf eine eher schlechte Anpassung der Daten durch das Poissonmodell schließen lässt.

In den weiteren Kapiteln wird gleichzeitig zu den theoretischen Grundlagen der Diagnostiken immer wieder auf das konkrete Beispiel der Poissonregression anhand des Patentdatensatzes eingegangen, das die Theorie veranschaulichen und die praktische Anwendung zeigen soll.

5.2 Residuen und Hatmatrix

Ebenso wie bei den LM's sieht man sich die Residuen an. In GLM's werden aber nun weitere verallgemeinerte Residuen definiert. Zuerst kann man sich wieder die gewöhnlichen Residuen ansehen, die nachfolgend definiert werden:

$$r_i^r := y_i - \hat{\mu}_i, \quad i = 1, \dots, n. \quad (5.1)$$

Für das klassische Lineare Modell waren die Fehler $y_i - \mu_i$ normalverteilt, homoskedastisch und symmetrisch um Null mit konstanter Varianz verteilt. Für Zähldaten hingegen ist der Fehler $y_i - \mu_i$ heteroskedastisch und asymmetrisch. Deshalb werden hier die *Pearson-Residuen* angegeben, die wie folgt definiert sind:

$$r_i^P := \frac{y_i - \hat{\mu}_i}{\hat{v}_i}, \quad i = 1, \dots, n, \quad \hat{v}_i \text{ geschätzte Varianzfunktion.} \quad (5.2)$$

Um eine Standardisierung vorzunehmen, ist wie bei den LM's die Hatmatrix nötig. Hier wird nun eine verallgemeinerte Form der Hatmatrix hergeleitet, die auf der Maximum-Likelihood-Schätzung basiert. Wie schon in Kapitel 3 erwähnt, gibt sie Auskunft über high leverage Punkte. Der Schätzer $\hat{\beta}$ für β wurde bereits mittels IWLS in Kapitel 4.2 ausgerechnet. Nachdem der IWLS-Algorithmus konvergiert ist, gilt:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{Z},$$

wobei

$$\begin{aligned} \mathbf{W} &:= \mathbf{W}(\hat{\beta}) = \text{diag} \left(\left(\frac{d\mu_1}{d\eta_1} \right)^2 \Big/ v_1 \Big|_{\beta=\hat{\beta}}, \dots, \left(\frac{d\mu_n}{d\eta_n} \right) \Big/ v_n \Big|_{\beta=\hat{\beta}} \right) \\ \mathbf{Z} &:= \mathbf{Z}(\hat{\beta}) = \hat{\eta} + \mathbf{D}^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}}) = \mathbf{X}\hat{\beta} + \mathbf{D}^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}}) \\ \mathbf{D}^{-1} &:= \mathbf{D}^{-1}(\hat{\beta}) = \text{diag} \left(\frac{d\eta_1}{d\mu_1} \Big|_{\beta=\hat{\beta}}, \dots, \frac{d\eta_n}{d\mu_n} \Big|_{\beta=\hat{\beta}} \right). \end{aligned}$$

Nach der GLM-Theorie gilt, dass der Schätzer $\hat{\beta}$ ein gewichteter Kleinste-Quadrate Schätzer mit Zielvariable $Z_i = \mathbf{x}_i^T \beta + (y_i - \mu_i) \frac{d\eta_i}{d\mu_i}$ ist. Der Schätzer $\hat{\beta}$ ist zudem ein ungewichteter Kleinste-Quadrate Schätzer des linearen Problems $\mathbf{W}^{1/2} \mathbf{Z} = \mathbf{W}^{1/2} \mathbf{X} \beta + \mathbf{W}^{1/2} \boldsymbol{\epsilon}$. Die Hatmatrix für dieses LM und damit für das GLM hat folgende Form:

$$H = \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2}. \quad (5.3)$$

Es gilt wie schon in den LM's $H = H^T$ und $H^2 = H$, d.h. die Hatmatrix H in den GLM's ist idempotent und symmetrisch. Dies lässt sich leicht zeigen, indem man folgendes ausführt:

$$\begin{aligned} H^T &= (\mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2})^T = \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2} = H, \\ H^2 &= (\mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2}) (\mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2})^T \\ &= \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2} \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2} \\ &= \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2} = H. \end{aligned}$$

Außerdem gilt $\sum_{i=1}^n h_{ii} = p$, was mit Hilfe der in Kapitel 3 aufgeführten Spur einer Matrix bewiesen werden kann. Somit gilt:

$$\begin{aligned} \text{Spur}(H) &= \text{Spur}(\mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2}) \\ &= \text{Spur}((\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2} \mathbf{W}^{1/2} \mathbf{X}) = \text{Spur}(I_p) = p. \end{aligned}$$

Falls $h_{ii} > 2p/n$, wird die i -te Beobachtung als high leverage Punkt (x-Ausreißer) angesehen. Im Gegensatz zum klassischen linearen Modell ist hier die Hatmatrix nicht nur von der Design-Matrix abhängig, sondern auch vom geschätzten Wert $\hat{\beta}$, der durch die Matrix \mathbf{W} eingeht. Deswegen müssen extreme Beobachtungen in der Design-Matrix nicht notwendigerweise einen hohen Wert h_{ii} aufweisen.

Nun lassen sich die standardisierten Residuen angeben. Um eine Standardisierung vorzunehmen, werden die Varianzen von y_i und $\hat{\mu}_i$ benötigt, welche wie folgt gegeben sind:

$$\begin{aligned} \text{Var}(y_i) &= v_i, \\ \text{Var}(\hat{\mu}_i) &\approx v_i h_{ii}. \end{aligned}$$

Da die Varianz von $\hat{\mu}_i$ nicht leicht ersichtlich ist, wird eine Taylorentwicklung von μ_i um $\hat{\mu}_i$ angewandt. Da

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

gilt nach Taylor erster Ordnung

$$\begin{aligned} \mu_i &\approx \hat{\mu}_i + h'(\mathbf{x}_i^T \boldsymbol{\beta}) \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} (\mu_i - \hat{\mu}_i), \\ h(\eta_i) &\approx h(\hat{\eta}_i) + h'(\eta_i) \Big|_{\eta=\hat{\eta}_i} (\eta_i - \hat{\eta}_i). \end{aligned}$$

Damit ist

$$\mu_i \approx \hat{\mu}_i + h'(\eta_i) \Big|_{\eta=\hat{\eta}_i} (\eta_i - \hat{\eta}_i)$$

und es gilt mit der Annahme, dass $h'(\eta_i) \Big|_{\eta=\hat{\eta}_i} =: h'(\hat{\eta}_i)$ konstant ist

$$\begin{aligned} \text{Var}(\mu_i - \hat{\mu}_i) = \text{Var}(\hat{\mu}_i) &\approx \text{Var}(h'(\hat{\eta}_i)(\eta_i - \hat{\eta}_i)) \\ &= h'(\hat{\eta}_i)^2 \text{Var}(\eta_i - \hat{\eta}_i) \\ &= h'(\hat{\eta}_i)^2 \text{Var}(\hat{\eta}_i). \end{aligned} \quad (5.4)$$

Ferner gilt mit $\frac{d\mu_i}{d\eta_i} = h'(\eta_i)$ und $w_i = \left(\frac{d\mu_i}{d\eta_i}\right)^2 / v_i$

$$h'(\hat{\eta}_i)^2 = \left(\frac{d\mu_i}{d\eta_i}\right)^2 = w_i v_i$$

und mit $\hat{\beta} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{Z}$ und $Z_i = \mathbf{x}_i^T \hat{\beta} + (y_i - \hat{\mu}_i) \frac{d\eta_i}{d\mu_i} \Big|_{\mu_i = \hat{\mu}_i}$ mit $\mathbf{W}(\hat{\beta})$ als konstant betrachtet

$$\begin{aligned} \text{Var}(\hat{\eta}_i) &= \text{Var}(\mathbf{x}_i^T \hat{\beta}) = \mathbf{x}_i^T \text{Var}(\hat{\beta}) \mathbf{x}_i \\ &= \mathbf{x}_i^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \underbrace{\text{Var}(\mathbf{Z})}_{= \text{diag}\left(v_i \left(\frac{d\eta_i}{d\mu_i}\right)^2\right)} \mathbf{W} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i \\ &= \mathbf{x}_i^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i \end{aligned} \quad (5.5)$$

Setzt man die letzten beiden Ausdrücke in die Gleichung (5.4) ein, so ergibt sich

$$\begin{aligned} \text{Var}(\hat{\mu}_i) &\approx h'(\hat{\eta}_i)^2 \text{Var}(\hat{\eta}_i) \\ &\approx v_i w_i \mathbf{x}_i^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i \\ &= v_i w_i^{1/2} \mathbf{x}_i^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i w_i^{1/2} \\ &= v_i h_{ii} \end{aligned} \quad (5.6)$$

Damit ist die Approximation gezeigt und es ergibt sich für das standardisierte Pearson-Residuum:

$$r_i^{P*} := \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{v}_i(1 - h_{ii})}}, \quad i = 1, \dots, n. \quad (5.7)$$

Es gilt folgende hilfreiche Beziehung zwischen der sogenannten *Pearson-Statistik* \mathcal{X}^2 und dem Pearson-Residuum:

$$\sum_{i=1}^n (r_i^P)^2 =: \mathcal{X}^2.$$

Diese Beziehung mit der Pearson-Statistik war auch für die Namensgebung verantwortlich. Ein Nachteil der Pearson-Residuen ist, dass die Verteilung für $\mathbf{r}^P = (r_1^P, \dots, r_n^P)^T$ für nicht normalverteilte Verteilungen oft deutlich schief ist, was die Eigenschaften ähnlich zur Normalverteilungstheorie verhindert. In McCullagh und Nelder (1989) wird darauf hingewiesen, dass Anscombe vorschlug, deshalb ein Residuum mit Hilfe einer Funktion $A(y)$ statt y zu definieren, die *Anscombe-Residuen*. Diese Residuen werden im weiteren Verlauf aber nicht verwendet und werden deshalb hier nicht näher besprochen. Trotz der Unähnlichkeit zur Normalverteilungstheorie setzt man eine absolute Schranke von 2 für das Pearson-Residuum fest, da es durch den Erwartungswert und die Varianzfunktion eine Standardisierung erfährt und damit das 95%-Niveau der Standardnormalverteilung eine akzeptable Schranke ist. Desweiteren werden die *Devianz-Residuen* definiert:

$$r_i^D := \text{sign}(y_i - \hat{\mu}_i) \sqrt{d_i}, \quad \text{sign}(x) = \begin{cases} -1 & \text{für } x < 0 \\ 0 & \text{für } x = 0 \\ 1 & \text{für } x > 0 \end{cases} \quad (5.8)$$

wobei $d_i = 2(\mathcal{L}(y_i; y_i, \phi) - \mathcal{L}(y_i; \hat{\mu}_i, \phi))$ und $D := \sum d_i$.

Ebenso wie die Pearson-Residuen werden die Devianz-Residuen standardisiert, was in Cameron und Trivedi (1998) nachgelesen werden kann, und folgendes ergibt sich:

$$r_i^{D*} = \frac{d_i}{\sqrt{1 - h_{ii}}}. \quad (5.9)$$

Wenn die Devianz D als Diskrepanzmaß eines GLM's benutzt wird, trägt jede Einheit ein d_i zum Gesamtmaß bei und somit gilt $\sum d_i = D$. Wie man leicht sieht, gilt hier:

$$\sum_{i=1}^n (r_i^D)^2 = D.$$

Auch für das Devianz-Residuum gilt eine absolute Schranke von 2.

Nun werden die Residuen und die h_{ii} -Werte konkret für den Patentdatensatz berechnet. In der Abbildung 5.2 werden die Devianz- und Pearson-Residuen grafisch dargestellt.

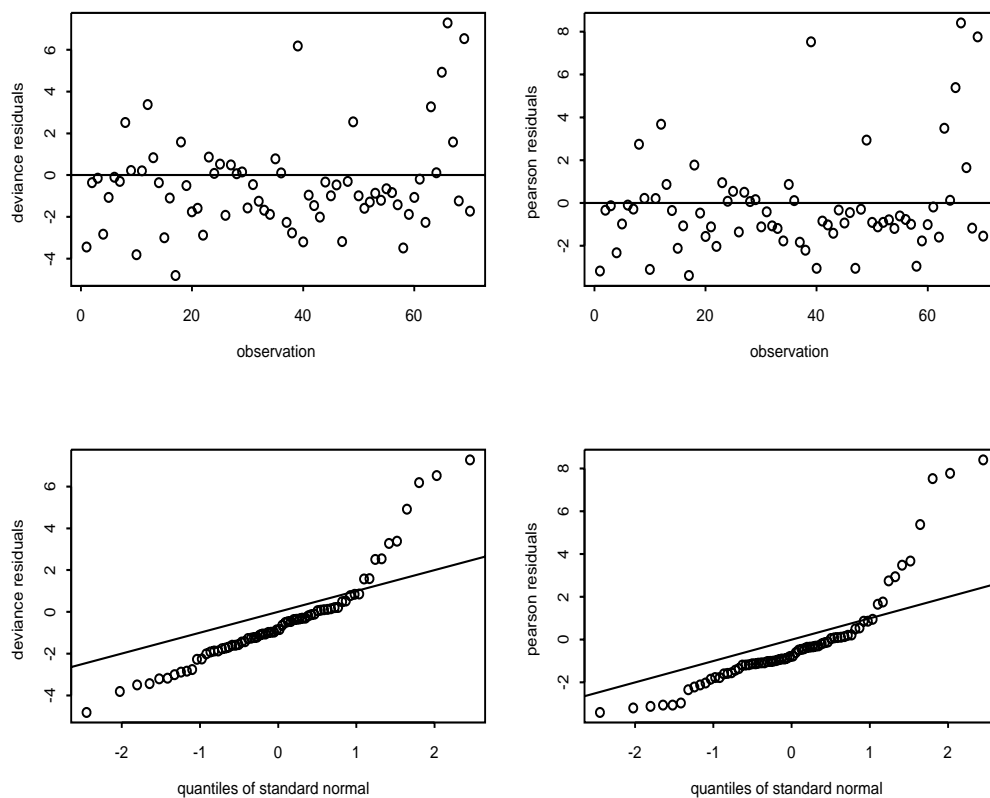


Abbildung 5.2: Devianz- und Pearson-Residuen des Poissonmodells für den Patentdatensatz (oben), Plot der Standardnormalverteilungsquantile gegen Devianz- und Pearson-Residuen des Poissonmodells für den Patentdatensatz (unten). Die durchgezogene Linie stellt die Winkelhalbierende dar.

In den oberen beiden Grafiken lässt sich erkennen, dass die Residuen zufällig um Null streuen, jedoch teilweise sehr hohe Werte annehmen. Die höchsten Werte der Devianz-Residuen liegen über

dem Wert 6 und die der Pearson-Residuen sogar bei einem Wert von 8. Die Beobachtungen, die sich um diese Werte bewegen, sind Beobachtung 39 (INTL FLAVORS & FRAGRANCES), 66 (SQUIBB CORP) und 69 (SYNTEX CORP). Die unteren beiden Grafiken zeigen die Quantile der Standardnormalverteilung gegenüber den beiden Residuen. Man sieht, dass sich die Residuen nicht nach der Normalverteilung verhalten, da sie deutlich von der durchgezogenen Linie, der Winkelhalbierenden abweichen.

Um die Diagonalelemente der Hatmatrix zu berechnen, wurde die SPlus-Funktion *diag.poi* geschrieben, die die entsprechenden Werte für die h_{ii} -Werte ausgibt und im Anhang definiert ist. Darüberhinaus werden die Werte der $DFFITs_i^1$ und der Cook's Distanz D_i^1 ausgegeben, was aus der Definition der Funktion hervorgeht und in den weiteren Kapiteln veranschaulicht wird. Wahlweise kann man sich die Werte auch in einer Grafik als Stabdiagramm anzeigen lassen. In Tabelle 5.2 und Abbildung 5.3 sieht man die Werte bzw. das Stabdiagramm der h_{ii} -Werte für den Patentdatensatz.

\$hat :

```
[1] 0.05325618 0.04116812 0.02463463 0.02237729 0.01810229 0.99797715
[7] 0.12956209 0.03431636 0.02095536 0.03058519 0.02150468 0.04455878
[13] 0.03764912 0.02049828 0.02123762 0.12121163 0.02734118 0.02067569
[19] 0.02495286 0.02394835 0.05494450 0.01896500 0.03695230 0.01967470
[25] 0.02926326 0.03389782 0.02092847 0.03307461 0.02448449 0.05071626
[31] 0.03909024 0.02687873 0.03994471 0.05566642 0.04934053 0.02251017
[37] 0.01965655 0.01872574 0.03721016 0.31200964 0.02886077 0.04946351
[43] 0.03126171 0.32673822 0.02644582 0.02002623 0.50704707 0.04220160
[49] 0.02119941 0.01912205 0.04341935 0.35933087 0.02356208 0.13871876
[55] 0.01961825 0.02147327 0.05184777 0.03307656 0.04366077 0.03166684
[61] 0.02993423 0.02805872 0.05133809 0.05850115 0.08860905 0.05760382
[67] 0.05018390 0.03605021 0.05847773 0.02205502
```

Tabelle 5.2: h_{ii} -Werte des Poissonmodells für den Patentdatensatzes

In der Grafik sind die Beobachtungen, die jeweils die kritische Schranke von $2p/n = 8/70 = 0.11$ überschreiten, angegeben. Insbesondere fällt Beobachtung 6 (ALZA CORP-CL A) auf. Der Wert für das Element $h_{66} = 0.9980$ ist außergewöhnlich hoch, was auf einen x-Ausreißer schließen lässt. Wenn man sich den Datensatz ansieht, bemerkt man, dass der Wert der Kovariable *rdsales* mit einem Wert von 3.33 ungewöhnlich hoch ist. Das bedeutet, dass die Investitionen für Forschung und Entwicklung gegenüber den Einnahmen der Patente sehr groß waren. Als eine weitere Erklärung wäre eine fehlerhafte Datenangabe denkbar. Desweiteren zeigen die Beobachtungen 40 (JOHNSON & JOHNSON), 44 (LILLY (ELI) & CO), 47 (MERCK & CO) und 52 (NEWPORT PHARMACEUTICALS INTL) einen Wert, der deutlich über der kritischen Schranke liegt. Diese Beobachtungen sind ebenso x-Ausreißer.

Um die auffallendsten Beobachtungen und ihre jeweiligen Werte nochmals explizit darzustellen, wurde eine weitere SPlus-Funktion für die h_{ii} -Werte geschrieben, welche die Beobachtungen ausgibt, die die signifikante Schranke überschreiten. Im Anhang wird die Funktion *einflbeob.hat* deklariert. In Tabelle 5.3 werden die auffälligsten Beobachtungen und dazugehörigen Werte für die h_{ii} -Werte aufgelistet.

Wie schon aus der Abbildung 5.3 erkennbar überschreiten auch die Beobachtungen 7 (AMERICAN HOME PRODUCTS CORP), 16 (BRISTOL-MYERS CO) und 54 (PFIZER INC) den Wert 0.11. Nach der Definition eines x-Ausreißers, d.h. von $h_{ii} \geq 2p/n$, sind auch diese Beob-

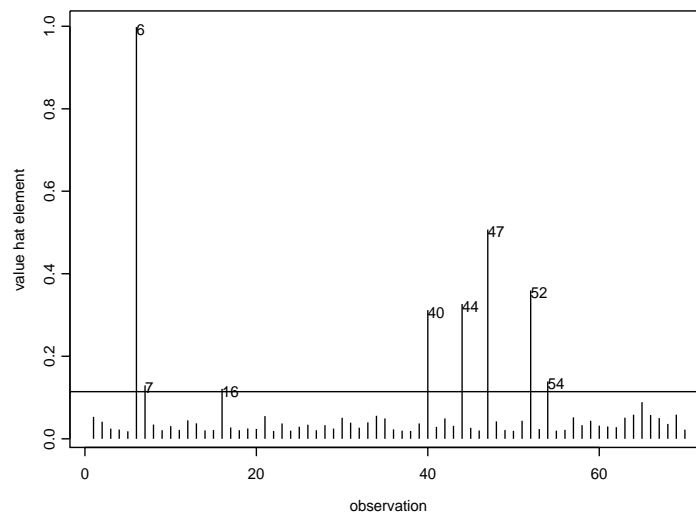


Abbildung 5.3: Stabdiagramm der h_{ii} -Werte des Poissonmodells für den Patentdatensatz. Die durchgezogene Linie markiert die Schranke $2p/n = 8/70 = 0.11$.

h_{ii} -Werte

\$eb:

```
[1] 0.9979771 0.5070471 0.3593309 0.3267382 0.3120096 0.1387188 0.1295621
[8] 0.1212116
```

\$b:

```
[1] 6 47 52 44 40 54 7 16
```

Tabelle 5.3: h_{ii} -Werte mit $h_{ii} \geq 2p/n$ des Poissonmodells für den Patentdatensatz

achtungen x-Ausreißer. Da in diesem Datensatz $p = 4$, also p klein, werden eventuell zu viele Beobachtungen angezeigt. Deshalb sind letztere drei Beobachtungen eventuell zu vernachlässigen.

5.3 Ein-Schritt-Approximation von $\hat{\beta}_{-i}$ bei GLM's

Einflußmaße der i -ten Beobachtung auf den ML-Schätzer $\hat{\beta}$ können auf der stichprobenbasierten Influenzkurve $SIK_i \propto \hat{\beta} - \hat{\beta}_{-i}$ basieren. Es ist rechnerisch aber sehr aufwendig, da man $n + 1$ ML-Schätzungen benötigt. Deswegen ist es nützlich, eine quadratische Approximation des Log-Likelihood ohne i -te Beobachtung $\mathcal{L}_{-i}(\beta)$ um $\hat{\beta}$ zu machen. Diese Approximation basiert auf Taylor und lautet

$$\mathcal{L}_{-i}(\beta) \approx \mathcal{L}_{-i}(\hat{\beta}) + (\beta - \hat{\beta})^T S_{-i}(\hat{\beta}) - \frac{1}{2}(\beta - \hat{\beta})^T FI_{-i}(\hat{\beta})(\beta - \hat{\beta}), \quad (5.10)$$

wobei $S_{-i}(\hat{\beta}) = \left. \frac{\partial \mathcal{L}_{-i}(\beta)}{\partial \beta} \right|_{\beta=\hat{\beta}}$ die Scoregleichungen ohne i -te Beobachtung und $FI_{-i}(\hat{\beta}) =$

$-E \left[\frac{\partial^2 \mathcal{L}_{-i}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_{-i}} \right]$ die Fisher-Informationsmatrix ohne i -te Beobachtung sind. Es muss folgendes gelten:

$$\frac{\partial}{\partial \boldsymbol{\beta}} \mathcal{L}_{-i}(\boldsymbol{\beta}) \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_{-i}} = \mathbf{0}.$$

Also gilt für die Approximation in (5.10) abgeleitet nach $\boldsymbol{\beta}$ und ausgewertet an der Stelle $\hat{\boldsymbol{\beta}}_{-i}$

$$S_{-i}(\hat{\boldsymbol{\beta}}) - FI_{-i}(\hat{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}}_{-i} - \hat{\boldsymbol{\beta}}) \approx \mathbf{0}.$$

Falls die Fisher-Informationsmatrix $FI_{-i}(\hat{\boldsymbol{\beta}})$ positiv definit ist, ist die quadratische Approximation durch

$$\hat{\boldsymbol{\beta}}_{-i}^1 := \hat{\boldsymbol{\beta}} + (FI_{-i}(\hat{\boldsymbol{\beta}}))^{-1} S_{-i}(\hat{\boldsymbol{\beta}}) \quad (5.11)$$

gegeben. Der Wert $\hat{\boldsymbol{\beta}}_{-i}^1$ ist die sogenannte *Ein-Schritt-Approximation* für $\hat{\boldsymbol{\beta}}_{-i}$. Dieser Wert ist derselbe, den man durch die Newton-Methode mit Startwert $\hat{\boldsymbol{\beta}}$ nach einem Schritt erhält, um die Funktion $\mathcal{L}_{-i}(\boldsymbol{\beta})$ zu maximieren. Falls $\hat{\boldsymbol{\beta}}_{-i}^1$ nicht zu verschieden von $\hat{\boldsymbol{\beta}}$ und $\mathcal{L}_{-i}(\boldsymbol{\beta})$ lokal quadratisch ist, ist die Ein-Schritt-Approximation in der Nähe des exakt iterierten Werts. Für einflußreiche Beobachtungen ist $\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{-i}$ 'groß'. Die Genauigkeit der Ein-Schritt-Schätzung ist geringer, aber eine genaue Approximation für $\hat{\boldsymbol{\beta}}_{-i}$ ist nicht nötig, wenn $\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{-i}^1$ genügend groß ist.

Wenn nun das Newton-Raphson-Verfahren auf das GLM angewendet wird, so ergibt sich für die Scoregleichungen und für die Fisher-Informationsmatrix analog zu den Gleichungen (4.6) und (D.3)

$$\begin{aligned} S_{-i}(\hat{\boldsymbol{\beta}}) &= \mathbf{X}_{-i}^T \mathbf{W}_{-i} \mathbf{D}_{-i}^{-1} \mathbf{r}_{-i}^r, \\ FI_{-i}(\hat{\boldsymbol{\beta}}) &= \mathbf{X}_{-i}^T \mathbf{W}_{-i} \mathbf{X}_{-i}, \end{aligned}$$

wobei \mathbf{X}_{-i} die Design-Matrix mit gelöschter i -ten Zeile, \mathbf{r}_{-i}^r der Vektor mit entfernter i -ten Komponente, \mathbf{W}_{-i} und \mathbf{D}_{-i}^{-1} die Matrizen mit i -ter gelöschter Zeile sind, also

$$\begin{aligned} \mathbf{W}_{-i} &:= \text{diag}(w_1, \dots, w_{i-1}, w_{i+1}, \dots, w_n), \\ \mathbf{D}_{-i}^{-1} &:= \text{diag} \left(\frac{d\eta_1}{d\mu_1}, \dots, \frac{d\eta_{i-1}}{d\mu_{i-1}}, \frac{d\eta_{i+1}}{d\mu_{i+1}}, \dots, \frac{d\eta_n}{d\mu_n} \right). \end{aligned}$$

Nach Anwendung von (5.11) erhält man damit

$$\hat{\boldsymbol{\beta}}_{-i}^1 = \hat{\boldsymbol{\beta}} + (\mathbf{X}_{-i}^T \mathbf{W}_{-i} \mathbf{X}_{-i})^{-1} \mathbf{X}_{-i}^T \mathbf{W}_{-i} \mathbf{D}_{-i}^{-1} \mathbf{r}_{-i}^r.$$

Da in Kapitel 4.2 innerhalb des IWLS $w_i := v_i^{-1} \left(\frac{d\mu_i}{d\eta_i} \right)^2$ definiert wurde, gilt $\mathbf{W} = \mathbf{V}^{-1} \mathbf{D}^2$ mit $\mathbf{W} := \text{diag}(w_1, \dots, w_n)$, $\mathbf{V} := \text{diag}(v_1, \dots, v_n)$ und $\mathbf{D} := \text{diag} \left(\frac{d\mu_1}{d\eta_1}, \dots, \frac{d\mu_n}{d\eta_n} \right)$. Damit gilt nach Entfernung der i -ten Zeile auch $\mathbf{W}_{-i} = \mathbf{V}_{-i}^{-1} \mathbf{D}_{-i}^2$ und aufgrund der Tatsache, dass diese Matrizen Diagonalmatrizen darstellen $\mathbf{W}_{-i}^{1/2} = \mathbf{V}_{-i}^{-1/2} \mathbf{D}_{-i}$. Also lässt sich damit folgendes schreiben:

$$\begin{aligned}\hat{\beta}_{-i}^1 &= \hat{\beta} + (\mathbf{X}_{-i}^T \mathbf{W}_{-i} \mathbf{X}_{-i})^{-1} \mathbf{X}_{-i}^T \mathbf{W}_{-i}^{1/2} \mathbf{V}_{-i}^{-1/2} \mathbf{r}_{-i} \\ &= \hat{\beta} + (\mathbf{X}_{-i}^T \mathbf{W}_{-i} \mathbf{X}_{-i})^{-1} \mathbf{X}_{-i}^T \mathbf{W}_{-i}^{1/2} \mathbf{r}_{-i}^P,\end{aligned}\quad (5.12)$$

wobei \mathbf{V}_{-i} die Diagonalmatrix der Varianzen von Y_j , $j = 1, \dots, n$, $j \neq i$, mit i -ter Zeile entfernt und $\mathbf{r}_{-i}^P = (r_1^P, \dots, r_{i-1}^P, r_{i+1}^P, \dots, r_n^P)$ ist. Mit Hilfe von Gleichung (3.15) und (3.16) aus Kapitel 3 kann man $(\mathbf{X}_{-i}^T \mathbf{W}_{-i} \mathbf{X}_{-i})^{-1}$ und $\mathbf{X}_{-i}^T \mathbf{W}_{-i}^{1/2} \mathbf{r}_{-i}^P$ als

$$\begin{aligned}(\mathbf{X}_{-i}^T \mathbf{W}_{-i} \mathbf{X}_{-i})^{-1} &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} + \frac{(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i w_i^{1/2} w_i^{1/2} \mathbf{x}_i^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}}{1 - \underbrace{w_i^{1/2} \mathbf{x}_i^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i w_i^{1/2}}_{=h_{ii}}}, \\ \mathbf{X}_{-i}^T \mathbf{W}_{-i}^{1/2} \mathbf{r}_{-i}^P &= \begin{pmatrix} \sum_{j=1, j \neq i}^n w_j^{1/2} r_j^P \\ \sum_{j=1, j \neq i}^n x_{j1} w_j^{1/2} r_j^P \\ \vdots \\ \sum_{j=1, j \neq i}^n x_{jk} w_j^{1/2} r_j^P \end{pmatrix} \\ &= \begin{pmatrix} \sum_{j=1}^n w_j^{1/2} r_j^P \\ \sum_{j=1}^n x_{j1} w_j^{1/2} r_j^P \\ \vdots \\ \sum_{j=1}^n x_{jk} w_j^{1/2} r_j^P \end{pmatrix} - \begin{pmatrix} w_i^{1/2} r_i^P \\ x_{i1} w_i^{1/2} r_i^P \\ \vdots \\ x_{ik} w_i^{1/2} r_i^P \end{pmatrix} = \mathbf{X}^T \mathbf{W}^{1/2} \mathbf{r}^P - w_i^{1/2} \mathbf{x}_i r_i^P\end{aligned}\quad (5.13)$$

schreiben. Letztere Gleichung lässt sich vereinfachen, da die Scoregleichungen für den Schätzer $\hat{\beta}$ darin vorkommen. Es gilt nämlich nach (4.6)

$$\mathbf{X}^T \mathbf{W}^{1/2} \mathbf{r}^P = \mathbf{X}^T \mathbf{W} \mathbf{D}^{-1} \mathbf{r} = \mathbf{0},$$

und somit

$$\mathbf{X}_{-i}^T \mathbf{W}_{-i}^{1/2} \mathbf{r}_{-i}^P = -w_i^{1/2} \mathbf{x}_i r_i^P. \quad (5.14)$$

Einsetzen der Gleichungen (5.13) und (5.14) in Gleichung (5.12) ergibt

$$\begin{aligned}\hat{\beta}_{-i}^1 &= \hat{\beta} + \left((\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} + \frac{(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i w_i^{1/2} w_i^{1/2} \mathbf{x}_i^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}}{1 - h_{ii}} \right) \left(-w_i^{1/2} \mathbf{x}_i r_i^P \right) \\ &= \hat{\beta} + \frac{-(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} w_i^{1/2} \mathbf{x}_i r_i^P + (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} w_i^{1/2} \mathbf{x}_i r_i^P h_{ii}}{1 - h_{ii}} + \\ &\quad + \frac{-(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i w_i^{1/2} \overbrace{w_i^{1/2} \mathbf{x}_i^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} w_i^{1/2} \mathbf{x}_i r_i^P}_{=h_{ii}}}{1 - h_{ii}} \\ &= \hat{\beta} - w_i^{1/2} \frac{(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i r_i^P}{(1 - h_{ii})},\end{aligned}\quad (5.15)$$

wobei \mathbf{W} , h_{ii} und r_i^P von $\hat{\beta}$ abhängen. Für weitere Berechnungen und das Auswerten der im folgenden definierten Diagnostiken wird diese Ein-Schritt-Approximation für $\hat{\beta}_{-i}$ verwendet, so dass für den Schätzer $\hat{\beta}_{-i}$ jeweils die Ein-Schritt-Approximation $\hat{\beta}_{-i}^1$ eingesetzt wird. Sollte dies einmal nicht der Fall sein, wird explizit darauf hingewiesen.

In folgender Anwendung wird der Patentdatensatz mit Hilfe des Poissonmodells beschrieben und verwendet. Es sollen nun die Ein-Schritt-Approximationen und die exakten Werte für jede Beobachtung berechnet werden und danach verglichen werden. Es werden also $n+1 = 71$ Regressionen durchgeführt, d.h. Regression des Poissonmodells basierend auf allen Daten und jeweils mit Weglassen der i -ten Beobachtung. Weiter werden die 70 Ein-Schritt-Approximationen nach Formel (5.15) berechnet und dann mit den wahren Werten verglichen. In Abbildung 5.4 sieht man die Differenz von Ein-Schritt-Approximation und exaktem Wert jeweils für jede Kovariable.

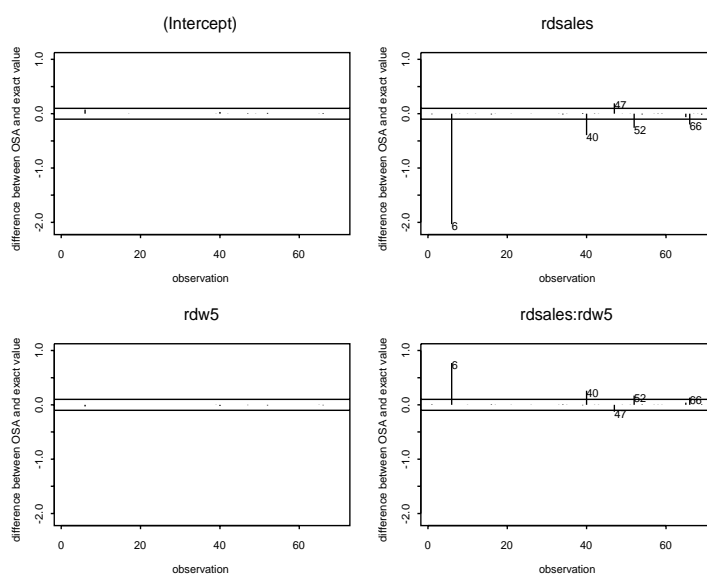


Abbildung 5.4: Stabdiagramm der Differenz von Ein-Schritt-Approximation $\hat{\beta}_{-i}^1$ und exaktem Wert $\hat{\beta}_{-i}$ des Poissonmodells für den Patentdatensatz

Die durchgezogenen Linien sind betragsmäßig gleich 0.1. Für den Intercept und die Kovariable *rdw5* ist die Differenz für alle Beobachtungen betragsmäßig unterhalb von 0.1, was eine gute Approximation bescheinigt. Die größten Abweichungen ergeben sich für die Beobachtung 6 und 40. Für die Kovariable *rdsales* und die Interaktion bewegt sich die Differenz für die Beobachtungen 40, 47, 52 und 66 betragsmäßig knapp über dem Wert von 0.1. Aber es sind noch annehmbare Approximationen. Lediglich die Beobachtung 6 weist hier höhere Werte auf. Die Differenz jeweils für die Kovariable *rdsales* und die Interaktion beträgt betragsmäßig 2.03 bzw. 0.77. Für die Beobachtung 6 ist die Ein-Schritt-Approximation also nicht ganz zufriedenstellend.

5.4 Bestimmung des Einflusses einer Beobachtung

Ebenso wie in den LM's kann man den Einfluß von einzelnen Beobachtungen auf vorhergesagte Werte und Regressionkoeffizienten feststellen, indem man geeignete Diagnostiken definiert. Wie man im weiteren Verlauf sehen wird, lassen sich die Diagnostiken nur teilweise vereinfachen. Komplexere Ausdrücke sind die Folge. Da man in dieser Arbeit im Besonderen Zähldaten be-

schreiben will, werden zur Veranschaulichung immer wieder Beispiele angeführt, die auf dem Poisson- und auf dem Negativ-Binomial-Fall basieren.

5.4.1 DFFITS und DFBETAS

In den LM's betrachtete man zuerst die $DFFITs_i$, welche die standardisierte Differenz der vorhergesagten Werte mit und ohne Betrachtung der i -ten Beobachtung waren. Im Fall der GLM's wird ebenfalls die standardisierte Differenz der vorhergesagten Werte mit und ohne Betrachtung der i -ten Beobachtung betrachtet. Die $DFFITs_i$ im GLM sind also durch

$$\begin{aligned} DFFITS_i &:= \frac{\hat{\mu}_i - \hat{\mu}_{i,-i}}{\sqrt{\hat{v}_{i,-i} h_{ii}}} \\ &= \frac{g^{-1}(\hat{\eta}_i) - g^{-1}(\hat{\eta}_{i,-i})}{\sqrt{\hat{v}_{i,-i} h_{ii}}} \\ &= \frac{g^{-1}(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}) - g^{-1}(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{-i})}{\sqrt{\hat{v}_{i,-i} h_{ii}}} \end{aligned}$$

definiert, wobei $\hat{v}_{i,-i} = a(\phi)b''(\hat{\theta}_{i,-i})$ die geschätzte Varianz von Y_i ohne Betrachtung der i -ten Beobachtung, $\hat{\theta}_{i,-i} = b'(\hat{\mu}_{i,-i})$, h_{ii} das i -te Diagonalelement der Hatmatrix H und $g(\cdot)$ die Linkfunktion sind. Dieser letzte Ausdruck wird durch die $DFFITs_i^1$ approximiert, indem für den Schätzer $\hat{\boldsymbol{\beta}}_{-i}$ die Ein-Schritt-Approximation (5.15) eingesetzt wird. Also gilt

$$DFFITs_i^1 := \frac{g^{-1}(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}) - g^{-1}(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{-i}^1)}{\sqrt{\hat{v}_{i,-i} h_{ii}}}. \quad (5.16)$$

Wie man sieht, ist es hier weit schwieriger als in den LM's den Ausdruck zu vereinfachen, da der Ausdruck (5.16) von mehreren Funktionen abhängt, wie z.B. von der Inversen der Linkfunktion. Zur Veranschaulichung wird ein Beispiel für den Poissonfall gezeigt.

Beispiel 5.1 (Poissonregression) Sei $Y_i \sim Poi(\mu_i)$. Im Poissonfall ist die Linkfunktion die Log-Link-Funktion, d.h. $\eta_i = \log(\mu_i)$, also $\mu_i = e^{\mathbf{x}_i^T \boldsymbol{\beta}}$. Es gilt zudem $\eta_i = \theta_i$, d.h. ein kanonischer Link, und die Varianz von Y_i ist $v_i = \mu_i$. Für die $DFFITs_i$ gilt

$$\begin{aligned} DFFITS_i &= \frac{g^{-1}(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}) - g^{-1}(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{-i})}{\sqrt{\hat{v}_{i,-i} h_{ii}}} \\ &= \frac{e^{\mathbf{x}_i^T \hat{\boldsymbol{\beta}}} - e^{\mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{-i}}}{\sqrt{e^{\mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{-i}} h_{ii}}}, \end{aligned}$$

wobei für $\hat{\boldsymbol{\beta}}_{-i}$ die Ein-Schritt-Approximation (5.15) eingesetzt wird, um die $DFFITs_i^1$ zu erhalten. Das Diagonalelement von H ist von der Schätzung $\hat{\boldsymbol{\beta}}$ abhängig und ergibt sich aus $h_{ii} = w_i^{1/2} \mathbf{x}_i^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i w_i^{1/2}$ mit

$$\begin{aligned} \mathbf{W} := \mathbf{W}(\hat{\boldsymbol{\beta}}) &= \text{diag} \left(\left(\frac{d\mu_1}{d\eta_1} \right)^2 \Big/ v_1 \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}, \dots, \left(\frac{d\mu_n}{d\eta_n} \right) \Big/ v_n \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \right) \\ &= \text{diag} (\hat{\mu}_1, \dots, \hat{\mu}_n) \\ &= \text{diag} \left(e^{\mathbf{x}_1^T \hat{\boldsymbol{\beta}}}, \dots, e^{\mathbf{x}_n^T \hat{\boldsymbol{\beta}}} \right). \end{aligned}$$

Somit sind alle Elemente von der Schätzung $\hat{\beta}$ abhängig, mit der die $DFFITS_i^1$ nun ausgerechnet werden können.

Der Wert $DFFITS_i$ gibt ebenso wie in den LM's die Zahl der geschätzten Standardfehler für den i-ten Punkt an, für die der geschätzte Wert geändert wird, wenn der i-te Punkt gelöscht wird. Es stellt sich nun aber auch die Frage, wann eine Beobachtung als einflussreich gilt. Dies erscheint in den GLM's auch eine schwierige Frage zu sein. Generell kann man aufgrund der Daten und der erhaltenen Werte $DFFITS_i, \forall i$, erkennen, welche Werte sich stark von den anderen unterscheiden und einen hohen Wert aufweisen. Als angemessene Schranke lässt sich ein Wert von 2 angeben, der auf dem 95%-Niveau der Standardnormalverteilung basiert, da der Ausdruck $DFFITS_i$ asymptotisch standardnormalverteilt ist. Es besteht außerdem die Möglichkeit, denselben Wert wie in den LM's zu verwenden, also $2\sqrt{p/n}$.

Eine weitere Diagnostik, nämlich die $DFBETAS_{j,i}$, die in den LM's definiert wurde, wird nun auch für die GLM's beschrieben. Die $DFBETAS_{j,i}$ in den GLM's sind nachfolgend definiert:

$$DFBETAS_{j,i} := \frac{\hat{\beta}_j - \hat{\beta}_{j,-i}}{\sqrt{c_{jj}}},$$

wobei $c_{jj} = ((\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1})_{jj} \approx \text{Var}(\hat{\beta}_j)$. Mit Hilfe von $\hat{\beta} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{Z}$ und $Z_i = \mathbf{x}_i^T \hat{\beta} + (y_i - \hat{\mu}_i) \frac{d\eta_i}{d\mu_i} \Big|_{\mu_i = \hat{\mu}_i}$ mit $\mathbf{W}(\hat{\beta})$ als konstant betrachtet, gilt

$$\begin{aligned} \text{Var}(\hat{\beta}) &= \text{Var}((\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{Z}) \\ &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \text{Var}(\mathbf{Z}) \mathbf{W} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}. \end{aligned}$$

Nach (5.5) ist $\text{Var}(\mathbf{Z}) \approx \mathbf{W}^{-1}$ und somit

$$\text{Var}(\hat{\beta}) \approx (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}.$$

Mit Hilfe der Ein-Schritt-Approximation (5.15) lässt sich $DFBETAS_{j,i}$ approximieren als

$$\begin{aligned} DFBETAS_{j,i}^1 &:= \frac{\hat{\beta}_j - \hat{\beta}_{j,-i}^1}{\sqrt{c_{jj}}} \\ &= \frac{\hat{\beta}_j - \hat{\beta}_j + \frac{r_{j,i} r_i^P}{1-h_{ii}}}{\sqrt{c_{jj}}} \\ &= \frac{\frac{r_{j,i} r_i^P}{1-h_{ii}}}{\sqrt{c_{jj}}} \\ &= \frac{\frac{r_{j,i} r_i^P}{1-h_{ii}}}{\sqrt{((\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1})_{jj} (\mathbf{X}^T \mathbf{W} \mathbf{X})_{jj} ((\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1})_{jj}}} \\ &= \frac{r_{j,i}}{\sqrt{\mathbf{r}_j^T \mathbf{r}_j}} \frac{r_i^P}{1-h_{ii}}, \end{aligned} \tag{5.17}$$

wobei $R = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2}$ und \mathbf{r}_j^T ist die j-te Zeile von R . Der Vorfaktor $\frac{r_{j,i}}{\sqrt{\mathbf{r}_j^T \mathbf{r}_j}}$ entspricht genauso wie in den LM's einem normalisierenden Maß der i-ten Beobachtung auf den j-ten Regressionskoeffizienten. Desweiteren sieht man, dass der Wert für die $DFBETAS_{j,i}$ von dem

Pearson-Residuum und dem h_{ii} -Wert. Auch hier stellt sich die Frage, ab wann man eine Beobachtung als einflussreich deklarieren kann. Wie bei den $DFFITs_i$ betrachtet man die errechneten Werte und untersucht Punkte näher, die sich von den anderen merklich unterscheiden und somit einen hohen Wert aufweisen. Hierbei kann basierend auf dem 95%-Niveau der Standardnormalverteilung ein kritischer Wert von 2 angegeben werden, da der Ausdruck $DFBETAS_{j,i}$ asymptotisch standardnormalverteilt ist. Es lässt sich darüberhinaus die Schranke der $DFBETAS_{j,i}$ bei den LM's verwenden, d.h. $2\sqrt{1/n}$.

Beispiel 5.2 (Poissonregression) Sei wieder $Y_i \sim Poi(\mu_i)$ und alles weitere wie in Beispiel 5.1. Für die $DFBETAS_{j,i}^1$ gilt

$$\begin{aligned} DFBETAS_{j,i}^1 &= \frac{r_{j,i}}{\sqrt{\mathbf{r}_j^T \mathbf{r}_j}} \frac{r_i^P}{1 - h_{ii}} \\ &= \frac{r_{j,i}}{\sqrt{\mathbf{r}_j^T \mathbf{r}_j}} \frac{r_i^{P*}}{\sqrt{1 - h_{ii}}} \end{aligned}$$

wobei $r_i^{P*} = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{v}_i(1 - h_{ii})}}$ das standardisierte Pearson-Residuum mit

$$\begin{aligned} \hat{\mu}_i &= \exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}), \\ \hat{v}_i &= \hat{\mu}_i = \exp(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}). \end{aligned}$$

Das Diagonalelement von H ist von der Schätzung $\hat{\boldsymbol{\beta}}$ abhängig und ergibt sich somit aus $h_{ii} = w_i^{1/2} \mathbf{x}_i^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i w_i^{1/2}$ mit \mathbf{W} wie in Beispiel 5.1. Somit sind wieder alle Elemente von der Schätzung $\hat{\boldsymbol{\beta}}$ abhängig und der Wert der $DFBETAS_{j,i}^1$ kann berechnet werden.

Als konkretes Datenbeispiel soll der Patentdatensatz dienen. In Tabelle 5.4 sind die Werte der $DFFITs_i^1$ angegeben, die durch die geschriebene SPlus-Funktion *diag.poi* ausgegeben werden können. In Abbildung 5.5 sind die Werte mittels eines Stabdiagramms veranschaulicht jeweils für die $DFFITs_i^1$ und die $DFFITs_i$ basierend auf dem exakten geschätzten Wert für $\hat{\boldsymbol{\beta}}_{-i}$. Die kritische Schranke von 2, welche einen absoluten Wert angibt, ist in der Abbildung 5.5 eingetragen. Gestrichelt ist ebenso die Schranke von $2\sqrt{p/n} = 2\sqrt{4/70} = 0.48$ eingetragen. Aufgrund des hohen Wertes für die Beobachtung 6 nicht eindeutig sichtbar. Deswegen sind in Abbildung 5.5 die unteren beiden Grafiken, die ebenso die $DFFITs_i^1$ und die $DFFITs_i$ aufzeigen, nochmals skaliert abgetragen, so dass die Schranken nun besser sichtbar sind.

Außerdem erkennt man aus der Grafik, dass die Ein-Schritt-Approximationen recht gut sind. Lediglich die Approximation für $DFFITs_6$ unterscheidet sich deutlich vom exakten Wert. Für die Approximation liegt der Wert absolut bei 368.38, der exakte Wert ist absolut dagegen 88. Um sich die Beobachtungen genauer anzusehen, die die kritische Schranke überschreiten, wurde eine weitere Funktion *inflbeob.dffits* geschrieben. Diese gibt die absoluten Werte und die Indizes der Beobachtungen aus. In Tabelle 5.5 erkennt man, dass die Beobachtung 6 mit einem absolutem Wert von 368.38 alle anderen Werte überragt. Darüberhinaus fallen wieder die Beobachtungen 47 und 40 auf. Wenn man sich den Datensatz ansieht, stellt man fest, dass diese Beobachtungen sehr hohe Werte im Response *patente* haben. Die Beobachtung 40 besitzt eine Anzahl von 105 Patenten und die Beobachtung 47 sogar den höchsten Wert mit einer Anzahl von 173 Patenten. Die gestrichelte Schranke überschreiten u.a. die Beobachtungen 65 (SMITHKLINE CORP) und 66, die ebenso hohe Werte in der Anzahl der Patente aufweisen.

\$dffits:

[1]	-0.69755704	-0.03391015	-0.01931477	-0.32488381	-0.13075873
[6]	-368.37880178	-0.12175262	0.39375800	0.03192132	-0.42642934
[11]	0.03040401	0.51367521	0.11202139	-0.03904194	-0.30888022
[16]	-0.42567988	-0.45735603	0.20934572	-0.05878027	-0.22465570
[21]	-0.05182741	-0.24710939	0.05690003	0.01222719	0.04211942
[26]	-0.14880926	0.07343683	0.00702559	0.01389784	-0.04824969
[31]	-0.04455050	-0.15046360	-0.08065738	-0.38497207	0.04430534
[36]	0.01457096	-0.26219363	-0.30790525	0.93131850	-2.00006531
[41]	-0.09922338	-0.06966501	-0.15188802	-0.27749581	-0.12750011
[46]	-0.06449559	-3.99146048	-0.03695000	0.43057517	-0.12583745
[51]	-0.09501825	-0.43767735	-0.11199350	-0.48701555	-0.08101786
[56]	-0.09608767	-0.05034985	-0.40800935	-0.26141551	-0.13128600
[61]	-0.02501937	-0.21232323	0.63344828	0.02019873	1.25384007
[66]	1.93490722	0.31617477	-0.15586496	1.01798841	-0.21731097

Tabelle 5.4: $DFFITS_i^1$ des Poissonmodells für den Patentdatensatz

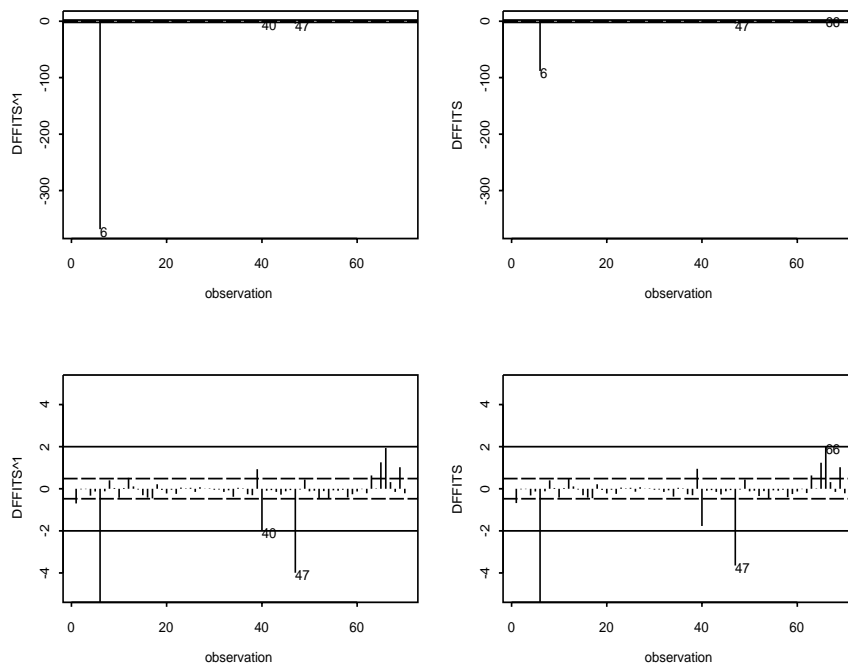


Abbildung 5.5: Stabdiagramm der $DFFITS_i^1$ (links) und der $DFFITS_i$ (rechts) des Poissonmodells für den Patentdatensatz ohne Skalierung (oben) und mit Skalierung ($|\cdot| \leq 5$) der y-Achse (unten). Die gestrichelte bzw. durchgezogene Linie gibt jeweils die absoluten Schranken $2\sqrt{p/n} = 2\sqrt{4/70} = 0.48$ und 2 an.

DFFITS

\$eb:

```
[1] 368.3788018  3.9914605  2.0000653
```

\$b:

```
[1]  6 47 40
```

Tabelle 5.5: $DFFITS_i^1$ mit $DFFITS_i^1 \geq 2$ des Poissonmodells für den Patentdatensatz

Nun lassen sich auch die $DFBETAS_{j,i}^1$ untersuchen. Die SPlus-Funktion, die dazu geschrieben wurde, wird mit `dfbetas.poi` aufgerufen und ist im Anhang definiert. In Abbildung 5.6 sind diese Werte zu sehen.

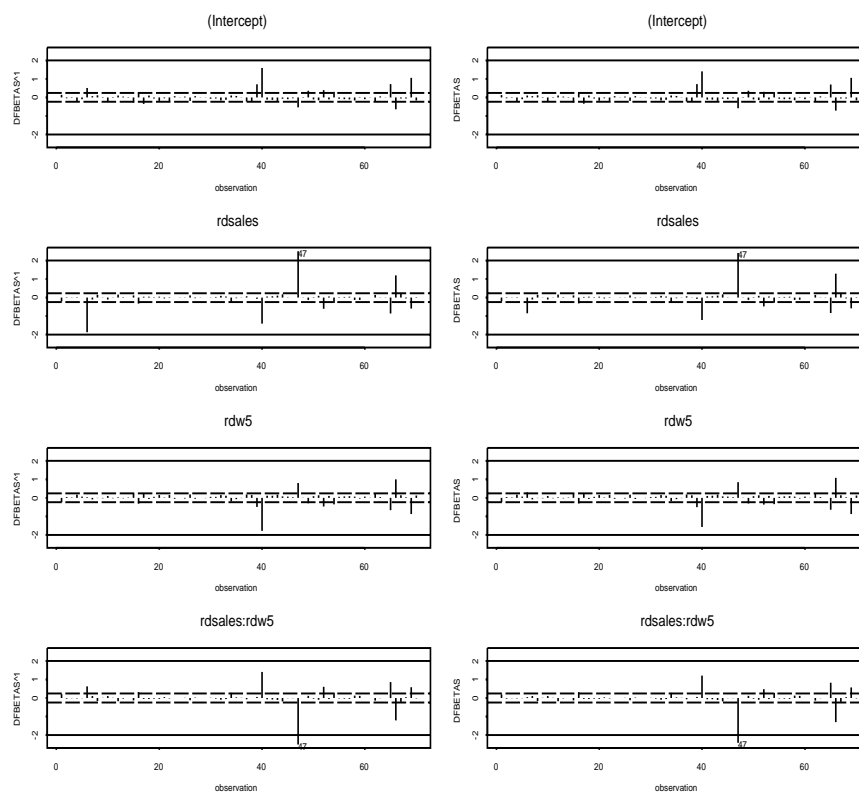


Abbildung 5.6: Stabdiagramme der $DFBETAS_{j,i}^1$ (links) und der $DFBETAS_{j,i}$ (rechts) des Poissonmodells für den Patentdatensatz. Die gestrichelte bzw. durchgezogene Linie markiert die absoluten Schranken $2\sqrt{1/n} = 2\sqrt{1/70} = 0.24$ und 2.

Die Abbildung 5.6 gibt für den Intercept und jede Kovariable die Veränderung an. Links sind jeweils die Ein-Schritt-Approximationen und rechts die exakten Werte angegeben. Die kritische Schranke von 2, welche eine absolute Schranke darstellt und in der Grafik jeweils eingezeichnet ist. Ebenso ist gestrichelt die Schranke $2\sqrt{1/n} = 2\sqrt{1/70} = 0.24$ angedeutet. Es fallen die Werte der Beobachtung 47 besonders auf, die für die Kovariable `rdsales` und die Interaktion die Schranke von 2 überschreiten. Berücksichtigt man die gestrichelte Schranke, so zeigen 6, 16, 40,

52, 54, 65, 66 und 69 in allen vier Regressionkoeffizienten signifikante Veränderungen auf. Die Approximationen sind bis auf die Beobachtung 6 akzeptabel.

5.4.2 Cook's Distanz

Nun wird die Cook's Distanz für GLM's entwickelt. Wie in Kapitel 3.5.2 angedeutet, kann diese auch über die Konfidenzbereiche hergeleitet werden. Dies wird im folgenden kurz erläutert. Es ist eine Taylor-Approximation nötig, um den Log-Likelihood an der Stelle β zu berechnen. Diese lautet wie folgt:

$$\mathcal{L}(\beta) \approx \mathcal{L}(\hat{\beta}) + (\beta - \hat{\beta})^T S(\hat{\beta}) - \frac{1}{2}(\beta - \hat{\beta})^T FI(\hat{\beta})(\beta - \hat{\beta}), \quad (5.18)$$

wobei $S(\hat{\beta}) = \left. \frac{\partial \mathcal{L}}{\partial \beta} \right|_{\beta=\hat{\beta}}$ der Vektor der Scorefunktionen und $FI(\hat{\beta}) = -E \left(\left. \frac{\partial^2 \mathcal{L}}{\partial \beta \partial \beta^T} \right|_{\beta=\hat{\beta}} \right)$ die Fisher-Informationsmatrix sind. Nach Definition maximiert der Schätzer $\hat{\beta}$ $\mathcal{L}(\hat{\beta})$ und somit $S(\hat{\beta}) = \mathbf{0}$. Mit Hilfe von Gleichung (5.18) bekommt man also

$$\mathcal{L}(\beta) - \mathcal{L}(\hat{\beta}) \approx -\frac{1}{2}(\beta - \hat{\beta})^T FI(\hat{\beta})(\beta - \hat{\beta}), \quad (5.19)$$

da $S(\hat{\beta}) = \mathbf{0}$. Da die asymptotische Verteilung von $\hat{\beta}$ eine p -dimensionale Normalverteilung ist, d.h. $\hat{\beta} \sim N_p(\beta, FI^{-1}(\beta))$, gilt asymptotisch

$$(\beta - \hat{\beta})^T FI(\hat{\beta})(\beta - \hat{\beta}) \sim \chi_p^2. \quad (5.20)$$

Ein approximativer $100(1 - \alpha)\%$ Likelihood-Konfidenzbereich für β ist gegeben durch

$$\left\{ \beta : -2 \log \left(\frac{l(\beta)}{l(\hat{\beta})} \right) \leq \chi_{1-\alpha, p}^2 \right\} = \left\{ \beta : -2 \left(\mathcal{L}(\beta) - \mathcal{L}(\hat{\beta}) \right) \leq \chi_{1-\alpha, p}^2 \right\}, \quad (5.21)$$

wobei $l(\beta)$ die Likelihood-Funktion ist und $\chi_{1-\alpha, p}^2$ das $(1 - \alpha)$ -Quantil einer χ_p^2 -Verteilung darstellt. Die Differenz zwischen den Log-Likelihoods jeweils ausgewertet für β und β_{-i} , ergibt die *Log-Likelihood Distanz*:

$$LD_i = 2 \left(\mathcal{L}(\hat{\beta}) - \mathcal{L}(\hat{\beta}_{-i}) \right) = -2 \log \left(\frac{l(\hat{\beta}_{-i})}{l(\hat{\beta})} \right). \quad (5.22)$$

Der Logarithmand der letzten Gleichung ist die Likelihood-Ratio-Statistik. Hier misst der Wert LD_i die Veränderung des Konfidenzbereichs für β , falls die i -te Beobachtung weggelassen wird. Nach obigen Berechnungen gilt

$$\begin{aligned} \left\{ \beta : -2 \left(\mathcal{L}(\beta) - \mathcal{L}(\hat{\beta}) \right) \leq \chi_{1-\alpha, p}^2 \right\} &\approx \left\{ \beta : (\beta - \hat{\beta})^T FI(\hat{\beta})(\beta - \hat{\beta}) \leq \chi_{1-\alpha, p}^2 \right\} \\ &\approx \left\{ \beta : (\beta - \hat{\beta})^T (\mathbf{X}^T \mathbf{W} \mathbf{X})(\beta - \hat{\beta}) \leq \chi_{1-\alpha, p}^2 \right\}. \end{aligned}$$

Im Artikel von Williams (1987) wird die Cook's Distanz für ein GLM definiert. Demnach lautet die Cook's Distanz im GLM nun folgendermaßen:

$$D_i \approx \frac{1}{p} (\hat{\beta} - \hat{\beta}_{-i})^T (\mathbf{X}^T \mathbf{W} \mathbf{X}) (\hat{\beta} - \hat{\beta}_{-i}).$$

Eine Teilung durch p bewirkt eine Normierung, da der Erwartungswert einer χ^2 -Verteilung gleich der Anzahl der Freiheitsgrade ist, also hier p . Dieser Schritt dient ebenso zur Festlegung des kritischen Werts. Um den Ausdruck etwas einfacher und interpretierbarer zu machen, verwendet man die Ein-Schritt-Approximation (5.15). Also ergibt sich

$$\begin{aligned} D_i^1 &= \frac{1}{p} (\hat{\beta} - \hat{\beta}_{-i}^1)^T (\mathbf{X}^T \mathbf{W} \mathbf{X}) (\hat{\beta} - \hat{\beta}_{-i}^1) \\ &\stackrel{(5.15)}{=} \frac{1}{p} \left(w_i^{1/2} \frac{(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i r_i^P}{1 - h_{ii}} \right)^T (\mathbf{X}^T \mathbf{W} \mathbf{X}) \left(w_i^{1/2} \frac{(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i r_i^P}{1 - h_{ii}} \right) \\ &= \frac{1}{p} \frac{w_i^{1/2} \mathbf{x}_i^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i w_i^{1/2} (r_i^P)^2}{(1 - h_{ii})^2} \\ &= \frac{1}{p} \left(\frac{h_{ii}}{1 - h_{ii}} \right) (r_i^{P*})^2. \end{aligned} \tag{5.23}$$

Ein kritischer Wert für eine einflußreiche Beobachtung ist der Wert 1. Eine Erklärung für den Wert 1 wurde bereits in Kapitel 3.5.2 gegeben. Wie bereits beschrieben, ist die Cook's Distanz, welche ein allgemeines Diskrepanzmaß ist und von Cook (1977) vorgeschlagen wurde, durch Konfidenzbereiche definiert. Im Artikel von Pregibon (1981) wird dies auch aufgegriffen und es werden zwei verschiedene Konfidenzbereiche definiert.

Für die Cook's Distanz soll ebenso das Patentdaten-Beispiel für das Poissonmodell betrachtet werden. In Tabelle 5.6 sind die Werte der Cook's Distanz für die Ein-Schritt-Approximation angegeben, die durch die geschriebene SPlus-Funktion *diag.poi* ausgegeben werden können. In Abbildung 5.7 sind die Werte mittels eines Stabdiagramms veranschaulicht. Da die Beobachtung 6 einen sehr hohen Wert (670.6) aufweist, werden die Grafiken skaliert, was durch die unteren beiden Grafiken dargestellt wird.

Auch die Beobachtungen 40, 47 und 66 werden angezeigt, was aus der geschriebenen SPlus-Funktion *inflbeob.cook* hervorgeht. In Tabelle 5.7 werden die geordneten Werte bzw. die Indizes der Beobachtungen, die die kritische Schranke von 1 überschreiten angegeben.

Die Approximationen zeigen hier ein Überschätzen der wahren Werte für die Cook's Distanz. Deutlich zeigt sich dies wieder bei der Beobachtung 6, aber auch die Beobachtungen 40, 47, 65 und 69 werden sichtbar überschätzt.

\$cook:

```
[1] 1.513621e-001 1.319100e-003 1.205642e-004 3.187319e-002 4.520184e-003
[6] 6.705644e+002 3.709225e-003 6.934207e-002 2.613606e-004 7.899560e-002
[11] 2.387027e-004 1.649655e-001 7.624032e-003 6.817241e-004 2.488997e-002
[16] 4.570633e-002 8.369754e-002 1.676681e-002 1.502955e-003 1.559113e-002
[21] 1.941389e-002 2.035940e-002 8.939442e-003 3.737928e-005 2.325230e-003
[26] 1.672173e-002 1.367597e-003 3.521860e-005 1.497674e-004 1.732239e-002
[31] 1.918734e-003 8.327096e-003 1.537642e-002 4.896365e-002 1.024328e-002
[36] 6.100980e-005 1.737183e-002 2.376379e-002 5.693840e-001 1.543202e+000
[41] 5.548226e-003 1.439934e-002 1.675573e-002 1.980237e-002 6.155641e-003
[46] 1.085905e-003 4.880744e+000 1.041205e-003 4.780025e-002 4.082509e-003
[51] 1.510971e-002 1.811727e-001 3.855256e-003 6.625494e-002 1.885977e-003
[56] 3.331417e-003 1.457636e-002 7.735644e-002 3.745634e-002 8.739019e-003
[61] 2.897904e-004 1.900533e-002 1.735651e-001 2.084404e-004 7.735563e-001
[66] 1.145749e+000 3.798477e-002 1.337077e-002 9.948464e-001 1.380990e-002
```

Tabelle 5.6: Cook's Distanz D_i^1 des Poissonmodells für den Patentdatensatz

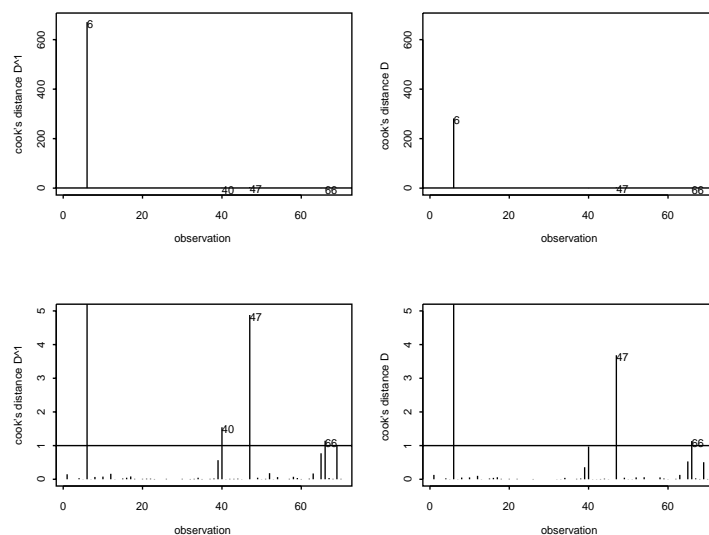


Abbildung 5.7: Stabdiagramm der Cook's Distanz D_i^1 (links) und D_i (rechts) des Poissonmodells für den Patentdatensatz ohne Skalierung (oben) und mit Skalierung ($|\cdot| \leq 5$) der y-Achse (unten). Die durchgezogene Linie markiert die Schranke 1.

Cook's Distanz

\$eb:

```
[1] 670.564419 4.880744 1.543202 1.145749
```

\$b:

```
[1] 6 47 40 66
```

Tabelle 5.7: Cook's Distanz D_i^1 mit $D_i^1 > 1$ des Poissonmodells für den Patentdatensatz

5.4.3 COVRATIO

Genau wie in den LM's lässt sich die Frage nach der 'besseren' Anpassung eines Modells mit oder ohne i -te Beobachtung noch nicht beantworten. Darum betrachtet man auch in den GLM's die allgemeine Varianz AG der Regressionskoeffizienten. Nach der Verteilungstheorie des ML-Schätzers, welche in Fahrmeir und Kaufmann (1985) (Seite 362) angegeben ist, gilt asymptotisch

$$\hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, FI^{-1}(\boldsymbol{\beta}))$$

mit $FI^{-1}(\boldsymbol{\beta}) = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}$ und $\mathbf{W} := \mathbf{W}(\boldsymbol{\beta})$. Also ist die allgemeine Varianz AG im GLM definiert durch

$$AG := \det(\text{Var}(\hat{\boldsymbol{\beta}})) = \det(FI^{-1}(\boldsymbol{\beta})). \quad (5.24)$$

Um den Einfluß der i -ten Beobachtung zu erfassen, definiert man wie in den LM's einen Kovarianz-Ratio mit Hilfe der geschätzten Varianz $\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}$ mit $\mathbf{W} := \mathbf{W}(\hat{\boldsymbol{\beta}})$

$$\begin{aligned} COVRATIO_i &:= \frac{\det(\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_{-i}))}{\det(\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}))} \\ &= \frac{\det((\mathbf{X}_{-i}^T \mathbf{W}_{-i} \mathbf{X}_{-i})^{-1})}{\det((\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1})}, \end{aligned} \quad (5.25)$$

wobei $\mathbf{W}_{-i} := \mathbf{W}_{-i}(\hat{\boldsymbol{\beta}}_{-i})$. Mit Hilfe der Eigenschaften der Determinante, welche z.B. in Rade und Westergren (1997) (Seite 90) aufgeführt sind, gilt für den $COVRATIO_i$:

$$\begin{aligned} COVRATIO_i &= \frac{\det(\mathbf{X}^T \mathbf{W} \mathbf{X})}{\det(\mathbf{X}_{-i}^T \mathbf{W}_{-i} \mathbf{X}_{-i})} \\ &= \frac{\det(\mathbf{X}^T \mathbf{X})}{\det(\mathbf{X}_{-i}^T \mathbf{X}_{-i})} \frac{\det(\mathbf{W})}{\det(\mathbf{W}_{-i})}. \end{aligned}$$

Das Lemma 3.2 kann hier für den ersten Faktor verwendet werden. Der zweite Faktor entspricht einem Quotient von Determinanten von Diagonalmatrizen. Weiter gilt also:

$$COVRATIO_i = \frac{1}{1 - h_{ii}^{LM}} \frac{\prod_{j=1}^n w_j(\hat{\boldsymbol{\beta}})}{\prod_{j=1, j \neq i}^n w_j(\hat{\boldsymbol{\beta}}_{-i})}, \quad (5.26)$$

wobei $h_{ii}^{LM} = \mathbf{x}_i (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i^T$ und $w_j(\hat{\boldsymbol{\beta}})$ und $w_j(\hat{\boldsymbol{\beta}}_{-i})$ die Diagonalelemente der Matrizen $\mathbf{W}(\hat{\boldsymbol{\beta}})$ bzw. $\mathbf{W}_{-i}(\hat{\boldsymbol{\beta}}_{-i})$ sind. Dieser letzte Ausdruck wird durch den $COVRATIO_i^1$ approximiert, indem für den Schätzer $\hat{\boldsymbol{\beta}}_{-i}$ die Ein-Schritt-Approximation (5.15) eingesetzt wird. Also gilt

$$COVRATIO_i^1 = \frac{1}{1 - h_{ii}^{LM}} \frac{\prod_{j=1}^n w_j(\hat{\boldsymbol{\beta}})}{\prod_{j=1, j \neq i}^n w_j(\hat{\boldsymbol{\beta}}_{-i}^1)}. \quad (5.27)$$

Im folgenden Beispiel wird der $COVRATIO_i$ anhand des Poissonfalls ausgerechnet.

Beispiel 5.3 (Poissonregression) Sei wieder $Y_i \sim Poi(\mu_i)$ und alles weitere wie in Beispiel 5.1. Für den $COVRATIO_i$ gilt

$$\begin{aligned} COVRATIO_i &= \frac{1}{1 - h_{ii}^{LM}} \frac{\prod_{j=1}^n w_j(\hat{\boldsymbol{\beta}})}{\prod_{j=1, j \neq i}^n w_j(\hat{\boldsymbol{\beta}}_{-i})} \\ &= \frac{1}{1 - h_{ii}^{LM}} \frac{\prod_{j=1}^n e^{\mathbf{x}_j^T \hat{\boldsymbol{\beta}}}}{\prod_{j=1, j \neq i}^n e^{\mathbf{x}_j^T \hat{\boldsymbol{\beta}}_{-i}}}, \end{aligned}$$

wobei sich der Wert für $w_j(\hat{\boldsymbol{\beta}})$ aus dem Diagonalelement der in Beispiel 5.1 berechneten Matrix \mathbf{W} ergibt. Nun wird für $\hat{\boldsymbol{\beta}}_{-i}$ die Ein-Schritt-Approximation (5.15) eingesetzt und damit ergibt sich

$$\begin{aligned} COVRATIO_i^1 &= \frac{1}{1 - h_{ii}^{LM}} \frac{\prod_{j=1}^n e^{\mathbf{x}_j^T \hat{\boldsymbol{\beta}}}}{\prod_{j=1, j \neq i}^n e^{\mathbf{x}_j^T \left(\hat{\boldsymbol{\beta}} - w_i^{1/2} \frac{(\mathbf{x}^T \mathbf{W} \mathbf{x})^{-1} \mathbf{x}_i r_i^P}{1 - h_{ii}} \right)}} \\ &= \frac{1}{1 - h_{ii}^{LM}} \frac{\prod_{j=1}^n e^{\mathbf{x}_j^T \hat{\boldsymbol{\beta}}}}{e^{-\mathbf{x}_i^T \hat{\boldsymbol{\beta}}} \prod_{j=1}^n e^{\mathbf{x}_j^T \hat{\boldsymbol{\beta}}} \prod_{j=1, j \neq i}^n e^{-w_i^{1/2} \frac{\mathbf{x}_j^T (\mathbf{x}^T \mathbf{W} \mathbf{x})^{-1} \mathbf{x}_i r_i^P}{1 - h_{ii}}}} \\ &= \frac{1}{1 - h_{ii}^{LM}} e^{\mathbf{x}_i^T \hat{\boldsymbol{\beta}}} \prod_{j=1, j \neq i}^n e^{w_i^{-1/2} \frac{h_{ji} r_i^P}{1 - h_{ii}}}, \end{aligned}$$

wobei h_{ii} und r_i^P wie in Beispiel 5.1 bzw. 5.2 gegeben sind und das Diagonalelement w_i aus der Ein-Schritt-Approximation, an der Stelle $\hat{\boldsymbol{\beta}}$ berechnet wird, d.h. $w_i := w_i(\hat{\boldsymbol{\beta}})$. Das Element h_{ji} aus der Hatmatrix H ist das Element in der j -ten Zeile und i -ten Spalte der Hatmatrix $H = \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2}$.

Falls die i -te Beobachtung geringen Einfluß hat, sollte sich der $COVRATIO_i$ in der Nähe von 1 bewegen. Eine allgemein gültige Schranke ist schwierig anzugeben. Die erhaltenen Werte müssen verglichen und auf mögliche auffallende Werte untersucht werden.

Die Anpassung an ein Modell mit Hilfe des $COVRATIO_i$ wird konkret anhand des Patentdatensatzes für das Poissonmodell untersucht. Um für einen Datensatz die Werte des $COVRATIO_i^1$ auszurechnen, wurde die SPlus-Funktion `covratio.poi` geschrieben, welche im Anhang erklärt wird. In Tabelle 5.8 stehen in der ersten Komponente die geordneten Werte für den $COVRATIO_i^1$ und in der zweiten der zugehörige Index der Beobachtung. Es zeigt sich, dass fast alle Beobachtungen einen sehr verschiedenen Wert zu 1 besitzen. In Abbildung 5.8 sind die Werte in einer Grafik jeweils für die Approximationen und die exakten Werte als Stabdiagramm veranschaulicht.

Die Approximationen für die Beobachtungen 6 und 47 sind wieder sehr groß im Gegensatz zum exakten Wert. Die übrigen Werte werden gut approximiert wie man aus der Grafik erkennen kann.

\$covratio:

[1]	2936.4757751	1998.9702267	1021.6233088	955.4296829	889.8945483
[6]	579.2399981	563.7636835	198.4512960	192.6052463	181.2163197
[11]	128.9147035	75.3281310	72.2043454	69.3774436	67.5294860
[16]	63.5428537	34.2679845	33.9945473	28.5434272	23.4909337
[21]	19.4161321	16.0597423	12.6370156	11.6954672	10.1679979
[26]	9.2225316	8.9971679	7.2942304	6.9883633	6.3415999
[31]	6.1338419	5.8680432	5.1286870	4.1760560	4.0716846
[36]	3.9873563	3.9095980	3.3993236	3.3353831	3.3015032
[41]	3.2720607	3.2326992	3.1231240	2.9899790	2.8506932
[46]	2.7818904	2.5380291	2.4782163	2.4611995	2.4120599
[51]	2.3273189	2.1891120	2.0398879	2.0355904	1.7753457
[56]	1.7139656	1.6600755	1.6401576	1.5433140	1.3508667
[61]	1.2975442	1.2948508	1.1556408	1.0941204	0.9499861
[66]	0.9341416	0.8960693	0.8758496	0.7721484	0.7553846

\$beob:

[1]	47	69	6	39	40	65	66	44	63	12	54	16	67	8	64	7	13	1	49	48	34	59	61	68
[25]	52	18	60	9	45	27	24	46	11	58	36	70	20	5	50	10	55	14	3	29	4	25	23	56
[49]	53	38	19	37	28	17	32	22	15	41	35	31	62	2	43	26	33	21	30	51	42	57		

Tabelle 5.8: Geordnete Werte des $COVRATIO_i^1$ des Poissonmodells für den Patentdatensatz

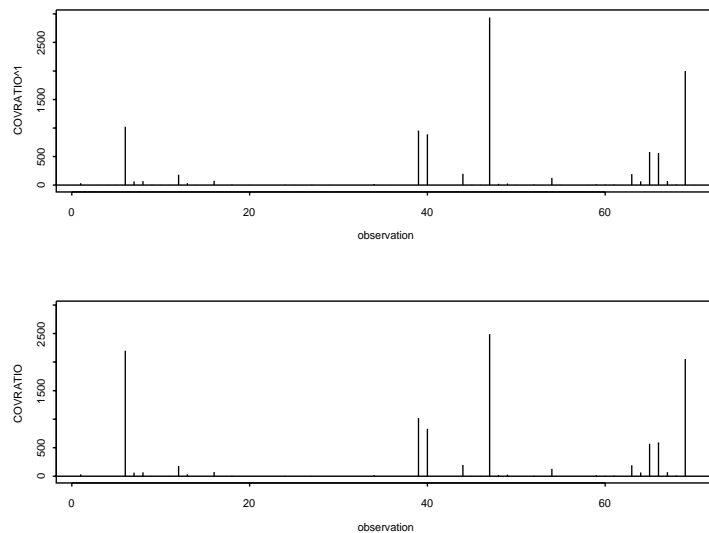


Abbildung 5.8: Stabdiagramm des $COVRATIO_i^1$ (oben) und $COVRATIO_i$ (unten) des Poissonmodells für den Patentdatensatz

5.4.4 Reduktion der Devianz

Ein wichtiger Faktor bei der Aufdeckung von Ausreißern und einflußreichen Beobachtungen in den GLM's ist die Veränderung der Anpassungsgüte. Man ist daher interessiert, wie sich die Devianz ändert, wenn man die i -te Beobachtung bei der Schätzung weglässt. Man will also solche Punkte identifizieren, die einen wesentlichen Einfluß auf die gefitteten Werte haben. Veränderungen in der Devianz können zwei grundsätzliche Formen haben:

1. Falls der i -te Punkt nicht gut vom Modell angepasst wird, sind die Änderungen in D gewöhnlich in der einzelnen Komponente d_i sichtbar, d.h. ist die Komponente d_i groß, so ist auch der absolute Wert für das Devianz-Residuum groß und damit ist die i -te Beobachtung ein y -Ausreißer.
2. Falls der i -te Punkt einen extremen Wert in der Design-Matrix aufweist, resultieren die Veränderungen in D aus allen einzelnen geänderten Komponenten d_j , $j = 1, \dots, n$.

Eine große Änderung sagt nicht aus, welcher von den obigen Fällen eingetreten ist. Aber weitere Diagnostiken, wie z.B. h_{ij} , sollten Aufschluss darüber geben.

Um die Devianzänderungen auszurechnen, wird zuerst die Devianz wie in (4.12) geschrieben, also

$$D(\mathbf{y}, \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^n [\mathcal{L}(y_i; y_i) - \mathcal{L}(y_i; \hat{\mu}_i)]. \quad (5.28)$$

Da im weiteren Verlauf mit der Ein-Schritt-Approximation gearbeitet wird und diese Abhängigkeit direkt sichtbar sein soll, definiert man die Devianz als

$$D(\mathbf{y}, \hat{\boldsymbol{\mu}}) := D(\mathbf{y}, \hat{\boldsymbol{\beta}}),$$

wobei $\hat{\boldsymbol{\eta}} = g(\hat{\boldsymbol{\mu}}) = \mathbf{X}\hat{\boldsymbol{\beta}}$ gilt. Die Darstellung soll die Abhängigkeit von dem Schätzwert $\hat{\boldsymbol{\beta}}$ widerspiegeln. Die Devianz ohne i -te Beobachtung wird als

$$D_{-i}(\mathbf{y}, \hat{\boldsymbol{\mu}}) := 2 \sum_{j \neq i} [\mathcal{L}(y_j; y_j) - \mathcal{L}(y_j; \hat{\mu}_j)]$$

definiert.

Die Devianz kann nun für die Ein-Schritt-Approximation $\hat{\boldsymbol{\beta}}_{-i}^1$ ausgewertet werden. Um einen approximativen Wert zu erhalten, wendet man eine Taylorentwicklung zweiter Ordnung um $\hat{\boldsymbol{\beta}}$ an. Man erhält

$$\begin{aligned} D_{-i}(\mathbf{y}, \hat{\boldsymbol{\beta}}_{-i}^1) &\approx D_{-i}(\mathbf{y}, \hat{\boldsymbol{\beta}}) + (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{-i}^1)^T \left. \frac{\partial D_{-i}(\mathbf{y}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \\ &\quad + \frac{1}{2} (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{-i}^1)^T \left. \frac{\partial^2 D_{-i}(\mathbf{y}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{-i}^1). \end{aligned} \quad (5.29)$$

Da die Devianz $D_{-i}(\mathbf{y}, \hat{\boldsymbol{\beta}}_{-i}^1)$ auf Basis von $n - 1$ Beobachtungen ausgerechnet wird, wird die Approximation schwieriger und mehrere Schritte sind nötig. Es gilt:

$$\begin{aligned}
 D_{-i}(\mathbf{y}, \hat{\boldsymbol{\beta}}) &= 2 \sum_{j \neq i} [\mathcal{L}(y_j; y_j) - \mathcal{L}(y_j; \hat{\mu}_j)] \\
 &= 2 \sum_{j=1}^n [\mathcal{L}(y_j; y_j) - \mathcal{L}(y_j; \hat{\mu}_j)] - \underbrace{2(\mathcal{L}(y_i; y_i) - \mathcal{L}(y_i; \hat{\mu}_i))}_{=d_i^2} \\
 &= D(\mathbf{y}, \hat{\boldsymbol{\beta}}) - d_i^2.
 \end{aligned} \tag{5.30}$$

Dies lässt sich in die obige Taylorentwicklung (5.29) einsetzen, so dass folgendes gilt:

$$\begin{aligned}
 D_{-i}(\mathbf{y}, \hat{\boldsymbol{\beta}}_{-i}^1) &\approx D(\mathbf{y}, \hat{\boldsymbol{\beta}}) - d_i^2 + (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{-i}^1)^T \left(\frac{\partial D(\mathbf{y}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} - \frac{\partial d_i^2}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \right) + \\
 &\quad + \frac{1}{2} (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{-i}^1)^T \left(\frac{\partial^2 D(\mathbf{y}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} - \frac{\partial^2 d_i^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \right) (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{-i}^1).
 \end{aligned} \tag{5.31}$$

Dieser Ausdruck lässt sich mit Hilfe der Ein-Schritt-Approximation (5.15) und einfachen Rechenschritten vereinfachen. Es gilt nämlich

$$\begin{aligned}
 \frac{\partial D(\mathbf{y}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} &= \mathbf{0} \\
 2 \frac{\partial d_i^2}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} &= 2 \frac{\partial (\mathcal{L}(y_i; y_i) - \mathcal{L}(y_i; \mu_i))}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = \mathbf{0} \\
 -E \left(\frac{\partial^2 D(\mathbf{y}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \right) &= 2 \mathbf{X}^T \mathbf{W} \mathbf{X} \\
 -E \left(2 \frac{\partial^2 d_i^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \right) &= 2 \frac{\partial^2 (\mathcal{L}(y_i; y_i) - \mathcal{L}(y_i; \mu_i))}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = 2 \mathbf{x}_i w_i \mathbf{x}_i^T.
 \end{aligned}$$

Die ersten beiden Gleichungen ergeben sich aus den Scoregleichungen (4.5) und aus der Eigenschaft $D = \sum d_i$. Die letzten beiden Gleichungen wurden nach der Fisher-Scoring-Methode berechnet, d.h. man benutzt anstelle der beobachteten Informationsmatrix die erwartete Informationsmatrix. Dies lässt sich nun in obigen Term (5.31) einsetzen. Damit ergibt sich

$$\begin{aligned}
 D_{-i}(\mathbf{y}, \hat{\boldsymbol{\beta}}_{-i}^1) &\approx D(\mathbf{y}, \hat{\boldsymbol{\beta}}) - d_i^2 - \\
 &\quad - \left(w_i^{1/2} \frac{(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i r_i^P}{1 - h_{ii}} \right)^T (\mathbf{X}^T \mathbf{W} \mathbf{X}) \left(w_i^{1/2} \frac{(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i r_i^P}{1 - h_{ii}} \right) + \\
 &\quad + \left(w_i^{1/2} \frac{(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i r_i^P}{1 - h_{ii}} \right)^T (\mathbf{x}_i w_i \mathbf{x}_i^T) \left(w_i^{1/2} \frac{(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i r_i^P}{1 - h_{ii}} \right) \\
 &= D(\mathbf{y}, \hat{\boldsymbol{\beta}}) - d_i^2 - \overbrace{\frac{w_i^{1/2} \mathbf{x}_i^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i w_i^{1/2} (r_i^P)^2}{(1 - h_{ii})^2}}^{=h_{ii}} +
 \end{aligned}$$

$$\begin{aligned}
& + \frac{\overbrace{w_i^{1/2} \mathbf{x}_i^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i w_i^{1/2}}^{=h_{ii}} \overbrace{w_i^{1/2} \mathbf{x}_i^T (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i w_i^{1/2}}^{=h_{ii}} (r_i^P)^2}{(1 - h_{ii})^2} \\
& = D(\mathbf{y}, \hat{\boldsymbol{\beta}}) - d_i^2 - \frac{h_{ii}(1 - h_{ii})(r_i^{P*})^2}{1 - h_{ii}} \\
& = D(\mathbf{y}, \hat{\boldsymbol{\beta}}) - (d_i^2 + h_{ii}(r_i^{P*})^2). \tag{5.32}
\end{aligned}$$

Nun lässt sich die Differenz der beiden Devianzen berechnen, es gilt also

$$\Delta D_i^1 := D(\mathbf{y}, \hat{\boldsymbol{\beta}}) - D_{-i}(\mathbf{y}, \hat{\boldsymbol{\beta}}_{-i}^1) \approx d_i^2 + h_{ii}(r_i^{P*})^2. \tag{5.33}$$

Dies ergibt die Veränderung bzw. die Reduktion der Devianz, wenn man die i -te Beobachtung löscht. Ist dieser Wert sehr hoch, so kann man von einem ungewöhnlichen Ausmaß an Einfluß der i -ten Beobachtung ausgehen. Als kritische Schranke für die Reduktion der Devianz kann mit Hilfe der χ^2 -Verteilung ein Wert von ungefähr 4 angegeben werden, welcher auf dem 95%-Quantil der χ_1^2 -Verteilung basiert.

Um die Reduktion der Devianz an einem Datensatz zu sehen, wird wieder auf das übliche Beispiel verwiesen. Dazu wurde die Funktion *devred.poi* geschrieben, die im Anhang nachgelesen werden kann. Ebenso wie die Funktion *covratio.poi* gibt diese die geordneten Werte für die Approximation der Devianzreduktion und die zugehörigen geordneten Indizes der Beobachtungen an. In Tabelle 5.9 sind diese Werte in der ersten Komponente geordnet gegeben und in der zweiten werden die Indizes der Beobachtungen dazu aufgelistet. In Abbildung 5.9 sieht man die Werte wieder anhand einer Grafik.

Man erkennt, dass sich die vier vorher auffälligsten Beobachtungen (6, 40, 47, 66) auch hier im ersten Drittel der geordneten Werte befinden. Die größte Devianzreduktion wird durch Weglassen der Beobachtung 66 erreicht, welche die Devianz um 57.31 mindert. Wird dagegen die Beobachtung 6 weggelassen fällt die Devianz 'nur' mehr um den Wert 5.44. Wie man aus den Grafiken sehen kann, ist hier die Approximation ΔD_i^1 für den exakten Wert ΔD_i besonders gut. In Tabelle 5.10 sind die besprochenen Diagnostiken zur besseren Übersicht nochmals aufgelistet.

\$devred:

[1]	57.310946166	46.404287236	40.464433541	27.006701278	23.522771123
[6]	19.701422744	14.846946036	14.532799401	12.500074146	12.408726441
[11]	12.033050771	11.354829547	9.096835923	8.360255898	8.121671378
[16]	7.768675273	6.671027759	6.573534550	5.436876722	5.205678351
[21]	5.189600957	4.099493230	3.758228600	3.690826237	3.666203968
[26]	3.169920148	3.046101593	2.905170784	2.660306292	2.613073736
[31]	2.603187654	2.551354732	2.533746269	2.165566771	2.116425110
[36]	2.083109824	1.703416811	1.601516439	1.569716951	1.376863551
[41]	1.175263000	1.160402460	1.004604719	0.979982397	0.944984779
[46]	0.773264714	0.754393339	0.737659719	0.714461431	0.641797136
[51]	0.420668963	0.277768012	0.261161146	0.239320534	0.225797050
[56]	0.214714115	0.164160095	0.139302507	0.136681885	0.100771099
[61]	0.096406905	0.047600690	0.042217304	0.038249204	0.023228580
[66]	0.019628543	0.013362851	0.010442045	0.007376559	0.004069211

\$beob:

[1]	66	69	39	65	17	47	10	40	58	1	12	63	15	22	4	38	49	8	6	62	37	43	26	34
[25]	59	20	70	33	67	51	21	18	30	42	52	57	54	32	68	16	60	5	45	50	41	23	53	13
[49]	56	35	55	25	19	27	46	31	44	14	2	7	48	9	11	61	29	3	64	36	24	28		

Tabelle 5.9: Geordnete Werte der Devianzreduktion ΔD_i^1 des Poissonmodells für den Patentdatensatz

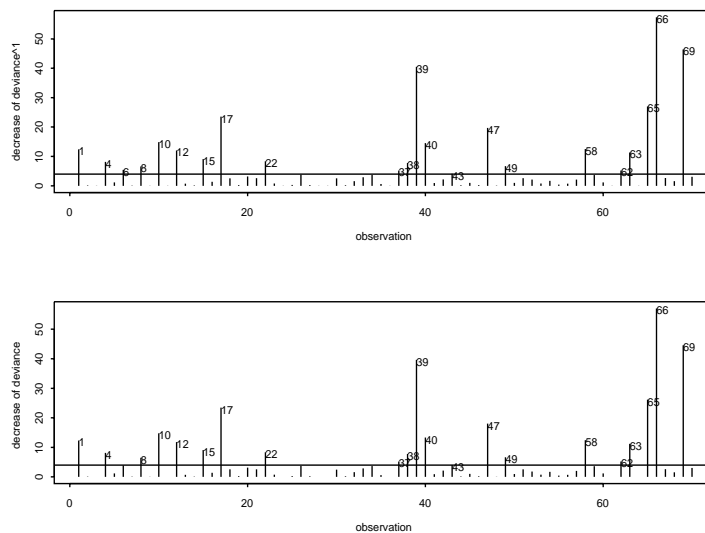


Abbildung 5.9: Stabdiagramm der Werte der Devianzreduktion ΔD_i^1 (oben) und ΔD_i (unten) des Poissonmodells für den Patentdatensatz. Die durchgezogene Linie markiert die Schranke 4.

Zusammenfassung: Regressionsdiagnostiken für GLM's				
Bezeichnung	Formel	(Nr.)	$ \cdot \leq \text{Schranke}$ $ \cdot \geq \text{Schranke}$	Zweck
h_{ii}	$w_i^{1/2} \mathbf{x}_i^T (\mathbf{X}^T \mathbf{W} \mathbf{X}) \mathbf{x}_i w_i^{1/2}$	(5.3)	$\geq 2p/n$	Hatmatrix-Diagonalelemente zur Identifizierung von x-Ausreißern
r_i^r	$y_i - \hat{\mu}_i$	(5.1)	-	Gewöhnliche Residuen zur Entdeckung von y-Ausreißern
r_i^P	$\frac{y_i - \hat{\mu}_i}{\sqrt{\hat{v}_i}}$	(5.2)	≥ 2	Pearson-Residuen zur Entdeckung von y-Ausreißern
r_i^D	d_i	(5.8)	≥ 2	Devianz-Residuen zur Entdeckung von y-Ausreißern
$DFFITs_i^1$	$\frac{g^{-1}(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}) - g^{-1}(\mathbf{x}_i^T \hat{\boldsymbol{\beta}}_{-i}^1)}{\sqrt{\hat{v}_{i,-i} h_{ii}}}$	(5.16)	≥ 2 ($\geq 2\sqrt{p/n}$)	Standardisierte Veränderung der geschätzten Werte (Fit) bei Löschung der i-ten Beobachtung
$DFBETAS_{j,i}^1$	$\frac{r_{j,i}}{\sqrt{r_j^T r_j}} \frac{r_i^P}{(1-h_{ii})}$	(5.17)	≥ 2 ($\geq 2\sqrt{1/n}$)	Standardisierte Veränderung von β_j bei Löschung der i-ten Beobachtung
D_i^1	$\frac{1}{p} \left(\frac{h_{ii}}{1-h_{ii}} \right) (r_i^{P*})^2$	(5.23)	≥ 1	Cook's Distanz zur Aufdeckung des Einflusses der i-ten Beobachtung auf die Regressionskoeffizienten
$COVRATIO_i^1$	$\frac{1}{1-h_{ii}} \frac{\sum_{j=1}^n w_j(\hat{\boldsymbol{\beta}})}{\sum_{j=1, j \neq i}^n w_j(\hat{\boldsymbol{\beta}}_{-i}^1)}$	(5.27)	-	Untersuchung des Einflusses der i-ten Beobachtung auf das Modell
ΔD_i^1	$d_i^2 + h_{ii}(r_i^{P*})^2$	(5.33)	≥ 4	Einfluß der i-ten Beobachtung auf die Devianz

Tabelle 5.10: Die wichtigsten Diagnostiken für GLM's sind hier aufgeführt. In der vierten Spalte basieren einige Schranken auf signifikanten Werten von bestimmten Verteilungen. Die Pearson- und Devianz-Residuen verhalten sich ähnlich zur Standardnormalverteilung, genauso wie die $DFFITs_i$ und die $DFBETAS_{j,i}$. Die Cook's Distanz D_i ist ähnlich zur F -Verteilung und die Devianzreduktion ΔD_i zur χ^2 -Verteilung.

5.5 Regressionsdiagnostiken bei Gewichtung einer Beobachtung

5.5.1 Gewichtung

Nun kann man genau wie bei Linearen Modellen eine Gewichtung der einzelnen Beobachtung durchführen. Es lassen sich also einzelne Beobachtungen verschieden gewichten, so dass Veränderungen in den geschätzten Werten sichtbar werden. Wie in Kapitel 3 wird definiert

$$u_i = \begin{cases} u & \text{für } i = l \\ 1 & \text{für sonst} \end{cases}$$

mit $0 \leq u \leq 1$. Der Maximum-Likelihood-Schätzer $\hat{\beta}$ von β ist nun eine Funktion in Abhängigkeit von u und wird erhalten durch Maximieren der Log-Likelihood-Funktion mit Gewichtung der l -ten Beobachtung $\mathcal{L}_u := \mathcal{L}_u(\mathbf{y}; \theta, \phi)$, also

$$\mathcal{L}_u(\mathbf{y}; \theta, \phi) = \sum_{i=1}^n u_i \mathcal{L}(y_i; \theta, \phi) = \sum_{i=1}^n u_i \left(\frac{\theta_i y_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right), \quad (5.34)$$

wobei β durch die Beziehung $\theta_i = \theta_i(\mu_i)$, $\mu_i = h(\theta_i)$ und $g(\mu_i) = \eta_i = \mathbf{x}_i^T \beta$ in die Funktion eingeht. Dies ist äquivalent zur Lösung des Problems in Kapitel 4.2. Die Scoregleichungen lauten dann

$$s_j(\beta) = \frac{\partial \mathcal{L}_u}{\partial \beta_j} = \sum_{i=1}^n u_i w_i (y_i - \mu_i) \frac{d\eta_i}{d\mu_i} x_{ij} = 0, \quad \forall j$$

und die erwartete Fisher-Informationsmatrix kann als

$$FI_u(\beta) := \left[-E \left(\frac{\partial^2 \mathcal{L}_u}{\partial \beta_r \partial \beta_s} \right) \right]_{\substack{r=1, \dots, p \\ s=1, \dots, p}} = \mathbf{X}^T \mathbf{W}^{1/2} \mathbf{U} \mathbf{W}^{1/2} \mathbf{X}$$

dargestellt werden. Die Matrix $\mathbf{U} = \text{diag}(u_1, \dots, u_n)$ stellt die Gewichtungsmatrix dar. Die Newton-Raphson-Methode erhält dann eine Folge von Schätzungen

$$\hat{\beta}_i^{t+1}(\mathbf{u}) = \hat{\beta}_i^t(\mathbf{u}) + (\mathbf{X}^T (\mathbf{W}^t)^{1/2} \mathbf{U} (\mathbf{W}^t)^{1/2} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^t (\mathbf{D}^t)^{-1} \mathbf{U} (\mathbf{y} - \hat{\mu}^t), \quad (5.35)$$

wobei

$$\begin{aligned} \mathbf{W}^t &:= \mathbf{W}(\hat{\beta}^t) = \text{diag}(w_1^t, \dots, w_n^t), \\ (\mathbf{D}^t)^{-1} &:= \mathbf{D}^{-1}(\hat{\beta}^t) = \text{diag} \left(\left. \frac{d\eta_1}{d\mu_1} \right|_{\beta=\hat{\beta}^t}, \dots, \left. \frac{d\eta_n}{d\mu_n} \right|_{\beta=\hat{\beta}^t} \right). \end{aligned}$$

Für $u = 1$ ist \mathbf{U} die Identitätsmatrix und der obige Ausdruck vereinfacht sich zum gewöhnlichen ML-Schätzer $\hat{\beta}^{t+1}$ in Gleichung (4.10). Das Ziel ist nun einzelne Einflüsse zu bestimmen. Für große Veränderungen reicht es, die Richtung zu kennen, für kleine jedoch ist Genauigkeit gefragt. Nützliche Resultate können wiederum mit dem gewöhnlichen ML-Wert $\hat{\beta}$ als Startwert erreicht werden. Dazu beendet man die Reihe in (5.35) nach dem ersten Schritt, d.h man erhält wieder die Ein-Schritt-Approximation. Somit ist

$$\begin{aligned} \hat{\beta}_i^1(\mathbf{u}) &:= \hat{\beta} + (\mathbf{X}^T \mathbf{W}^{1/2} \mathbf{U} \mathbf{W}^{1/2} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2} \mathbf{D}^{-1} \mathbf{W}^{1/2} \mathbf{U} (\mathbf{y} - \hat{\mu}) \\ &= (\mathbf{X}^T \mathbf{W}^{1/2} \mathbf{U} \mathbf{W}^{1/2} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2} \mathbf{U} \mathbf{W}^{1/2} \mathbf{Z}, \end{aligned} \quad (5.36)$$

wobei $\mathbf{Z} := \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{D}^{-1}(\mathbf{y} - \hat{\boldsymbol{\mu}})$. Diese Gleichung ist der exakten nichtiterativen Lösung für das klassische Lineare Modell mit $\mathbf{W}^{1/2}\mathbf{X}$ als Design-Matrix für die Zielvariable $\mathbf{W}^{1/2}\mathbf{Z}$ identisch. Dies ist nützlich, da man nun die in Kapitel 3 hergeleitete Formel (3.32) auf diese Ein-Schritt-Schätzung anwenden kann. Es gilt also

$$\hat{\beta}_i^1(\mathbf{u}) = \hat{\beta} - w^{1/2} \frac{(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i r_i^P (1-u)}{(1 - (1-u)h_{ii})}. \quad (5.37)$$

Für $u = 1$ ergibt sich der gewöhnliche ML-Wert $\hat{\beta}$. Wenn man dagegen u gegen Null gehen lässt und somit der i -ten Beobachtung immer weniger Gewicht gibt, kann man etwaige Veränderungen beobachten. Wenn die Veränderung in den Koeffizienten mit sinkendem u vernachlässigbar ist, so hat die i -te Beobachtung wenig Einfluß auf die Koeffizienten und daher auf den geschätzten Wert selbst. Andererseits wenn kleine Veränderungen in u große Änderungen in den Koeffizienten zur Folge haben, so ist die i -te Beobachtung einflußreich. Man kann sich dann die unstabilsten Regressionskoeffizienten und Komponenten der angepassten Werte bezüglich dieses Punktes ansehen. Eine Auswertung von $u = 0$ ergibt den in Gleichung (5.15) ausgerechneten Wert

$$\hat{\beta}_i^1(0) = \hat{\beta}_{-i}^1 = \hat{\beta} - \frac{(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i w^{1/2} r_i^P}{(1 - h_{ii})}.$$

5.5.2 Influenzkurven

Zunächst kann man sich ebenso wie in Kapitel 2 infinitesimale Änderungen ansehen. Dazu sieht man sich die erste Ableitung von $\hat{\beta}_i^1(\mathbf{u})$ bzgl. u an. Im allgemeinen ergibt sich dafür

$$\begin{aligned} \frac{\partial}{\partial u} \hat{\beta}_i^1(\mathbf{u}) &= (\hat{\beta}_i^1)'(\mathbf{u}) \\ &= - \frac{(1 - (1-u)h_{ii})(-(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i w_i^{1/2} r_i^P) - (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i w_i^{1/2} r_i^P (1-u)h_{ii}}{(1 - (1-u)h_{ii})^2} \\ &= \frac{(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i w_i^{1/2} r_i^P}{(1 - (1-u)h_{ii})^2}. \end{aligned}$$

Eine Auswertung für $u = 0, 1, \bar{u}$ ergibt verschiedene Informationen bezüglich des Einflusses der i -ten Beobachtung auf den angepassten Wert, wie man schon in Kapitel 3 gesehen hat. Man bekommt wieder drei Influenzkurven die nachfolgend angegeben werden:

$$(\hat{\beta}_i^1)'(1) = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i w_i^{1/2} r_i^P, \quad (5.38)$$

$$(\hat{\beta}_i^1)'(0) = \frac{(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i w_i^{1/2} r_i^P}{(1 - h_{ii})^2}. \quad (5.39)$$

Falls $\hat{\beta}_i^1(\mathbf{u})$ auf dem Einheitsintervall differenzierbar ist, kann man den Mittelwertsatz anwenden. Es ergibt sich die dritte Influenzkurve als

$$\frac{\hat{\beta}_i^1(1) - \hat{\beta}_i^1(0)}{1 - 0} = \frac{(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i w_i^{1/2} r_i^P}{1 - h_{ii}} = (\hat{\beta}_i^1)'(\bar{u}). \quad (5.40)$$

Ein Plot der einzelnen Komponenten von $(\hat{\beta}_i^1)'(\bar{u})$ gegen die Standardabweichung der einzelnen Komponenten von $\hat{\beta}$ deckt mögliche Instabilitäten in den ausgewählten Koeffizienten auf. Darüberhinaus lässt sich $(\hat{\beta}_i^1)'(\bar{u})$ genau wie in den LM's als Linearkombination von $(\hat{\beta}_i^1)'(1)$ und $(\hat{\beta}_i^1)'(0)$ schreiben, d.h.

$$\alpha(\hat{\beta}_i^1)'(1) + (1 - \alpha)(\hat{\beta}_i^1)'(0) = (\hat{\beta}_i^1)'(\bar{u})$$

mit $\alpha = -\frac{1}{h_{ii}-2}$ für $1/n \leq h_{ii} \leq 1$.

Um die Influenzkurven zu veranschaulichen, wird ein Beispiel mit Hilfe des Poissonmodells des Patentdatensatzes konstruiert. Es ergeben sich hierbei dreidimensionale Grafiken, da man die Abhängigkeit von der Gewichtung u und des Index i zeigt. In Abbildung 5.10 und Abbildung 5.11 sind die Influenzkurven sichtbar, einmal unskaliert und einmal mit Skalierung der z -Achse, um die Kurven genauer zu verdeutlichen. Aufgrund der Abhängigkeit des Index i erhält man eine Oberfläche. Da ein Vektor nach u abgeleitet wird, ergibt sich jeweils für jede Kovariable eine Grafik.

Aus den Grafiken und den Werten fallen die Beobachtungen 6, 40, 44 und 47 auf, deren Rate mit anwachsendem u absolut gesehen abnimmt. Falls diese Beobachtungen also jeweils weggelassen oder weniger gewichtet werden, ist die Änderungsrate der Regressionskoeffizienten groß, was auf einflussreiche Beobachtungen schließen lässt.

5.5.3 Reduktion der Devianz

Ein weiterer bereits angesprochener Punkt ist der sogenannte Goodness-of-Fit des Modells. Dieser kann leicht durch Ansehen der Devianz D nach Änderung von u untersucht werden. Diejenigen Punkte, die das Modell stark beeinflussen, können auch Veränderungen in der Qualität des Fit herbeiführen, was durch die Devianz sichtbar wird. Somit sollten vorwiegend Punkte identifiziert werden, die einen starken Einfluß auf den Fit haben. Wie bereits bekannt, können Veränderungen in den Statistiken zwei mögliche Formen haben. Falls erstens der i -te Punkt nicht gut vom Modell angepasst wird, sind die Änderungen in D gewöhnlich in den einzelnen Komponenten d_i sichtbar. Zweitens, falls der i -te Punkt ein Ausreißer ist, resultieren die Veränderungen in D aus allen einzelnen geänderten Komponenten d_i . Welcher der beiden Fälle tatsächlich eingetreten ist, muss mit Hilfe der anderen Diagnostiken ermittelt werden. Die Elemente der Hatmatrix können hier verwendet werden.

Zuerst wird die Devianz für beliebige Werte von u behandelt. Benutzt man (5.34), dann bekommt man für die entsprechende Devianz

$$D_u(\mathbf{y}, \hat{\beta}_i(\mathbf{u})) = 2 \sum_{j=1}^n u_j [\mathcal{L}(y_j, y_j) - \mathcal{L}(y_j, \hat{\mu}_j)].$$

Dies ist die Devianz mit Gewichtung der i -ten Beobachtung. Wenn man dies für die Ein-Schritt-Schätzung $\hat{\beta}_i^1(\mathbf{u})$ auswertet, wird diese Devianz approximiert. Insbesondere bewirkt eine Entwicklung zweiter Ordnung von $D_u(\mathbf{y}, \hat{\beta}_i^1(\mathbf{u}))$ um $\hat{\beta}$, wie sie bereits für den Spezialfall $u = 0$ gemacht wurde, folgendes:

$$\begin{aligned} D_u(\mathbf{y}, \hat{\beta}_i^1(\mathbf{u})) &\approx D_u(\mathbf{y}, \hat{\beta}) + (\hat{\beta} - \hat{\beta}_i^1(\mathbf{u}))^T \left. \frac{\partial D_u(\mathbf{y}, \beta)}{\partial \beta} \right|_{\beta=\hat{\beta}} \\ &\quad + \frac{1}{2} (\hat{\beta} - \hat{\beta}_i^1(\mathbf{u}))^T \left. \frac{\partial^2 D_u(\mathbf{y}, \beta)}{\partial \beta \partial \beta^T} \right|_{\beta=\hat{\beta}} (\hat{\beta} - \hat{\beta}_i^1(\mathbf{u})). \end{aligned} \quad (5.41)$$

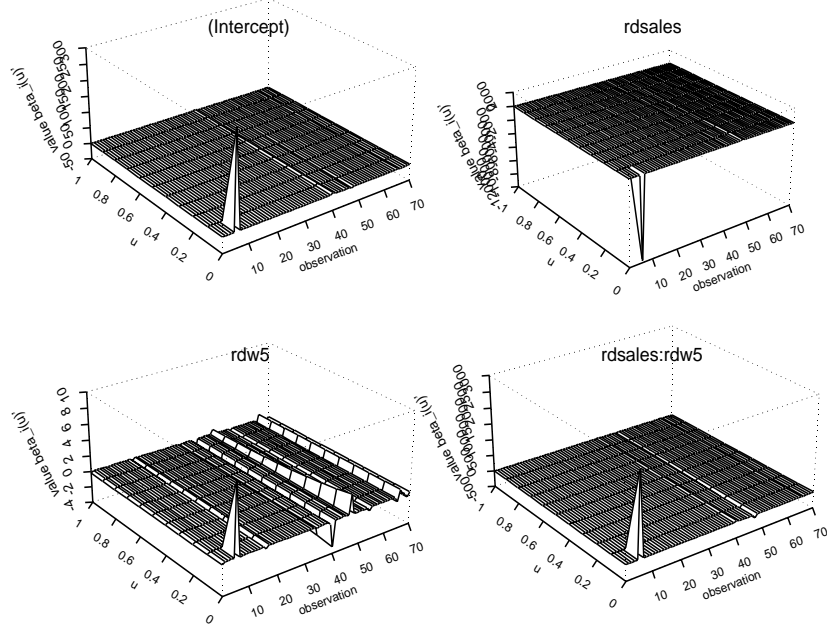


Abbildung 5.10: Influenzkurven des Poissonmodells für den Patentdatensatz

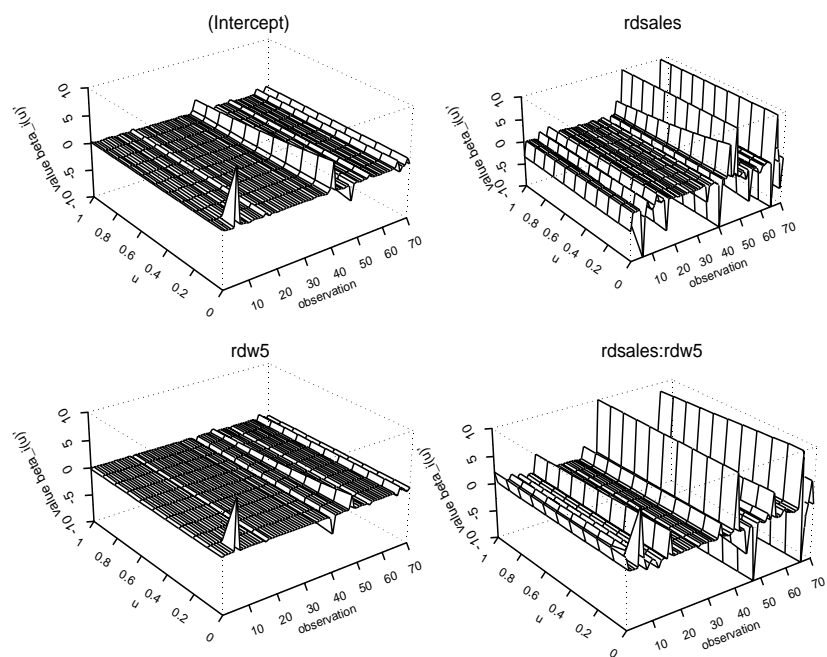


Abbildung 5.11: Influenzkurven des Poissonmodells für den Patentdatensatz mit Skalierung ($|\cdot| \leq 10$) der z-Achse

Es gilt:

$$D_u(\mathbf{y}, \hat{\boldsymbol{\beta}}) = D(\mathbf{y}, \hat{\boldsymbol{\beta}}) - (1-u)d_i^2.$$

Nun kann man analog zum Spezialfall von $u = 0$ im vorigen Kapitel vorgehen. Man setzt den letzten Ausdruck und die gewichtete Ein-Schritt-Approximation (5.37) in (5.41) ein und rechnet aus. Es ergibt sich daraus

$$\begin{aligned} D_u^1 &:= D_u^1(\mathbf{y}, \hat{\boldsymbol{\beta}}_i^1(\mathbf{u})) \approx D(\mathbf{y}, \hat{\boldsymbol{\beta}}) - (1-u)d_i^2 - \\ &\quad - \left(w_i^{1/2} \frac{(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i r_i^P (1-u)}{(1-(1-u)h_{ii})} \right)^T (\mathbf{X}^T \mathbf{W} \mathbf{X}) \left(w_i^{1/2} \frac{(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i r_i^P (1-u)}{(1-(1-u)h_{ii})} \right) + \\ &\quad + \left(w_i^{1/2} \frac{(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i r_i^P (1-u)}{(1-(1-u)h_{ii})} \right)^T ((1-u)\mathbf{x}_i w_i \mathbf{x}_i^T) \left(w_i^{1/2} \frac{(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{x}_i r_i^P (1-u)}{(1-(1-u)h_{ii})} \right) \\ &= D(\mathbf{y}, \hat{\boldsymbol{\beta}}) - (1-u)d_i^2 - \left(\frac{(1-u)^2 h_{ii} (r_i^P)^2}{(1-(1-u)h_{ii})^2} - \frac{(1-u)^3 h_{ii}^2 (r_i^P)^2}{(1-(1-u)h_{ii})^2} \right) \\ &= D(\mathbf{y}, \hat{\boldsymbol{\beta}}) - (1-u)d_i^2 - \frac{(1-u)^2 h_{ii} (r_i^P)^2 (1-(1-u)h_{ii})}{(1-(1-u)h_{ii})^2} \\ &= D(\mathbf{y}, \hat{\boldsymbol{\beta}}) - \left[(1-u)d_i^2 + \frac{(1-u)^2 h_{ii} (r_i^P)^2}{(1-(1-u)h_{ii})} \right]. \end{aligned} \quad (5.42)$$

Die gewichtete Devianz erreicht bei $u = 0$ ihr Minimum, welches durch

$$D(\mathbf{y}, \hat{\boldsymbol{\beta}}) - (d_i^2 + h_{ii}(r_i^{P*})^2)$$

gegeben ist und bei $u = 1$ ihr Maximum, welches bei

$$D(\mathbf{y}, \hat{\boldsymbol{\beta}})$$

liegt. Darum reduziert ein Untergewichten der i -ten Beobachtung die Devianz.

Die Änderungsrate der Devianz bzgl. Änderungen in u für den i -ten Punkt wird einfach durch Differenzieren von (5.42) erhalten:

$$\begin{aligned} \frac{\partial}{\partial u} D_u(\mathbf{y}, \hat{\boldsymbol{\beta}}^1(\mathbf{u})) &= -(-d_i^2) - \frac{(1-(1-u)h_{ii})(r_i^P)^2 h_{ii} 2(1-u)(-1) - (r_i^P)^2 (1-u)^2 h_{ii} h_{ii}}{(1-(1-u)h_{ii})^2} \\ &= d_i^2 + \frac{(r_i^P)^2 [2(1-u)h_{ii} - 2(1-u)^2 h_{ii}^2 + (1-u)^2 h_{ii}^2]}{(1-(1-u)h_{ii})^2} \\ &= d_i^2 + \frac{(r_i^P)^2 (1-u)h_{ii}(2-(1-u)h_{ii})}{(1-(1-u)h_{ii})^2}. \end{aligned}$$

Für verschiedene Auswertungen an der Stelle u ergeben sich die zugehörigen Änderungsraten der Devianz. Der Wert für die Differentiation der gewichteten Devianz gibt Auskunft über die Änderung der Devianz. Falls ein hoher Wert angenommen wird, bestätigt dies ein ungewöhnliches Ausmaß des i -ten Punkts auf die Devianz.

Eine Auswertung der gewichteten Devianz für die Ein-Schritt-Approximation ohne i -te Beobachtung ergibt

$$\Delta_i D^1 := D_1^1(\mathbf{y}, \hat{\boldsymbol{\beta}}) - D_0^1(\mathbf{y}, \hat{\boldsymbol{\beta}}^1(0)) = d_i^2 + \frac{(r_i^P)^2 h_{ii}}{1-h_{ii}}.$$

Um diesen Beitrag nimmt die Devianz ab, falls man die i -te Beobachtung löscht. Alle wichtigen Diagnostiken spielen hier eine Rolle, d.h. der Ausdruck ist von den d_i , r_i^P und h_{ii} abhängig. Falls

1. $u = 0$ mit der Löschung der i -ten Beobachtung einhergeht,
2. die Löschung der i -ten Beobachtung die Design-Matrix \mathbf{X} um eine Dummyvariable vergrößert und
3. die Devianzreduzierung durch Addieren einer Variablen asymptotisch χ_1^2 -verteilt ist,

dann sind die einzelnen Werte $\Delta_i D^1$ asymptotisch χ_1^2 -verteilt. Insbesondere führt ein Ersetzen von r_i^P durch d_i in der Definition von $\Delta_i D^1$ zu

$$\Delta_i D^1 \approx \frac{d_i^2}{1 - h_{ii}}$$

und zum approximativen standardisierten Residuum $r_i^{D*} = \frac{d_i}{\sqrt{1-h_{ii}}}$. In Pregibon (1981) wird die Verwendung der geordneten Werte von $\frac{d_i}{\sqrt{1-h_{ii}}}$ statt r_i^P als Ordinaten in Wahrscheinlichkeitsplots vorgeschlagen.

Kapitel 6

Negativ-Binomial-Regression mit geschätztem α

In diesem Kapitel sollen nun Regressionsdiagnostiken für den Spezialfall der Negativ-Binomial-Verteilung mit geschätztem α entwickelt werden. Die Negativ-Binomial-Verteilung ist im Anhang B definiert. Desweiteren werden dort wichtige Eigenschaften angegeben.

Hierbei tritt ein entscheidendes Problem auf. Wenn der Parameter α unbekannt ist, ist die Negativ-Binomial-Verteilung nicht mehr in der Klasse der Exponentiellen Familie wie in Kapitel 3 beschrieben. Die Negativ-Binomial-Regression ist somit kein GLM mehr. Im folgenden sollen nun geeignete Schätzer für β und α gefunden werden. Dabei stützt man sich zuerst auf einige feste Werte α , für die man den Log-Likelihood $\mathcal{L}(\beta, \alpha)$ bzgl. β maximiert. Dieses Vorgehen wird als *Profiling* bezeichnet. Man erhält also Schätzungen $\hat{\beta}(\alpha)$ und den *Profile-Log-Likelihood* $\mathcal{L}(\hat{\beta}(\alpha), \alpha)$ in Abhängigkeit von α , von dem man schließlich einen Schätzer $\hat{\alpha}$ durch eindimensionales Maximieren bestimmen kann.

6.1 Profiling

Das Ziel ist es nun, den gemeinsamen ML-Schätzer $(\hat{\beta}, \hat{\alpha})$ für (β, α) zu berechnen. Die einfachste Methode, die Schätzer $(\hat{\beta}, \hat{\alpha})$ zu erhalten, ist, $\mathcal{L}(\beta, \alpha)$ bezüglich β für bestimmte, feste Werte α zu maximieren. Es ergibt sich die Schätzung $\hat{\beta}(\alpha)$ und der sogenannte Profile-Log-Likelihood $\mathcal{L}(\hat{\beta}(\alpha), \alpha)$ in Abhängigkeit von α . Schließlich erhält man einen Schätzer $\hat{\alpha}$ durch eindimensionales Maximieren des Profile-Log-Likelihood. Die Maximierung von $\mathcal{L}(\beta, \alpha)$ bezüglich β kann mittels des in Anhang D vorgestellten Newton-Raphson-Verfahrens erfolgen. Um die angeführten Schätzungen durchzuführen, ist die Fisher-Informationsmatrix zu berechnen. Dies soll im folgenden getan werden.

Es wird für die weiteren Berechnungen wieder die Log-Link-Funktion herangezogen, so dass $Y_i \sim NB(\mu_i, \alpha)$, $i = 1, \dots, n$, unabhängig mit $\mu_i = \exp(\mathbf{x}_i^T \beta)$. Für andere Linkspezifikationen ergeben sich analoge Ergebnisse. Die Likelihood-Funktion ist gegeben durch

$$l(\beta, \alpha) = \prod_{i=1}^n \left[\left(\frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y_i + 1)} \right) \left(\frac{1}{1 + \alpha\mu_i} \right)^{\alpha^{-1}} \left(\frac{\alpha\mu_i}{1 + \alpha\mu_i} \right)^{y_i} \right]. \quad (6.1)$$

Da man im folgenden nach α ableiten soll, wird der Ausdruck $\frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1})}$ in einer anderen Darstellung geschrieben. Da $\Gamma(x + 1) = x\Gamma(x) = x!$ für $x > 0$, gilt

$$\begin{aligned} \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1})} &= \frac{(y_i + \alpha^{-1} - 1)\Gamma(y_i + \alpha^{-1} - 1)}{(\alpha^{-1} - 1)\Gamma(\alpha^{-1} - 1)} \\ &= (y_i + \alpha^{-1} - 1) \cdot \dots \cdot (y_i + \alpha^{-1} - y_i) = \prod_{j=0}^{y_i-1} (j + \alpha^{-1}), \end{aligned}$$

wobei $\prod_{j=0}^{y_i-1}$ Null ist, falls $y_i < 0$. Somit kann die Likelihood-Funktion umgeschrieben werden in

$$\begin{aligned} l(\boldsymbol{\beta}, \alpha) &= \prod_{i=1}^n \left[\left(\prod_{j=0}^{y_i-1} (j + \alpha^{-1}) \right) \frac{1}{y_i!} \left(\frac{1}{1 + \alpha\mu_i} \right)^{\alpha^{-1}} \left(\frac{\alpha\mu_i}{1 + \alpha\mu_i} \right)^{y_i} \right] \\ &= \prod_{i=1}^n \left[\left(\prod_{j=0}^{y_i-1} (\alpha j + 1) \right) \frac{1}{y_i!} \left(\frac{1}{1 + \alpha\mu_i} \right)^{\alpha^{-1}} \left(\frac{\mu_i}{1 + \alpha\mu_i} \right)^{y_i} \right]. \end{aligned} \quad (6.2)$$

Durch Logarithmieren erhält man die Log-Likelihood-Funktion

$$\begin{aligned} \mathcal{L}(\boldsymbol{\beta}, \alpha) &= \sum_{i=1}^n \left[\left(\sum_{j=0}^{y_i-1} \log(\alpha j + 1) \right) - \log(y_i!) + \alpha^{-1} \log \left(\frac{1}{1 + \alpha\mu_i} \right) + y_i \log \left(\frac{\mu_i}{1 + \alpha\mu_i} \right) \right] \\ &= \sum_{i=1}^n \left[\left(\sum_{j=0}^{y_i-1} \log(\alpha j + 1) \right) - \log(y_i!) - (y_i + \alpha^{-1}) \log(1 + \alpha\mu_i) + y_i \log(\mu_i) \right]. \end{aligned} \quad (6.3)$$

Zunächst bildet man die ersten Ableitungen von $\mathcal{L} := \mathcal{L}(\boldsymbol{\beta}, \alpha)$ nach $\boldsymbol{\beta}$ und α . Die erste Ableitung von \mathcal{L} nach $\boldsymbol{\beta}$ ist aufgrund von $\mu_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$ und nach Beispiel 4.6

$$\frac{\partial \mathcal{L}}{\partial \beta_j} = \sum_{i=1}^n \frac{1}{1 + \alpha\mu_i} (y_i - \mu_i) x_{ij}.$$

Die Ableitung nach α ist

$$\frac{\partial \mathcal{L}}{\partial \alpha} = \sum_{i=1}^n \left[\left(\sum_{j=0}^{y_i-1} \left(\frac{j}{\alpha j + 1} \right) \right) + \frac{1}{\alpha^2} \log(1 + \alpha\mu_i) - (y_i + \alpha^{-1}) \frac{\mu_i}{1 + \alpha\mu_i} \right].$$

Um die Fisher-Informationsmatrix zu bekommen, sind außerdem die zweiten Ableitungen nötig:

$$\frac{\partial^2 \mathcal{L}}{\partial \beta_r \partial \beta_s} = \sum_{i=1}^n \frac{-x_{ir} x_{is} \mu_i (\alpha y_i + 1)}{(1 + \alpha\mu_i)^2} \quad (6.4)$$

$$\frac{\partial^2 \mathcal{L}}{\partial \beta_r \partial \alpha} = \sum_{i=1}^n \frac{-x_{ir} \mu_i (y_i - \mu_i)}{(1 + \alpha\mu_i)^2} \quad (6.5)$$

$$\begin{aligned}
\frac{\partial^2 \mathcal{L}}{\partial \alpha^2} &= \sum_{i=1}^n \left[\left(\sum_{j=0}^{y_i-1} \left(\frac{-j^2}{(\alpha j + 1)^2} \right) \right) - 2 \frac{1}{\alpha^3} \log(1 + \alpha \mu_i) + 2 \frac{1}{\alpha^2} \frac{\mu_i}{1 + \alpha \mu_i} + \frac{(y_i + \alpha^{-1}) \mu_i^2}{(1 + \alpha \mu_i)^2} \right] \\
&= - \sum_{i=1}^n \left[\left(\sum_{j=0}^{y_i-1} \left(\frac{j}{\alpha j + 1} \right)^2 \right) + 2 \frac{1}{\alpha^3} \log(1 + \alpha \mu_i) - 2 \frac{1}{\alpha^2} \frac{\mu_i}{1 + \alpha \mu_i} - \frac{(y_i + \alpha^{-1}) \mu_i^2}{(1 + \alpha \mu_i)^2} \right] \quad (6.6)
\end{aligned}$$

Das Ziel ist nun, aus diesen Ableitungen die Fisher-Informationsmatrix FI herzuleiten. Somit ist die Fisher-Informationsmatrix eine $(p+1) \times (p+1)$ -Matrix, deren erster Block aus den Erwartungswerten der negativen zweiten Ableitungen nach β besteht. In der letzten Spalte und der letzten Zeile sind die gemischten Ableitungen eingetragen, die wegen der Symmetrie der Fisher-Informationsmatrix gleich sind. Das Element an der Stelle $(p+1, p+1)$ der Matrix ist der Erwartungswert aus der negativen zweiten Ableitung nach α . Man erhält aus (6.4)

$$\begin{aligned}
FI_{r,s}(\beta, \alpha) &= E \left[- \sum_{i=1}^n \frac{-x_{ir} x_{is} \mu_i (\alpha y_i + 1)}{(1 + \alpha \mu_i)^2} \right] \\
&= \sum_{i=1}^n \frac{x_{ir} x_{is} \mu_i (1 + \alpha \overbrace{E(y_i)}^{=\mu_i})}{(1 + \alpha \mu_i)^2} \\
&= \sum_{i=1}^n \frac{x_{ir} x_{is} \mu_i}{1 + \alpha \mu_i}, \quad r, s = 1, \dots, p. \quad (6.7)
\end{aligned}$$

Die Einträge der letzten Spalte bzw. der letzten Zeile sind nach (6.5)

$$\begin{aligned}
FI_{r,p+1}(\beta, \alpha) &= E \left[- \sum_{i=1}^n \frac{-x_{ir} \mu_i (y_i - \mu_i)}{(1 + \alpha \mu_i)^2} \right] \\
&= \sum_{i=1}^n \frac{x_{ir} \mu_i (E(y_i) + \mu_i)}{(1 + \alpha \mu_i)^2} = \mathbf{0}, \quad r = 1, \dots, p. \quad (6.8)
\end{aligned}$$

Nun ist noch der $(p+1, p+1)$ -Eintrag zu bestimmen. Da der Erwartungswert einer Summe, welche von y abhängt, zu bestimmen ist, stellt sich dieser Rechenschritt etwas schwieriger dar. In Lawless (1987) wird dazu Stellung genommen. Der Ausdruck für das Element an der Stelle $(p+1, p+1)$ wird durch Umschreiben des Log-Likelihoods mit Variablen β und $k = \alpha^{-1}$ erhalten. Man berechnet sodann die zweiten Ableitungen nach k , also $\frac{\partial^2 \mathcal{L}}{\partial k^2}$, und benutzt folgendes Lemma.

Lemma 6.1 Sei $k = \frac{1}{\alpha}$. Dann gilt die Gleichheit

$$E \left[- \frac{\partial^2 \mathcal{L}(\beta, \alpha)}{\partial \alpha^2} \right] = \alpha^{-4} E \left[- \frac{\partial^2 \mathcal{L}(\beta, k)}{\partial k^2} \right].$$

Beweis: Da $\mathcal{L} := \mathcal{L}(\beta, \alpha) = \log(l(\beta, \alpha))$ gilt, lässt sich der linke Term ohne Erwartungswert wie folgt berechnen:

$$\frac{\partial^2 \mathcal{L}}{\partial \alpha^2} = \frac{\partial^2}{\partial \alpha^2} (\log(l(\beta, \alpha))) = \frac{\partial}{\partial \alpha} \left(\frac{l'(\beta, \alpha)}{l(\beta, \alpha)} \right) = \frac{l''(\beta, \alpha) l(\beta, \alpha) - (l'(\beta, \alpha))^2}{(l(\beta, \alpha))^2},$$

wobei l' und l'' jeweils die erste und zweite Ableitung nach α bezeichnen. Nach Umparameterisierung der Likelihood-Funktion durch $k = g(\alpha) = 1/\alpha$, lässt sie sich als eine Funktion $l(\boldsymbol{\beta}, k) = l(\boldsymbol{\beta}, g(\alpha))$ auffassen, für die allgemein die folgenden Formeln gelten:

$$\frac{\partial}{\partial \alpha} l(\boldsymbol{\beta}, k) = \frac{\partial}{\partial \alpha} l(\boldsymbol{\beta}, g(\alpha)) = \frac{\partial}{\partial g(\alpha)} l(\boldsymbol{\beta}, g(\alpha)) \frac{\partial}{\partial \alpha} g(\alpha) = l'(\boldsymbol{\beta}, g(\alpha)) g'(\alpha), \quad (6.9)$$

$$\begin{aligned} \frac{\partial^2}{\partial \alpha^2} l(\boldsymbol{\beta}, k) &= \frac{\partial^2}{\partial \alpha^2} l(\boldsymbol{\beta}, g(\alpha)) \\ &= \frac{\partial}{\partial \alpha} (l'(\boldsymbol{\beta}, g(\alpha)) g'(\alpha)) \\ &= l''(\boldsymbol{\beta}, g(\alpha)) (g'(\alpha))^2 + l'(\boldsymbol{\beta}, g(\alpha)) g''(\alpha). \end{aligned} \quad (6.10)$$

Die Ableitungen von $g(\alpha)$ sind

$$g'(\alpha) = -\frac{1}{\alpha^2} \quad \text{und} \quad g''(\alpha) = \frac{2}{\alpha^3}. \quad (6.11)$$

Mit dieser Umparameterisierung berechnet sich die linke Seite der Behauptung ohne Erwartungswert wie folgt:

$$\begin{aligned} \frac{\partial^2 \mathcal{L}(\boldsymbol{\beta}, k)}{\partial \alpha^2} &= \frac{\partial^2 \mathcal{L}(\boldsymbol{\beta}, g(\alpha))}{\partial \alpha^2} = \frac{\partial^2}{\partial \alpha^2} (\log(l(\boldsymbol{\beta}, g(\alpha)))) = \frac{\partial}{\partial \alpha} \left(\frac{\frac{\partial}{\partial \alpha} l(\boldsymbol{\beta}, g(\alpha))}{l(\boldsymbol{\beta}, g(\alpha))} \right) \\ &\stackrel{(6.9)}{=} \frac{\partial}{\partial \alpha} \left(\frac{l'(\boldsymbol{\beta}, g(\alpha)) g'(\alpha)}{l(\boldsymbol{\beta}, g(\alpha))} \right) \\ &\stackrel{(6.10)}{=} \frac{(l''(\boldsymbol{\beta}, g(\alpha)) (g'(\alpha))^2 + l'(\boldsymbol{\beta}, g(\alpha)) g''(\alpha)) l(\boldsymbol{\beta}, g(\alpha)) - (l'(\boldsymbol{\beta}, g(\alpha)) g'(\alpha))^2}{(l(\boldsymbol{\beta}, g(\alpha)))^2} \\ &\stackrel{(6.11)}{=} \frac{1}{\alpha^4} \frac{(l''(\boldsymbol{\beta}, g(\alpha)) + 2\alpha l'(\boldsymbol{\beta}, g(\alpha))) l(\boldsymbol{\beta}, g(\alpha)) - (l'(\boldsymbol{\beta}, g(\alpha)))^2}{(l(\boldsymbol{\beta}, g(\alpha)))^2}. \end{aligned} \quad (6.12)$$

Nimmt man nun den Erwartungswert und spaltet den Term auf, so ergibt sich

$$\frac{1}{\alpha^4} E \left[\frac{l''(\boldsymbol{\beta}, g(\alpha)) l(\boldsymbol{\beta}, g(\alpha)) - (l'(\boldsymbol{\beta}, g(\alpha)))^2}{(l(\boldsymbol{\beta}, g(\alpha)))^2} \right] + \frac{1}{\alpha^4} E \left[\frac{2\alpha l'(\boldsymbol{\beta}, g(\alpha)) l(\boldsymbol{\beta}, g(\alpha))}{(l(\boldsymbol{\beta}, g(\alpha)))^2} \right]$$

Für den zweiten Term gilt hier

$$\begin{aligned} \frac{1}{\alpha^4} E \left[\frac{2\alpha l'(\boldsymbol{\beta}, g(\alpha)) l(\boldsymbol{\beta}, g(\alpha))}{(l(\boldsymbol{\beta}, g(\alpha)))^2} \right] &= \frac{2}{\alpha^3} E \left[\frac{l'(\boldsymbol{\beta}, g(\alpha))}{l(\boldsymbol{\beta}, g(\alpha))} \right] \\ &= \frac{2}{\alpha^3} \int \frac{l'(\boldsymbol{\beta}, g(\alpha))}{l(\boldsymbol{\beta}, g(\alpha))} l(\boldsymbol{\beta}, g(\alpha)) dx \\ &= \frac{2}{\alpha^3} \int l'(\boldsymbol{\beta}, g(\alpha)) dx \\ &= \frac{2}{\alpha^3} \frac{\partial}{\partial g(\alpha)} \underbrace{\int l(\boldsymbol{\beta}, g(\alpha)) dx}_{=1, \text{ da } l \text{ eine Dichte.}} = 0, \end{aligned} \quad (6.13)$$

wobei im vorletzten Schritt die Vertauschung von Integration und Differentiation unter Regularitätsbedingungen durchgeführt werden darf. Zusammenfassend ergibt sich also

$$\begin{aligned}
E \left[-\frac{\partial^2 \mathcal{L}(\boldsymbol{\beta}, \alpha)}{\partial \alpha^2} \right] &= E \left[\frac{l''(\boldsymbol{\beta}, \alpha)l(\boldsymbol{\beta}, \alpha) - (l(\boldsymbol{\beta}, \alpha))^2}{(l(\boldsymbol{\beta}, \alpha))^2} \right] \\
&\stackrel{(6.12)}{=} \frac{1}{\alpha^4} E \left[-\frac{(l''(\boldsymbol{\beta}, g(\alpha)) + 2\alpha l'(\boldsymbol{\beta}, g(\alpha)))l(\boldsymbol{\beta}, g(\alpha)) - (l(\boldsymbol{\beta}, g(\alpha)))^2}{(l(\boldsymbol{\beta}, g(\alpha)))^2} \right] \\
&\stackrel{(6.13)}{=} \frac{1}{\alpha^4} E \left[-\frac{l''(\boldsymbol{\beta}, g(\alpha))l(\boldsymbol{\beta}, g(\alpha)) - (l(\boldsymbol{\beta}, g(\alpha)))^2}{(l(\boldsymbol{\beta}, g(\alpha)))^2} \right] \\
&= \frac{1}{\alpha^4} E \left[-\frac{\partial^2 \mathcal{L}(\boldsymbol{\beta}, g(\alpha))}{\partial (g(\alpha))^2} \right] \\
&= \alpha^{-4} E \left[-\frac{\partial^2 \mathcal{L}(\boldsymbol{\beta}, k)}{\partial k^2} \right] \tag{6.14}
\end{aligned}$$

□

Die Likelihood-Funktion ergibt sich durch die Umparameterisierung als

$$\begin{aligned}
l(\boldsymbol{\beta}, k) &= \prod_{i=1}^n \left[\left(\frac{\Gamma(y_i + k)}{\Gamma(k)\Gamma(y_i + 1)} \right) \left(\frac{1}{1 + \frac{\mu_i}{k}} \right)^k \left(\frac{\frac{\mu_i}{k}}{1 + \frac{\mu_i}{k}} \right)^{y_i} \right] \\
&= \prod_{i=1}^n \left[\left(\prod_{j=0}^{y_i-1} (j + k) \right) \frac{1}{y_i!} \left(\frac{1}{1 + \frac{\mu_i}{k}} \right)^k \left(\frac{\mu_i}{k + \mu_i} \right)^{y_i} \right] \tag{6.15}
\end{aligned}$$

Die Log-Likelihood-Funktion $\mathcal{L}(\boldsymbol{\beta}, k)$ lautet dann

$$\mathcal{L}(\boldsymbol{\beta}, k) = \sum_{i=1}^n \left[\left(\sum_{j=0}^{y_i-1} \log(j + k) \right) - \log(y_i!) - y_i \log(k + \mu_i) - k \log \left(1 + \frac{\mu_i}{k} \right) + y_i \log(\mu_i) \right].$$

Durch Bildung der ersten und zweiten Ableitung nach k erhält man

$$\begin{aligned}
\frac{\partial \mathcal{L}(\boldsymbol{\beta}, k)}{\partial k} &= \sum_{i=1}^n \left[\left(\sum_{j=0}^{y_i-1} \frac{1}{j + k} \right) - y_i \frac{1}{k + \mu_i} - \log \left(1 + \frac{\mu_i}{k} \right) - k \frac{1}{1 + \frac{\mu_i}{k}} \left(-\frac{\mu_i}{k^2} \right) \right] \\
&= \sum_{i=1}^n \left[\left(\sum_{j=0}^{y_i-1} \frac{1}{j + k} \right) - \frac{y_i}{k + \mu_i} - \log \left(1 + \frac{\mu_i}{k} \right) + \frac{\mu_i}{k + \mu_i} \right], \\
\frac{\partial^2 \mathcal{L}(\boldsymbol{\beta}, k)}{\partial k^2} &= \sum_{i=1}^n \left[\left(\sum_{j=0}^{y_i-1} \frac{-1}{(j + k)^2} \right) + \frac{y_i}{(k + \mu_i)^2} - \frac{1}{1 + \frac{\mu_i}{k}} \left(-\frac{\mu_i}{k^2} \right) + \frac{-\mu_i}{(k + \mu_i)^2} \right] \\
&= -\sum_{i=1}^n \left[\left(\sum_{j=0}^{y_i-1} \frac{1}{(j + k)^2} \right) - \frac{y_i}{(k + \mu_i)^2} - \frac{\mu_i}{(k + \mu_i)k} + \frac{\mu_i}{(k + \mu_i)^2} \right]. \tag{6.16}
\end{aligned}$$

Mittels (6.16) und des Lemmas 6.1 ermittelt man nun das Element der Fisher-Informationsmatrix an der Stelle $(p+1, p+1)$. Es ergibt sich

$$\begin{aligned}
FI_{p+1,p+1}(\beta, \alpha) &= E \left[-\frac{\partial^2 \mathcal{L}}{\partial \alpha^2} \right] = \alpha^{-4} E \left[-\frac{\partial^2 \mathcal{L}(\beta, k)}{\partial k^2} \right] \\
&= \alpha^{-4} E \left[\sum_{i=1}^n \left[\left(\sum_{j=0}^{Y_i-1} \frac{1}{(j+k)^2} \right) - \frac{y_i}{(k+\mu_i)^2} - \frac{\mu_i}{(k+\mu_i)k} + \frac{\mu_i}{(k+\mu_i)^2} \right] \right] \\
&= \alpha^{-4} \sum_{i=1}^n \left[E \left(\sum_{j=0}^{Y_i-1} \frac{1}{(j+k)^2} \right) - \frac{\overbrace{E(y_i)}^{\mu_i}}{(k+\mu_i)^2} - \frac{\mu_i}{(k+\mu_i)k} + \frac{\mu_i}{(k+\mu_i)^2} \right] \\
&= \alpha^{-4} \sum_{i=1}^n \left[E \left(\sum_{j=0}^{Y_i-1} \frac{1}{(j+\alpha^{-1})^2} \right) - \frac{\mu_i}{(\alpha^{-1}+\mu_i)\alpha^{-1}} \right]. \tag{6.17}
\end{aligned}$$

Der i -te Term in Gleichung (6.17) lässt sich nach Lawless (1987) durch

$$\alpha^{-4} \left(\sum_{j=0}^{\infty} \frac{1}{(j+\alpha^{-1})^2} P_{NB}(Y_i \geq j) - \frac{\mu_i}{(\alpha^{-1}+\mu_i)\alpha^{-1}} \right)$$

berechnen. Um dies zu zeigen, ist speziell die Gleichheit von

$$E \left(\sum_{j=0}^{Y_i-1} \frac{1}{(j+\alpha^{-1})^2} \right) = \sum_{j=0}^{\infty} \frac{1}{(j+\alpha^{-1})^2} P_{NB}(Y_i \geq j)$$

zu zeigen. Nach Resnick (1992) (Seite 2) gilt allgemein für eine nichtnegative ganzzahlige Zufallsvariable Y und für die Funktion $f: \mathbb{R}_0^+ \mapsto \mathbb{R}^+$ die Formel

$$E(f(Y)) = \sum_{y=0}^{\infty} f(y) p_y,$$

wobei $p_y = P(Y = y)$. Damit gilt mit $y^* = y - 1$

$$\begin{aligned}
E(f(Y)) &= E \left(\sum_{j=0}^{Y_i-1} \frac{1}{(j+\alpha^{-1})^2} \right) \\
&= \sum_{y^*=0}^{\infty} \sum_{j=0}^{y^*} \frac{1}{(j+\alpha^{-1})^2} P_{NB}(Y = y^*) \\
&= \frac{1}{(0+\alpha^{-1})^2} P_{NB}(Y \geq 0) + \frac{1}{(1+\alpha^{-1})^2} P_{NB}(Y \geq 1) + \dots \\
&= \sum_{j=0}^{\infty} \frac{1}{(j+\alpha^{-1})^2} P_{NB}(Y_i \geq j)
\end{aligned} \tag{6.18}$$

Somit ist die Gleichheit gezeigt und die Behauptung folgt.

Nun sind alle Einträge berechnet und die Fisher-Informationsmatrix kann in Blockform angegeben werden. Sie ergibt sich zu

$$FI(\boldsymbol{\beta}, \alpha) = \begin{bmatrix} FI_{r,s}(\boldsymbol{\beta}, \alpha) & \mathbf{0} \\ \mathbf{0}^T & FI_{p+1,p+1}(\boldsymbol{\beta}, \alpha) \end{bmatrix}, \quad r, s = 1, \dots, p.$$

Die Matrix $FI_{r,s}(\boldsymbol{\beta}, \alpha)$ ist eine Blockmatrix, $FI_{p+1,p+1}(\boldsymbol{\beta}, \alpha)$ ein skalarer Wert und $\mathbf{0}$ beschreibt den Nullvektor der Länge p .

Die Asymptotik der Schätzer wird darüberhinaus in Lawless (1987) behandelt. Dieser gibt an, dass für $\alpha > 0$ und unter schwachen Bedingungen an die Kovariablenvektoren \mathbf{x}_i , die die Approximation von $n^{-1}FI(\boldsymbol{\beta}, \alpha)$ gegen einen positiven Grenzwert für $n \rightarrow \infty$ sichern, die ML-Schätzer $(\hat{\boldsymbol{\beta}}, \hat{\alpha})$ als asymptotisch normalverteilt behandelt werden können, d.h. man kann für große n zeigen, dass

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}, \hat{\alpha} - \alpha) \sim N_p(\mathbf{0}, nFI^{-1}(\hat{\boldsymbol{\beta}}, \hat{\alpha}))$$

gilt, wobei die Kovarianzmatrix durch

$$nFI^{-1}(\hat{\boldsymbol{\beta}}, \hat{\alpha}) = \begin{bmatrix} FI_{r,s}^{-1}(\hat{\boldsymbol{\beta}}, \hat{\alpha}) & \mathbf{0} \\ \mathbf{0}^T & FI_{p+1,p+1}^{-1}(\hat{\boldsymbol{\beta}}, \hat{\alpha}) \end{bmatrix}, \quad r, s = 1, \dots, p$$

gegeben ist mit Einträgen $FI_{r,s}^{-1}(\hat{\boldsymbol{\beta}}, \hat{\alpha})$ und $FI_{p+1,p+1}^{-1}(\hat{\boldsymbol{\beta}}, \hat{\alpha})$, die sich aus den Gleichungen (6.7) und (6.17) ausgewertet an $(\boldsymbol{\beta}, \alpha) = (\hat{\boldsymbol{\beta}}, \hat{\alpha})$ ergeben. Ein wesentlicher Punkt ist die asymptotische Unabhängigkeit zwischen $\hat{\boldsymbol{\beta}}$ und $\hat{\alpha}$.

6.2 Regressionsdiagnostiken

Um die Theorie der Negativ-Binomial-Regression mit geschätztem α abzuschließen, werden nun die Regressionsdiagnostiken für die Negativ-Binomial-Verteilung mit geschätztem α aufgeführt. Die Diagnostiken berechnen sich analog zu Kapitel 5, einziger Unterschied ist nun die Schätzung von α zu berücksichtigen. Deshalb werden der Form halber und der besseren Nachvollziehbarkeit die Diagnostiken hier beschrieben. Dabei bleibt anzumerken, dass die Unsicherheit in der Schätzung von α vernachlässigt wird.

Die Pearson-Residuen sehen demnach wie folgt aus:

$$r_i^P := \frac{y_i - \hat{\mu}_i(\hat{\alpha}, \hat{\boldsymbol{\beta}}(\hat{\alpha}))}{\hat{v}_i(\hat{\alpha}, \hat{\boldsymbol{\beta}}(\hat{\alpha}))}, \quad i = 1, \dots, n, \quad \hat{v}_i(\hat{\alpha}, \hat{\boldsymbol{\beta}}(\hat{\alpha})) \text{ geschätzte Varianzfunktion.}$$

Um eine Standardisierung durchzuführen, muss die Hatmatrix H definiert werden, welche nun vom geschätzten Wert $\hat{\boldsymbol{\beta}}(\hat{\alpha})$ und somit vom geschätzten Wert $\hat{\alpha}$ abhängt. Also ist die Hatmatrix durch

$$H = \mathbf{W}(\hat{\alpha}, \hat{\boldsymbol{\beta}}(\hat{\alpha}))^{1/2} \mathbf{X}(\mathbf{X}^T \mathbf{W}(\hat{\alpha}, \hat{\boldsymbol{\beta}}(\hat{\alpha})) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(\hat{\alpha}, \hat{\boldsymbol{\beta}}(\hat{\alpha}))^{1/2}$$

definiert. Eine Standardisierung ergibt also folgendes:

$$r_i^{P*} := \frac{y_i - \hat{\mu}_i(\hat{\alpha}, \hat{\beta}(\hat{\alpha}))}{\sqrt{\hat{v}_i(\hat{\alpha}, \hat{\beta}(\hat{\alpha}))(1 - h_{ii})}}, \quad i = 1, \dots, n.$$

Desweiteren können die Devianz-Residuen angegeben werden:

$$r_i^D := \text{sign}(y_i - \hat{\mu}_i(\hat{\alpha}, \hat{\beta}(\hat{\alpha})))\sqrt{d_i}, \quad \text{sign}(x) = \begin{cases} -1 & \text{für } x < 0 \\ 0 & \text{für } x = 0 \\ 1 & \text{für } x > 0 \end{cases},$$

wobei $d_i = 2(\mathcal{L}(y_i; y_i, \phi) - \mathcal{L}(y_i; \hat{\mu}_i(\hat{\alpha}, \hat{\beta}(\hat{\alpha})), \phi))$.

Nach einer Standardisierung erhält man folgende Formel:

$$r_i^{D*} = \frac{d_i}{\sqrt{1 - h_{ii}}}.$$

Als Schranke wird wieder der absolute Wert von 2 angesehen.

Nun sollen wiederum die komplexeren Regressionsdiagnostiken definiert werden. Zunächst wird die Formel für die $DFFITs_i$ angegeben, die wie folgt lautet:

$$DFFITs_i := \frac{\hat{\mu}_i(\hat{\alpha}, \hat{\beta}(\hat{\alpha})) - \hat{\mu}_{i,-i}(\hat{\alpha}_{-i}, \hat{\beta}_{-i}(\hat{\alpha}_{-i}))}{\sqrt{\hat{v}_{i,-i}(\hat{\alpha}_{-i}, \hat{\beta}_{-i}(\hat{\alpha}_{-i}))h_{ii}}},$$

wobei $\hat{v}_{i,-i}(\hat{\alpha}_{-i}, \hat{\beta}_{-i}(\hat{\alpha}_{-i})) = a(\phi)b''(\hat{\theta}_{i,-i}(\hat{\alpha}_{-i}, \hat{\beta}_{-i}(\hat{\alpha}_{-i})))$ die geschätzte Varianz von Y_i ohne Betrachtung der i -ten Beobachtung, $\hat{\theta}_{i,-i}(\hat{\alpha}_{-i}, \hat{\beta}_{-i}(\hat{\alpha}_{-i})) = b'(\hat{\mu}_{i,-i}(\hat{\alpha}_{-i}, \hat{\beta}_{-i}(\hat{\alpha}_{-i})))$ und h_{ii} das i -te Diagonalelement der Hatmatrix H sind. Schranken werden durch die absoluten Werte von 2 und $2\sqrt{p/n}$ gegeben.

Die $DFBETAs_{j,i}$ sind durch

$$DFBETAs_{j,i} := \frac{\hat{\beta}_j(\hat{\alpha}) - \hat{\beta}_{j,-i}(\hat{\alpha}_{-i})}{\sqrt{c_{jj}}}$$

gegeben, wobei $c_{jj} = ((\mathbf{X}^T \mathbf{W}(\hat{\alpha}, \hat{\beta}(\hat{\alpha})) \mathbf{X})^{-1})_{jj} \approx \text{Var}(\hat{\beta}_j(\hat{\alpha}))$. Um letztere Approximation zu zeigen, geht man analog zu Kapitel 5.4.1 vor. Schranken werden durch die absoluten Werte von 2 und $2\sqrt{1/n}$ gegeben.

Die Cook's Distanz lässt sich schreiben als

$$D_i = \frac{1}{p} \left(\hat{\beta}(\hat{\alpha}) - \hat{\beta}_{-i}(\hat{\alpha}_{-i}) \right)^T (\mathbf{X}^T \mathbf{W}(\hat{\alpha}, \hat{\beta}(\hat{\alpha})) \mathbf{X}) \left(\hat{\beta}(\hat{\alpha}) - \hat{\beta}_{-i}(\hat{\alpha}_{-i}) \right).$$

Eine Schranke bildet der Wert 1.

Um den $COVRATIO_i$ anzugeben, sind mehrere Schritte nötig. Diese sollen nachfolgend kurz vollzogen werden. Zunächst gilt, dass

$$\begin{aligned} COVRATIO_i &:= \frac{\det(\widehat{Var}(\hat{\beta}_{-i}(\hat{\alpha}_{-i})))}{\det(\widehat{Var}(\hat{\beta}(\hat{\alpha})))} \\ &= \frac{\det((\mathbf{X}_{-i}^T \mathbf{W}_{-i}(\hat{\alpha}_{-i}, \hat{\beta}_{-i}(\hat{\alpha}_{-i})) \mathbf{X}_{-i})^{-1})}{\det((\mathbf{X}^T \mathbf{W}(\hat{\alpha}, \hat{\beta}(\hat{\alpha})) \mathbf{X})^{-1})}. \end{aligned}$$

Mit Hilfe der Eigenschaften der Determinante, welche z.B. in Rade und Westergren (1997) (Seite 90) aufgeführt sind, gilt für den $COVRATIO_i$:

$$\begin{aligned} COVRATIO_i &= \frac{\det(\mathbf{X}^T \mathbf{W}(\hat{\alpha}, \hat{\beta}(\hat{\alpha})) \mathbf{X})}{\det(\mathbf{X}_{-i}^T \mathbf{W}_{-i}(\hat{\alpha}_{-i}, \hat{\beta}_{-i}(\hat{\alpha}_{-i})) \mathbf{X}_{-i})} \\ &= \frac{\det(\mathbf{X}^T \mathbf{X})}{\det(\mathbf{X}_{-i}^T \mathbf{X}_{-i})} \frac{\det(\mathbf{W}(\hat{\alpha}, \hat{\beta}(\hat{\alpha})))}{\det(\mathbf{W}_{-i}(\hat{\alpha}_{-i}, \hat{\beta}_{-i}(\hat{\alpha}_{-i})))}. \end{aligned}$$

Das Lemma 3.2 kann hier für den ersten Faktor verwendet werden. Der zweite Faktor entspricht einem Quotient von Determinanten von Diagonalmatrizen. Weiter gilt also:

$$COVRATIO_i = \frac{1}{1 - h_{ii}^{LM}} \frac{\prod_{j=1}^n w_j(\hat{\beta}(\hat{\alpha}))}{\prod_{j=1, j \neq i}^n w_j(\hat{\beta}_{-i}(\hat{\alpha}_{-i}))},$$

wobei hier das Diagonalelement $h_{ii}^{LM} = \mathbf{x}_i (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i^T$ ist.

Um die Devianzreduktion festzustellen, berechnet man zuerst die Devianz des vollständigen Datensatzes und dann die Devianz des Datensatzes nach Löschung der i -ten Beobachtung. Man erhält also folgende Formel:

$$\Delta D_i^1 := D(\mathbf{y}, \hat{\beta}(\hat{\alpha})) - D_{-i}(\mathbf{y}, \hat{\beta}_{-i}(\hat{\alpha}_{-i})).$$

Eine Schranke stellt der Wert 4 dar.

Kapitel 7

Regressionsdiagnostik für Negativ-Binomial-Modelle

Hier sollen nun die theoretischen Ergebnisse, die in den vorhergehenden Kapiteln erhalten wurden, konkret auf den Patentdatensatz angewandt werden. Mit Hilfe der Poisson-Verteilung wurde der Datensatz bereits in Kapitel 5 modelliert. Hier ging man von Äquidispersion aus, d.h. der Erwartungswert und die Varianz sind gleich. Aufgrund des möglichen Auftretens von Überdispersion beschreibt man ein Modell mit Hilfe der Negativ-Binomial-Verteilung, wobei der Parameter α anfangs fest sein soll. In einer dritten Modellierung soll der Parameter α unbekannt sein. Durch Schätzung des Parameters α kann man danach wiederum die Regressionsdiagnostiken berechnen.

7.1 Negativ-Binomial-Regression mit bekanntem α für den Patentdatensatz

Nun will man ein Modell mit Hilfe der Negativ-Binomial-Verteilung mit bekanntem α anpassen. In SPlus ist zwar ein Programm zur Berechnung von Negativ-Binomial-Modellen enthalten, jedoch lässt es nur diskrete α -Werte zu. Deshalb wurde von der Internetseite

<http://lib.stat.cmu.edu/S/>

das Programm *negbin* von B. Venables installiert, welches auch stetige α -Werte zulässt. Zu beachten ist, dass die Negativ-Binomial-Verteilung in SPlus eine andere Parameterisierung besitzt als in dieser Arbeit verwendet. In SPlus wird die Negativ-Binomial-Verteilung durch

$$\begin{aligned} P(Y = y) &= \frac{\Gamma(\theta + y)}{\Gamma(\theta)y!} \frac{\theta^\theta \mu^y}{(\theta + \mu)^{\theta+y}} \\ &= \frac{\Gamma(\theta + y)}{\Gamma(\theta)y!} \left(\frac{\theta}{\theta + \mu}\right)^\theta \left(\frac{\mu}{\theta + \mu}\right)^y \end{aligned}$$

definiert. Dies entspricht der Parameterisierung von $\theta = 1/\alpha$. Da man hier das α fest wählt, wird man sich für mehrere Werte von α Modelle ansehen. Um einen angemessenen Wertebereich für α bzw. θ zu bekommen, bedient man sich des *Signal-to-Noise-Quotient*, der wie folgt gegeben ist:

$$SN(\alpha, \mu_i) = \frac{E(Y_i)}{\sqrt{Var(Y_i)}}. \quad (7.1)$$

Für den Fall der Negativ-Binomial-Verteilung ergibt sich daraus

$$SN_{NB}(\alpha, \mu_i) = \frac{E(Y_i)}{\sqrt{Var(Y_i)}} = \frac{\mu_i}{\sqrt{\mu_i + \alpha\mu_i^2}} = \sqrt{\frac{\mu_i}{1 + \alpha\mu_i}}$$

Der Signal-to-Noise-Quotient konvergiert schnell gegen Null, wenn α Werte annimmt, die größer als 0.6 sind. Für eine weitere Untersuchung wird der Wert 1.25 für den Signal-to-Noise-Quotienten betrachtet. Da der Mittelwert \bar{y} für die Patentdaten bei 23.46 liegt, berechnet sich eine obere Schranke für α aus

$$\sqrt{\frac{\bar{y}}{1 + \alpha\bar{y}}} = 1.25 \Leftrightarrow 0.4 = \alpha$$

Somit ergibt sich für θ ein Wert von $1/0.4 = 2.5$. Ausgehend von $\alpha = 0$ bis $\alpha = 0.4$ wird dieser Bereich in äquidistante Intervalle der Länge 0.05 zerlegt. Es ergeben sich also neun Werte, somit neun verschiedene Negativ-Binomial-Modelle. Nun müssen die Werte für θ bestimmt werden, um in SPlus die Modelle einzugeben. Für den ersten Wert von α , der Null ist, wird wegen der Nicht-Definiertheit des Teilens durch Null der Wert $\alpha = 10^{-9}$ gewählt. Also ergeben sich die folgenden Werte:

$$\begin{aligned}\alpha &= (10^{-9}, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4) \\ \theta &= (10^9, 10, 6.66, 5, 4, 3.33, 2.86, 2.5)\end{aligned}$$

Nun können die Modelle eingegeben werden. Für das Negativ-Binomial-Modell nimmt man dieselben Kovariableneffekte wie beim Poissonmodell an. Mit den entsprechenden Haupteffekten und Interaktionen kann man nun die Negativ-Binomial-Modelle in SPlus eingeben, die wie folgt lauten:

```
patentnba1_glm(patente~rdsales+rdw5+rdsales*rdw5,family=negative.binomial
(theta=theta[1]),x=T)
patentnba2_glm(patente~rdsales+rdw5+rdsales*rdw5,family=negative.binomial
(theta=theta[2]),x=T)
patentnba3_glm(patente~rdsales+rdw5+rdsales*rdw5,family=negative.binomial
(theta=theta[3]),x=T)
patentnba4_glm(patente~rdsales+rdw5+rdsales*rdw5,family=negative.binomial
(theta=theta[4]),x=T)
patentnba5_glm(patente~rdsales+rdw5+rdsales*rdw5,family=negative.binomial
(theta=theta[5]),x=T)
patentnba6_glm(patente~rdsales+rdw5+rdsales*rdw5,family=negative.binomial
(theta=theta[6]),x=T)
patentnba7_glm(patente~rdsales+rdw5+rdsales*rdw5,family=negative.binomial
(theta=theta[7]),x=T)
patentnba8_glm(patente~rdsales+rdw5+rdsales*rdw5,family=negative.binomial
(theta=theta[8]),x=T)
patentnba9_glm(patente~rdsales+rdw5+rdsales*rdw5,family=negative.binomial
(theta=theta[9]),x=T)
```

Mit Hilfe der programmierten SPlus-Funktionen *diag.nba*, *dfbetas.nba*, *covratio.nba*, *devred.nba*, *inflbeob.hat*, *inflbeob.dffits* und *inflbeob.cook*, welche im Anhang beschrieben sind, kann der

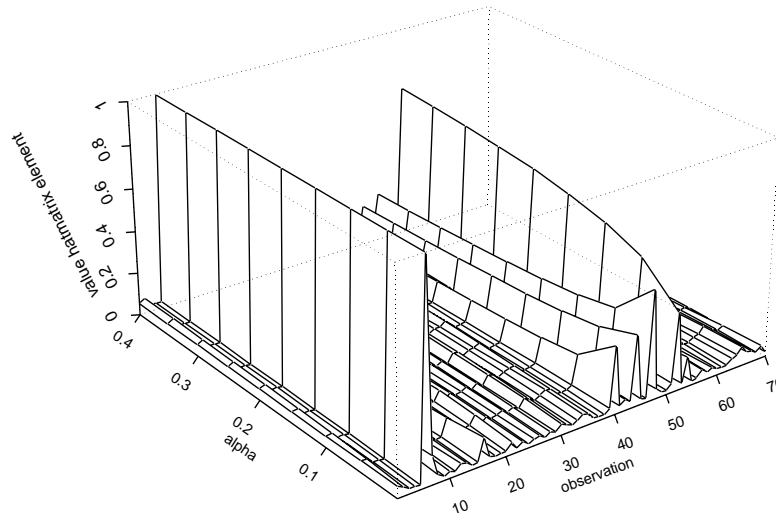


Abbildung 7.1: h_{ii} -Werte des Negativ-Binomial-Modells mit festem α für den Patentdatensatz für verschiedene Werte von α

Datensatz auf Ausreißer und einflußreiche Beobachtungen untersucht werden. In Abbildung 7.1 sieht man die h_{ii} -Werte der 70 Beobachtungen für die neun verschiedenen Werte von α .

Das h_{ii} -Element der Beobachtung 6 (ALZA CORP-CL A) ändert sich kaum. Es nimmt für jeden Wert α annähernd den Wert 1 an. Dagegen nehmen die Werte für die Beobachtungen 40 (JOHNSON & JOHNSON), 44 (LILLY (ELI) & CO) und 47 (MERCK & CO) mit zunehmendem Wert von α ab. Es stellt sich ab dem Wert 0.2 für α etwa ein konstanter Wert ein. Der h_{ii} -Wert der Beobachtung 52 nimmt im Gegensatz dazu mit steigendem α -Wert zu, es stellt sich ab $\alpha = 0.25$ wieder ein konstanter Wert ein.

In den Abbildungen 7.2 und 7.3 sieht man die Entwicklung der $DFFITs_i^1$ bzw. $DFFITs_i$ und der Cook's Distanz D_i^1 bzw. D_i für verschiedene Werte von α . Links jeweils die Werte für die Ein-Schritt-Approximation und rechts für den exakten Wert. Um die Veränderungen deutlich sichtbar zu machen, wurde die z-Achse skaliert wie in den jeweils unteren beiden Grafiken zu sehen ist.

In beiden Grafiken sticht die Beobachtung 6 heraus, deren absolute Werte für die $DFFITs_i^1$ bzw. $DFFITs_i$ und der Cook's Distanz D_i^1 bzw. D_i mit zunehmendem Wert für α abnehmen. Gleiches gilt ebenso für die Beobachtungen 40, 47 und 66 (SQUIBB CORP), die für kleines α relativ hohe Werte besitzen. Einzig die Beobachtung 52 nimmt für die Cook's Distanz mit steigendem α zu, wird also wie es scheint zunehmend einflußreicher. Die Approximationen sind gut, wie man v.a. aus den skalierten Grafiken sehen kann. Für den Wert $\alpha = 0$ (Poissonfall) wurde bereits in Kapitel 5 die ungenaue Approximation für die Beobachtung 6 angesprochen.

In Abbildung 7.4 wird nun der Wert des $COVRATIO_i^1$ bzw. $COVRATIO_i$ für die verschiedenen Werte von α abgetragen. Links sind die Werte der Approximation und rechts die exakten Werte zu sehen. Die unteren beiden Grafiken sind skaliert.

Der Wert für die Beobachtungen 40, 47, 66 und 69 (SYNTEX CORP) nimmt für steigendes α deutlich ab. Einzig der Wert für die Beobachtung 6 bleibt ab $\alpha = 0.15$ relativ hoch. Lässt man die hohen Werte für $\alpha = 0$ unberücksichtigt, so wird der exakte Wert gut durch die Ein-Schritt-Approximation approximiert.

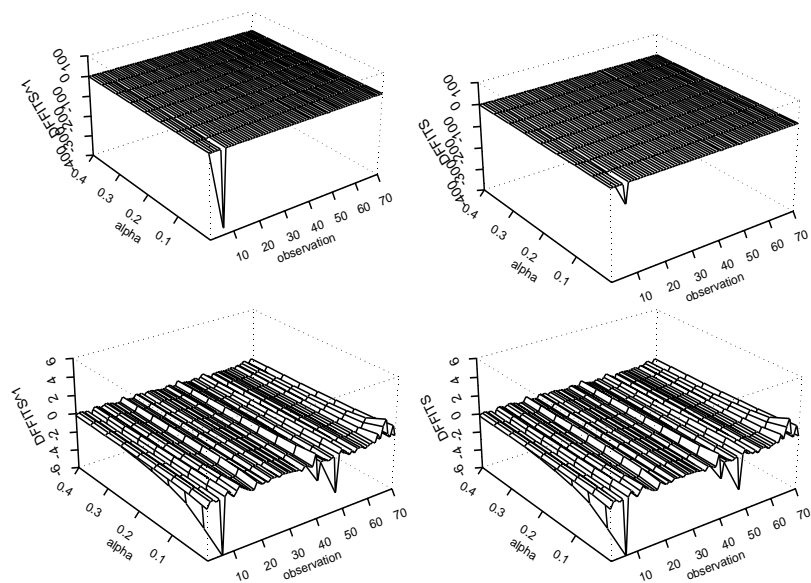


Abbildung 7.2: $DFFITs_i^1$ (links) und $DFFITs_i$ (rechts) des Negativ-Binomial-Modells mit festem α für den Patentdatensatz für verschiedene Werte von α ohne Skalierung (oben) und mit Skalierung ($|\cdot| \leq 6$) der z-Achse (unten)

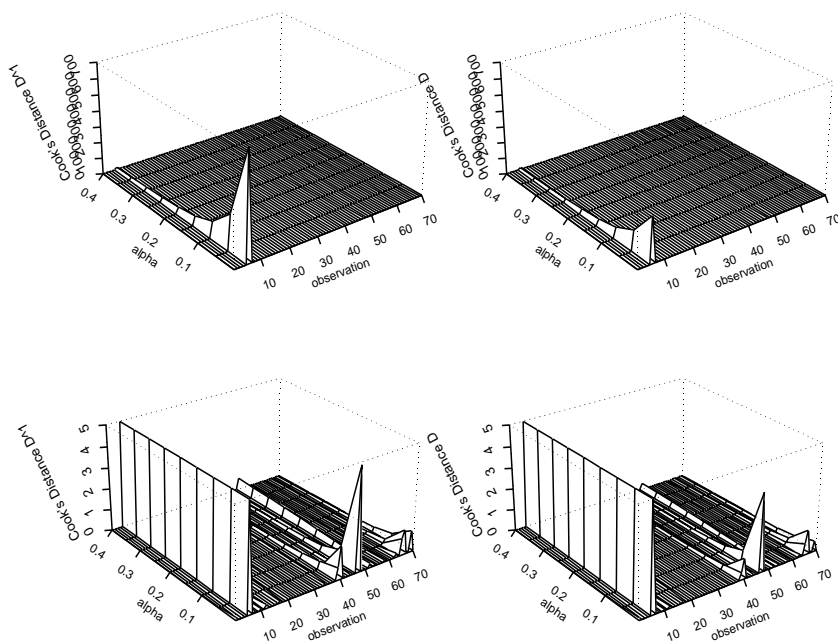


Abbildung 7.3: Cook's Distanz D_i^1 (links) und D_i (rechts) des Negativ-Binomial-Modells mit festem α für den Patentdatensatz für verschiedene Werte von α ohne Skalierung (oben) und mit Skalierung ($|\cdot| \leq 5$) der z-Achse (unten)

In der Abbildung 7.5 ist die Devianzreduktion ΔD_i^1 bzw. ΔD_i für alle 70 Beobachtungen und die verschiedenen Werte für α sichtbar. Die linke Grafik gibt die Werte für die Approximation aus, die rechte dagegen die exakten Werte. Die Approximationen spiegeln die exakten Werte recht gut.

Für die drei Beobachtungen 39 (INTL FLAVORS & FRAGRANCES), 47, 66 und 69 fällt die Devianzreduktion für $\alpha = 0.05$ schon rapide ab und wird auch mit zunehmendem α kleiner. Dagegen behalten die Beobachtungen 15 (BOCK DRUG-CL A), 17 (CARTER-WALLACE INC) und 22 (CODE LABORATORIES INC) ihr Niveau mit steigendem α -Wert. Wie aus der Grafik hervorgeht, sind die Devianzreduktionen ab einem Wert $\alpha = 0.25$ etwa konstant.

In den folgenden Tabellen 7.1, 7.2, 7.3 und 7.4, werden nochmals die auffälligsten Beobachtungen für die h_{ii} -Werte, die $DFITS_i^1$, die Cook's Distanz D_i^1 und die Devianzreduktion ΔD_i^1 für die verschiedenen Werte von α aufgelistet. Mit zunehmendem α werden die Beobachtungen, die die signifikanten Schranken überschreiten weniger, also haben nicht mehr so großen Einfluß auf die geschätzten Werte und das Modell.

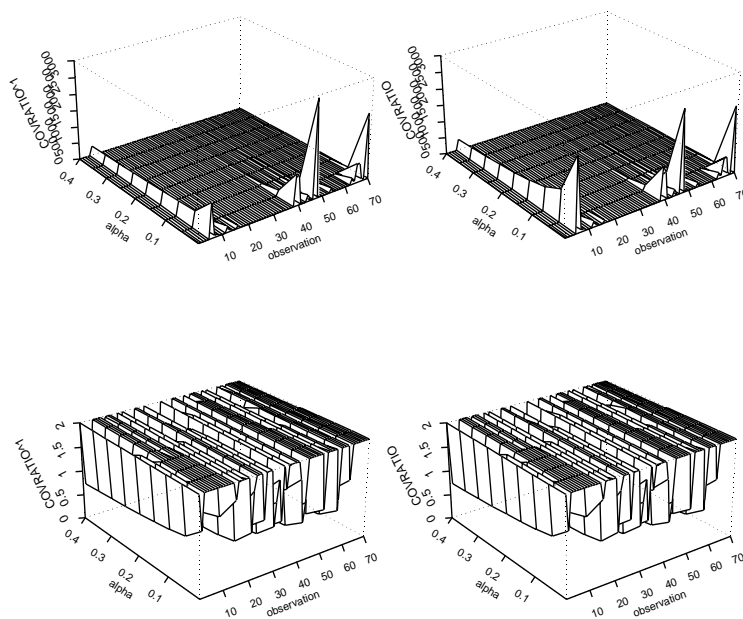


Abbildung 7.4: $COVRATIO_i^1$ (links) und $COVRATIO_i$ (rechts) des Negativ-Binomial-Modells mit festem α für den Patentdatensatz für verschiedene Werte von α ohne Skalierung (oben) und mit Skalierung ($|\cdot| \leq 2$) der z-Achse (unten)

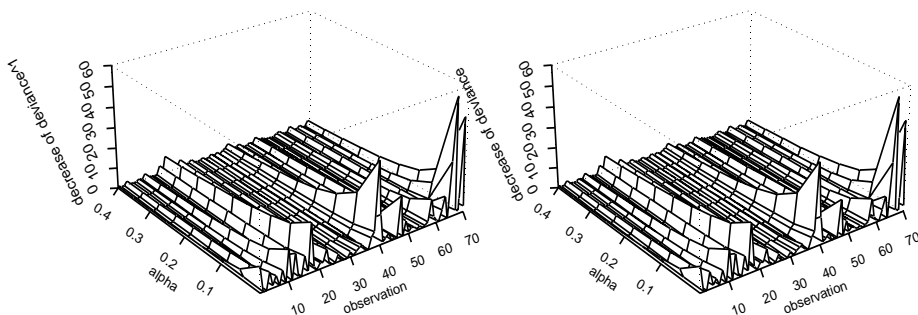


Abbildung 7.5: Werte der Devianzreduktion ΔD_i^1 (links) und ΔD_i (rechts) des Negativ-Binomial-Modells mit festem α für den Patentdatensatz für verschiedene Werte von α

<i>h_{ii}-Werte</i>	
α	<i>Beobachtungen</i>
0	6, 47, 52, 44, 40, 54, 7, 16
0.05	6, 52, 47, 44, 40, 7
0.1	6, 52, 47, 44, 40
0.15	6, 52, 47, 44, 40
0.2	6, 52, 47, 44, 40
0.25	6, 52, 47, 44, 40
0.3	6, 52, 47, 44, 40
0.35	6, 52, 47, 44, 40
0.4	6, 52, 47, 44, 40

Tabelle 7.1: Beobachtungen mit $h_{ii} \geq 2p/n$ des Negativ-Binomial-Modells mit festem α für den Patentdatensatz

<i>DFFIT_{S_i}</i>		
α	<i>Beobachtungen für Ein-Schritt-Approximation und exakten Wert</i>	
0	6, 47	6, 47, 66
0.05	6	6
0.1	6	6
0.15	6	6
0.2	6	6
0.25	-	-
0.3	-	-
0.35	-	-
0.4	-	-

Tabelle 7.2: Beobachtungen mit $DFFIT_{S_i}$ ¹ ≥ 2 bzw. $DFFIT_{S_i}$ ≥ 2 des Negativ-Binomial-Modells mit festem α für den Patentdatensatz

Cook's Distanz		
α	Beobachtungen für Ein-Schritt-Approximation und exakten Wert	
0	6, 47, 40, 66	6, 47, 66
0.05	6	6
0.1	6	6
0.15	6	6
0.2	6	6
0.25	6	6
0.3	6	6
0.35	6	6
0.4	6	6

Tabelle 7.3: Beobachtungen mit $D_i^1 > 1$ bzw. $D_i > 1$ des Negativ-Binomial-Modells mit festem α für den Patentdatensatz

Devianzreduktion		
α	Beobachtungen für Ein-Schritt-Approximation und exakten Wert	
0	66, 69, 39, 65, 17, 47 (19.7)	66, 69, 39, 65, 17, 47 (18.0)
0.05	39, 17, 69, 66, 10, 49 (7.1)	39, 17, 69, 66, 10, 49 (7.0)
0.1	17, 39, 69, 10, 15, 49 (6.0)	17, 39, 69, 10, 15, 49 (5.9)
0.15	17, 39, 69, 15, 10, 49 (5.1)	17, 39, 69, 15, 10, 49 (5.0)
0.2	17, 39, 69, 15, 22, 49 (4.5)	17, 39, 69, 15, 22, 10 (4.4)
0.25	17, 39, 15, 22, 69, 49 (4.0)	17, 39, 15, 22, 69, 49 (3.9)
0.3	17, 39, 15, 22	17, 39, 15, 22
0.35	17, 39, 15, 22	17, 39, 15, 22
0.4	17, 15, 39	17, 15, 39

Tabelle 7.4: Beobachtungen mit $\Delta D_i^1 \geq 4$ bzw. $\Delta D_i \geq 4$ des Negativ-Binomial-Modells mit festem α für den Patentdatensatz. Es sind bis $\alpha = 0.25$ jeweils die 6 höchsten Werte aufgeführt. Die Zahl in runden Klammern gibt jeweils die Devianzreduktion der 6-höchsten Beobachtung an.

7.2 Negativ-Binomial-Regression mit geschätztem α für den Patentdatensatz

In diesem Kapitel sollen nun die Patent-Daten mit Hilfe der Negativ-Binomial-Verteilung mit unbekanntem α modelliert werden. Hierzu wird wiederum das Programm *negbin* von B. Venables, welches im vorherigen Kapitel erwähnt wurde, benötigt. Das Modell, welches in SPlus eingegeben wird, lautet wie folgt:

```
patentnb_glm.nb(patente~rdsales+rdw5+rdsales*rdw5,link=log,x=T)
```

Mit Hilfe der *summary*-Funktion erhält man Informationen über die geschätzten Regressionskoeffizienten $\hat{\beta}_j$, $j = 1, \dots, 4$, und dem geschätztem $\hat{\alpha}$ bzw. $\hat{\theta} = 1/\hat{\alpha}$.

```
Call: glm.nb(formula = patente ~ rdsales + rdw5 + rdsales * rdw5, x = T,
link = log, init.theta = 3.40016179771382)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.045781	-1.033093	-0.254448	0.4505105	2.283661

Coefficients:

	Value	Std. Error	t value
(Intercept)	-1.855345	0.3980563	-4.661011
rdsales	-8.234816	5.9981504	-1.372893
rdw5	2.636001	0.2514156	10.484634
rdsales:rdw5	5.742963	3.9631379	1.449095

(Dispersion Parameter for Negative Binomial family taken to be 1)

Null Deviance: 524.2385 on 69 degrees of freedom

Residual Deviance: 78.53238 on 66 degrees of freedom

Number of Fisher Scoring Iterations: 1

Theta:	3.40016
Std. Err.:	0.9251

2 x log-likelihood: 10233.61782

Tabelle 7.5: *summary* des Negativ-Binomial-Modells mit geschätztem α für den Patentdatensatz

Aus der Tabelle 7.5 kann man erkennen, dass die Kovariable *rdsales* und die Interaktion einen nicht mehr so deutlichen Einfluß haben, wie es beim Poissonmodell noch war, da der Wert für den *t value* unterhalb von 2 liegt. Das Modell wird aber weiter betrachtet, da die Vergleichbarkeit mit vorherigen Modellen erhalten bleiben soll. Jedoch zeigt der Wert der Devianz mit 78.53 einen deutlich geringeren Wert als dies im Poissonmodell (377.18) der Fall war. Das Verhältnis von Devianz und der Anzahl der Freiheitsgrade mit 66 ist nun deutlich besser, was auf eine bessere Anpassung der Daten durch das Negativ-Binomial-Modell schließen lässt. Schließlich ist

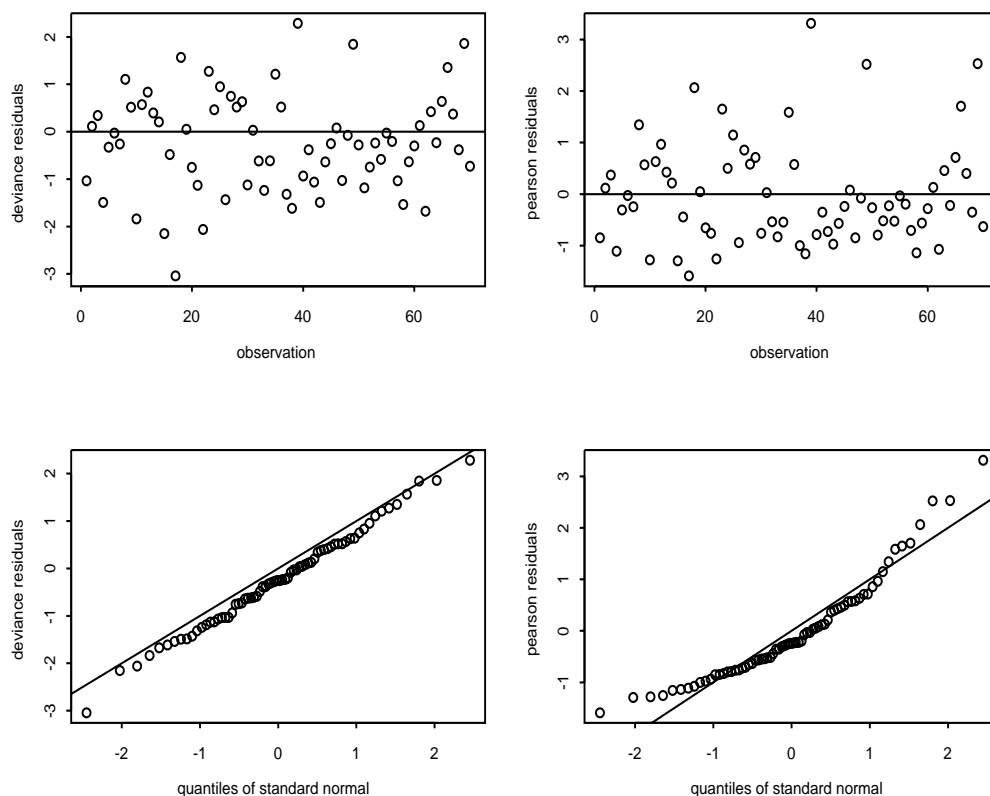


Abbildung 7.6: Devianz- und Pearson-Residuen des Negativ-Binomial-Modells mit geschätztem α für den Patentdatensatz (oben), Plot der Standardnormalverteilungsquantile gegen Devianz- und Pearson-Residuen (unten). Die durchgezogene Linie gibt die Winkelhalbierende an.

der geschätzte Wert $\hat{\theta}$ für θ mit 3.4 aufgeführt und die geschätzte Standardabweichung beträgt 0.93. Es ergibt sich also für $\hat{\alpha}$ ein Wert von $\hat{\alpha} = 1/\hat{\theta} = 1/3.4 = 0.29$, welcher relativ groß und ein Anzeichen für Überdispersion ist. Das Negativ-Binomial-Modell ist damit für den Patentdatensatz gut geeignet. Mit Hilfe der Delta-Methode, welche in Bickel und Doksum (1977) beschrieben ist, lässt sich auch die geschätzte Standardabweichung von $\hat{\alpha}$ berechnen. Diese ist 0.08.

Nun werden die Residuen für den Patentdatensatz berechnet. In der Abbildung 7.6 werden die Devianz- und Pearson-Residuen grafisch dargestellt.

In den oberen beiden Grafiken lässt sich erkennen, dass die Residuen zufällig um Null streuen. Die höchsten Werte der Devianz- und Pearson-Residuen liegen absolut über dem Wert 3 und folglich nicht mehr so hoch wie im Poissonmodell. Die unteren beiden Grafiken zeigen die Quantile der Standardnormalverteilung gegenüber den beiden Residuen. Hier zeigt sich ebenfalls eine deutliche Verbesserung zum Poissonmodell, da sich die Residuen nun deutlich entlang der Winkelhalbierenden bewegen. Sehr deutlich ist dies für die Devianz-Residuen zu sehen. Man kann also von einer Standardnormalverteilung der Residuen ausgehen.

Um nun einflussreiche Beobachtungen zu entdecken, wurden in SPlus ebenso Funktionen geschrieben, die die in Kapitel 6.2 aufgeführten Diagnostiken widerspiegeln. Anders als beim Poissonmodell und dem Negativ-Binomial-Modell geht man hier nicht von einer Ein-Schritt-Approximation

für β_{-i} aus, sondern man betrachtet die wahren geschätzten Werte $(\hat{\beta}_{-i}(\hat{\alpha}_{-i}), \hat{\alpha}_{-i})$ für $(\beta_{-i}(\alpha_{-i}), \alpha_{-i})$. Die Funktion *diag.nb* gibt die entsprechenden Werte für die h_{ii} -Werte, die $DFFITs_i$ und die Cook's Distanz D_i aus. Die Funktion wird im Anhang erklärt. In Abbildung 7.7 sieht man die Werte, die in Tabelle 7.6 aufgelistet sind, jeweils für die h_{ii} -Werte, die $DFFITs_i$ und die Cook's Distanz D_i anhand eines Stabdiagramms.

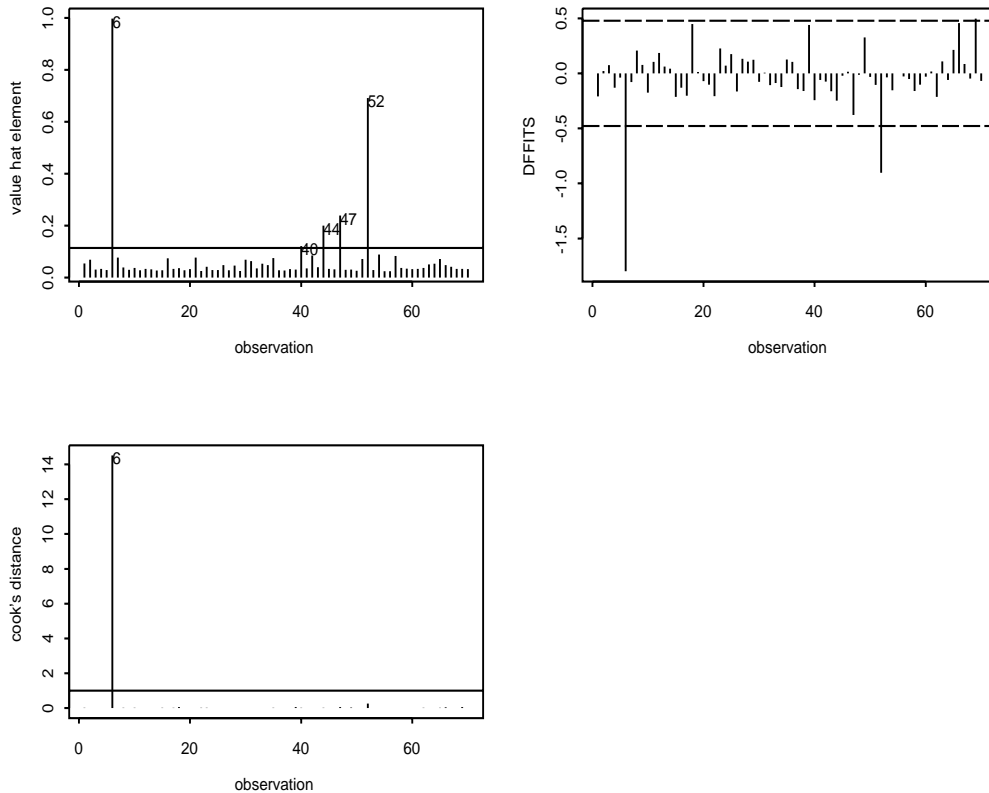


Abbildung 7.7: Stabdiagramme der h_{ii} -Werte (links oben), der $DFFITs_i$ (rechts oben) und der Cook's Distanz D_i (links unten) des Negativ-Binomial-Modells mit geschätztem α für den Patentdatensatz. Die durchgezogenen bzw. gestrichelten Linien markieren die absoluten Schranken (links oben $2p/n = 8/70 = 0.11$, rechts oben $2\sqrt{p/n} = 2\sqrt{4/70} = 0.48$, links unten 1)

Links oben werden die h_{ii} -Werte dargestellt. Die Beobachtungen 6(ALZA CORP-CL A), 40 (JOHNSON & JOHNSON), 44 (LILLY (ELI) & CO), 47 (MERCK & CO) und 52 (NEWPORT PHARMACEUTICALS INTL) überschreiten die Schranke von $2p/n = 0.11$. Diese Beobachtungen werden somit als x-Ausreißer identifiziert. In der Grafik rechts oben sind die Werte für die $DFFITs_i$ sichtbar. Keine Beobachtung überschreitet die Schranke von 2. Lediglich die Schranke von $2\sqrt{p/n} = 0.48$, die die gestrichelte Linie darstellt, wird von den Beobachtungen 6, 52 und 69 (SYNTEX CORP) überschritten. Die Grafik links unten zeigt die Werte der Cook's Distanz D_i . Die Beobachtung 6 ragt hier mit einem Wert von 14.5 hervor und wird somit aufgrund des Überschreitens des Werts 1 als einflußreich deklariert.

Die Funktion *dfbetas.nb* gibt die entsprechenden Werte für die $DFBETAS_{j,i}$ an. In Abbildung 7.8 sieht man die Werte jeweils für die entsprechenden Kovariablen anhand eines Stabdiagramms.

\$hat:

```
[1] 0.05474831 0.06872539 0.03119192 0.03367481 0.02884634 0.99678892
[7] 0.07750807 0.03899198 0.03041371 0.03735951 0.02840877 0.03332601
[13] 0.03235975 0.02717602 0.02785908 0.07438255 0.03347305 0.03643430
[19] 0.02864429 0.03310374 0.07705452 0.02537278 0.04155767 0.02885092
[25] 0.02964012 0.04815709 0.02821890 0.04613059 0.02538979 0.06897818
[31] 0.06340005 0.03530771 0.05331486 0.04799575 0.07517822 0.02829567
[37] 0.02759734 0.03323174 0.03070676 0.12188372 0.03589878 0.08568692
[43] 0.04005760 0.20001492 0.03402124 0.03194093 0.23947115 0.02995672
[49] 0.03059666 0.02690594 0.07141167 0.69109042 0.02945870 0.08909689
[55] 0.02498028 0.02423358 0.08329619 0.03720918 0.03425485 0.03292842
[61] 0.03355472 0.03700051 0.05083131 0.05405491 0.07195447 0.04795978
[67] 0.04160420 0.03400638 0.03371658 0.03312722
```

\$dffits:

```
[1] -0.2091285604 0.0201167696 0.0742570896 -0.1304770400 -0.0395659843
[6] -1.7986339252 -0.0786413989 0.2074739374 0.0759403335 -0.1750223483
[11] 0.1049821030 0.1857734902 0.0605461078 0.0426503498 -0.2145539120
[16] -0.1313902794 -0.2026994975 0.4470660572 0.0128244506 -0.0715610442
[21] -0.1036105045 -0.2067222989 0.2267377630 0.0705147023 0.1738576395
[26] -0.1641163947 0.1329956197 0.1059296277 0.1232819126 -0.0779513865
[31] 0.0070401240 -0.1067939937 -0.0879881882 -0.1235031845 0.1264547744
[36] 0.1047378918 -0.1431534913 -0.1600916727 0.4393735888 -0.2432540195
[41] -0.0607660279 -0.0742322762 -0.1635068490 -0.2479733786 -0.0223875700
[46] 0.0148549289 -0.3784401380 -0.0122823413 0.3252260555 -0.0324654061
[51] -0.1053514147 -0.9023085568 -0.0375369306 -0.1533563901 0.0003478223
[56] -0.0282057608 -0.0522212447 -0.1606823345 -0.1037626949 -0.0292925459
[61] 0.0176148757 -0.2144733429 0.1076909542 -0.0611102294 0.2131876760
[66] 0.4567967337 0.0846827902 -0.0476175183 0.4961477142 -0.0691202579
```

\$cook:

```
[1] 1.193130e-002 1.858159e-004 1.391186e-003 7.040152e-003 4.930790e-004
[6] 1.451194e+001 1.686621e-003 1.468121e-002 2.081055e-003 1.137966e-002
[11] 3.108931e-003 8.858583e-003 1.245805e-003 4.728526e-004 1.274806e-002
[16] 4.836408e-003 1.672632e-002 5.233484e-002 4.507453e-005 2.115375e-003
[21] 6.096607e-003 1.147578e-002 2.172798e-002 1.688640e-003 9.618806e-003
[26] 9.398987e-003 5.210608e-003 3.753254e-003 3.979337e-003 4.328296e-003
[31] 2.178712e-005 2.939866e-003 4.610308e-003 4.143966e-003 1.528556e-002
[36] 2.836613e-003 6.372159e-003 9.183052e-003 6.091734e-002 2.025323e-002
[41] 1.073862e-003 4.341449e-003 8.770536e-003 2.209112e-002 2.318702e-004
[46] 9.472729e-005 5.557775e-002 8.202301e-005 3.579187e-002 3.242706e-004
[51] 6.218327e-003 2.590153e-001 3.593712e-004 7.429005e-003 2.284554e-006
[56] 2.017559e-004 2.903064e-003 9.174647e-003 2.999265e-003 3.792669e-004
[61] 1.522349e-004 1.227835e-002 3.066790e-003 1.057105e-003 1.190280e-002
[66] 4.826345e-002 1.887117e-003 8.292532e-004 6.057710e-002 1.982057e-003
```

Tabelle 7.6: h_{ii} -Werte, $DFFITs_i$ und Cook's Distanz D_i des Negativ-Binomial-Modells mit geschätztem α für den Patentdatensatz

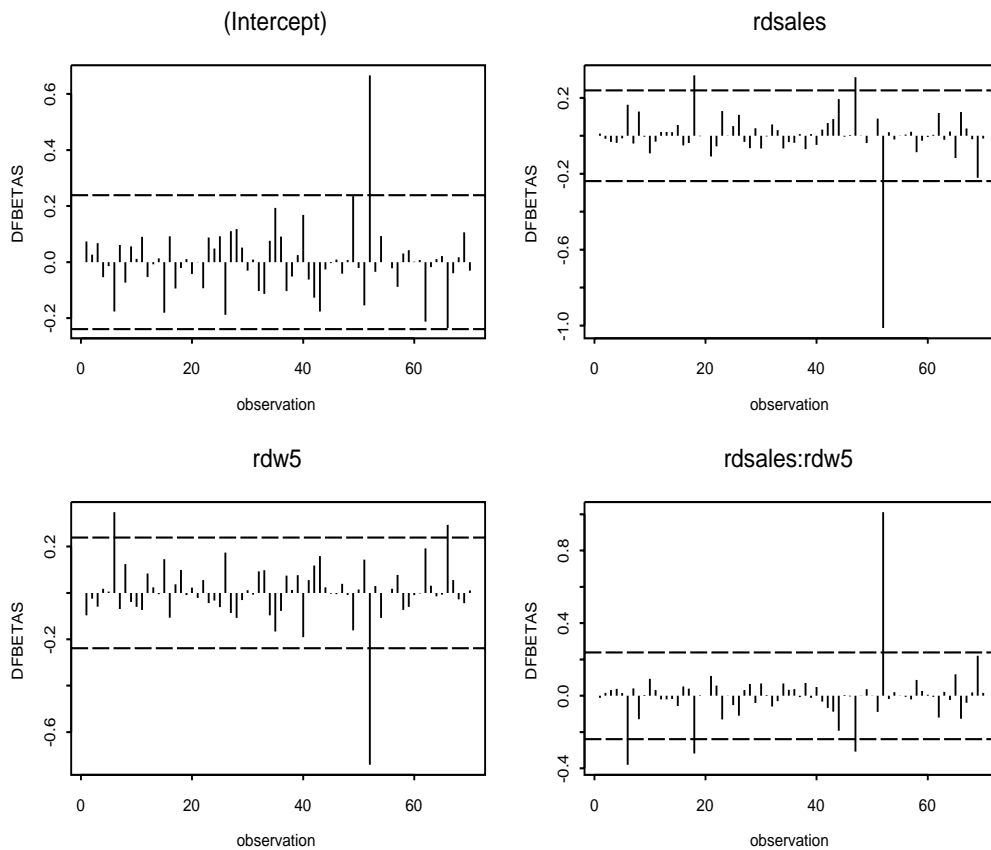


Abbildung 7.8: Stabdiagramme der $DFBETAS_{j,i}$ des Negativ-Binomial-Modells mit geschätztem α für den Patentdatensatz. Die gestrichelten Linien markieren die absoluten Schranken $2\sqrt{1/n} = 2\sqrt{1/70} = 0.24$.

Aus allen vier Grafiken lässt sich erkennen, dass die Schranke 2 von keinem Wert überschritten wird. Dagegen wird die Schranke von $2/\sqrt{n} = 0.24$ mehrmals überschritten. Auffallend ist die Beobachtung 52, die für jede Kovariable einen betragsmäßig größeren $DFBETAS_{j,i}$ -Wert als 0.24 zeigt. Folglich ändert sich also jeder Regressionskoeffizient nach Löschung der Beobachtung 52 signifikant. Die Beobachtung 6 zeigt für die Kovariable $rdw5$ und die Interaktion signifikante Werte, wohingegen die $DFBETAS_{j,i}$ der Kovariable $rdsales$ und der Interaktion für die Beobachtungen 18 (CAVITRON CORP) und 47 den Wert von 0.24 überschreitet. Für die Beobachtung 66 (SQUIBB CORP) übersteigt der $DFBETAS_{j,i}$ -Wert der Kovariable $rdw5$ die signifikante Schranke.

Die Funktion `covratio.nb` gibt die entsprechenden Werte für den $COVRATIO_i$ an. In Tabelle 7.7 sind die geordneten Werte des $COVRATIO_i$ aufgelistet, dazu die Indizes der Beobachtungen. In Abbildung 7.9 sieht man die Werte anhand eines Stabdiagramms. In der unteren Grafik ist die y-Achse skaliert worden, um die Werte besser sichtbar zu machen.

Aus der oberen Grafik geht hervor, dass der $COVRATIO_i$ für die Beobachtung 6 sehr groß ist. Man sieht in der unteren Grafik, dass ein Drittel der Werte einen Wert von 5 überschreitet und mehr als die Hälfte der Werte einen Wert von 3. Somit haben die meisten der Beobachtungen einen deutlichen Einfluss auf das Modell.

\$covratio.ord:

```
[1] 1.090250e+003 1.596191e+001 1.592663e+001 1.503486e+001 1.298314e+001
[6] 1.280177e+001 1.275659e+001 1.254720e+001 1.166176e+001 1.098107e+001
[11] 1.051104e+001 8.475655e+000 7.871194e+000 7.521420e+000 7.412612e+000
[16] 7.255684e+000 7.035488e+000 6.791412e+000 6.762348e+000 6.650704e+000
[21] 6.556757e+000 6.332678e+000 5.654627e+000 5.483123e+000 5.477931e+000
[26] 4.657106e+000 3.862056e+000 3.836762e+000 3.803114e+000 3.687074e+000
[31] 3.676993e+000 3.580804e+000 3.514158e+000 3.433209e+000 3.314835e+000
[36] 3.150906e+000 3.047561e+000 3.006270e+000 2.454584e+000 2.320209e+000
[41] 2.292928e+000 2.289791e+000 2.285208e+000 2.247277e+000 1.790927e+000
[46] 1.190818e+000 1.188095e+000 1.161468e+000 1.083805e+000 9.862259e-001
[51] 6.005215e-001 4.184347e-001 3.389991e-001 3.284051e-001 3.219158e-001
[56] 3.185977e-001 3.103199e-001 2.979718e-001 2.943146e-001 2.900790e-001
[61] 2.836598e-001 2.710160e-001 2.636436e-001 1.911171e-001 1.143411e-001
[66] 9.708546e-002 7.225838e-002 6.808715e-002 1.279560e-002 1.466222e-003
```

\$beob:

```
[1] 6 64 44 7 48 54 16 67 13 63 61 68 60 34 47 59 9 45 40 46 24 65 27 12
[25] 11 36 1 8 14 3 29 55 52 18 25 50 23 5 19 70 56 28 53 20 35 31 41 2
[49] 32 49 66 37 51 26 33 42 43 57 30 21 4 58 62 38 22 15 10 69 39 17
```

Tabelle 7.7: Geordnete Werte des $COVRATIO_i$ des Negativ-Binomial-Modells mit geschätztem α für den Patentdatensatz

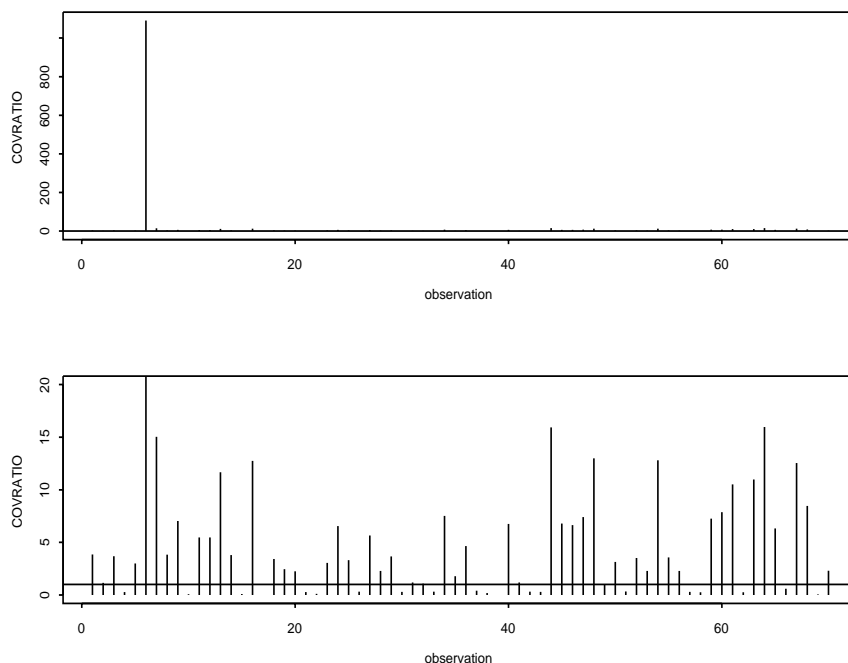


Abbildung 7.9: Stabdiagramm des $COVRATIO_i$ des Negativ-Binomial-Modells mit geschätztem α für den Patentdatensatz ohne Skalierung (oben) und mit Skalierung ($|\cdot| \leq 20$) der y-Achse (unten). Die durchgezogene Linie kennzeichnet den Wert 1.

Die Funktion *devred.nb* gibt die entsprechenden Werte für die Devianzreduktion an. In Tabelle 7.8 sieht man die geordneten Werte und die Indizes der Beobachtungen für diese Werte. Die Werte werden in Abbildung 7.10 grafisch anhand eines Stabdiagramms dargestellt.

Die Abbildung 7.10 zeigt, dass keiner der Werte den Wert 4 übersteigt. Die höchste Devianzreduktion erhält man für die Beobachtung 17 mit einem Wert von 3.32. Die weiteren Beobachtungen sind 15 (BOCK DRUG-CL A), 22 (CODE LABORATORIES INC), 62 (SCHERER (R.P.)), 23 (CONCEPT INC) und 18. Für die Beobachtung 18 bekommt man noch eine Reduktion der Devianz von 1.94. Für die Beobachtungen 39 (INTL FLAVORS & FRAGRANCES) und 69 erhöht sich dagegen die Devianz um 0.29 bzw. 0.06.

Der Patentdatensatz konnte mit den angegebenen Regressionsdiagnostiken sehr gut analysiert werden. Die Beobachtungspunkte waren mit $n = 70$ nicht zu niedrig und nicht zu hoch, so dass die Regressionsdiagnostiken für eine einzelne Beobachtung nach deren Weglassen sichtbaren Einfluß zeigten. Die zumeist auffallend hohe Überschätzung durch die Beobachtung 6 kann aus der ungenaueren Schätzung der Ein-Schritt-Approximation resultieren, die für diese Beobachtung am schlechtesten war. Aus Sicht der Modellanpassung ist das Negativ-Binomial-Modell zu präferieren.


```

$devred.ord:
 [1] 3.31553280 2.67044881 2.53676562 1.96877129 1.94439518 1.93846934
 [7] 1.84910919 1.64536396 1.59303304 1.54822846 1.47463363 1.45962729
[13] 1.42955305 1.41905740 1.41831984 1.38430565 1.37644229 1.37305367
[19] 1.36912421 1.34057674 1.25690167 1.25013885 1.24452721 1.23627171
[25] 1.21847748 1.21247599 1.20887643 1.20885756 1.18016569 1.17477893
[31] 1.16209895 1.14479539 1.13145356 1.13062313 1.10081501 1.09487192
[37] 1.09069875 1.08801668 1.08196761 1.02829561 1.02739943 1.02586268
[43] 1.02284882 1.01638111 1.01347075 0.99530764 0.93854609 0.89369547
[49] 0.87719430 0.84223090 0.83738406 0.83635072 0.83146268 0.80246297
[55] 0.79648156 0.74808230 0.73891048 0.73859490 0.60427544 0.59529430
[61] 0.59092981 0.58803807 0.55068027 0.50756057 0.47437327 0.40985741
[67] 0.39158120 0.37254995 -0.06329414 -0.29395039
    
```

```

$beob:
 [1] 17 15 22 62 23 18 35 43 49 26 64 7 44 16 54 67 25 48 63 13 34 61 59 68
[25] 40 65 33 27 6 60 12 9 47 51 8 24 11 45 1 38 29 21 46 30 36 52 42 37
[49] 57 3 20 70 28 4 14 55 50 5 19 58 53 56 10 32 66 41 2 31 69 39
    
```

Tabelle 7.8: Geordnete Werte der Devianzreduktion des Negativ-Binomial-Modells mit geschätztem α für den Patentdatensatz

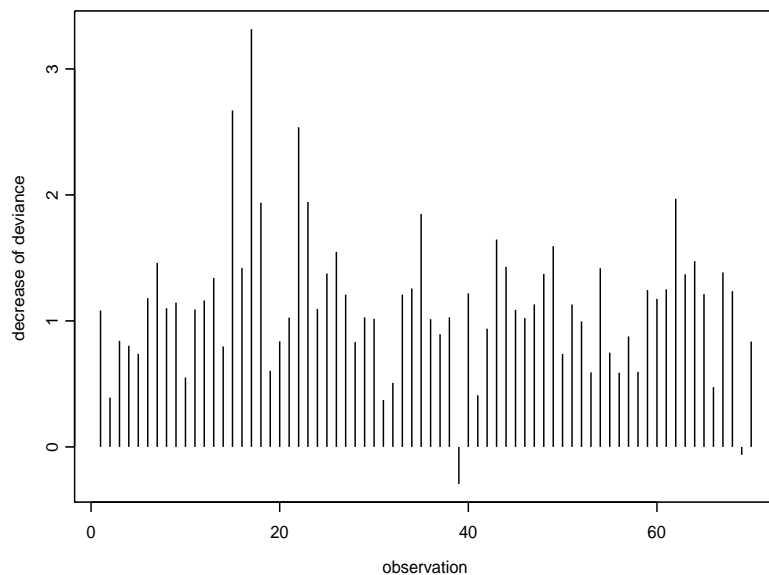


Abbildung 7.10: Stabdiagramm der Werte der Devianzreduktion des Negativ-Binomial-Modells mit geschätztem α für den Patentdatensatz

Kapitel 8

Fallstudie: Aggregierter KFZ-Haftpflichtdatensatz

8.1 Offset in der Regression

Bisher wurde angenommen, dass bei der Regression die Beobachtungszeiträume für die Beobachtungen y_i für alle $i = 1, \dots, n$ einheitlich sind. In der Praxis kommt es aber häufig vor, dass die Beobachtungszeiträume der analysierten Daten variieren. Deshalb wird eine zusätzliche Variable t_i als Gewichtung eingeführt, die bekannt ist und sich auf den jeweiligen Zeitraum der i -ten Beobachtung bezieht, um die Zeiteinheit zu standardisieren. Für die Poissonverteilung bedeutet dies, dass die zu untersuchende Variable Y_i poissonverteilt mit dem Parameter $t_i \mu_i^*$ ist, d.h.

$$P(Y_i = y_i) = e^{-t_i \mu_i^*} \frac{(t_i \mu_i^*)^{y_i}}{y_i!}, \quad i = 1, \dots, n, \quad y_i = 0, 1, 2, \dots,$$

wobei $\mu_i^* = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$ ist. Die Dichte kann in Form der exponentiellen Familie wie folgt geschrieben werden:

$$f(y_i; \theta, \phi) = \exp(-t_i \mu_i^* + y_i \log(t_i \mu_i^*) - \log(y_i!)).$$

Dabei ist $a(\phi) = 1$, $b(\theta) = e^\theta$ und $c(y_i, \phi) = -\log(y_i!)$. Der Parameter θ_i ist nun gegeben durch $\theta_i = \log(t_i \mu_i^*)$. Damit ergibt sich für den Erwartungswert und die Varianz

$$\begin{aligned} E(Y_i) &= \mu_i = b'(\theta_i) = e^{\theta_i} = t_i \mu_i^* = t_i \exp(\mathbf{x}_i^T \boldsymbol{\beta}) \\ \text{Var}(Y_i) &= b''(\theta_i) a(\phi) = e^{\theta_i} = t_i \mu_i^* = t_i \exp(\mathbf{x}_i^T \boldsymbol{\beta}). \end{aligned}$$

Für den linearen Prädiktor gilt dann

$$\eta_i = \mathbf{x}_i^T \boldsymbol{\beta} = \log(\mu_i) - \log(t_i) = \theta_i - \log(t_i),$$

wobei der letzte Term $\log(t_i)$ den Offset darstellt, welcher als bekannt vorausgesetzt wird. Ist $t_i = 1 \forall i$, so erhält man die kanonische Linkfunktion

$$g(\mu_i) = \log(\mu_i),$$

während man für den Fall $t_i \neq 1$ die kanonische Linkfunktion

$$g(\mu_i) = \log(\mu_i) - \log(t_i)$$

bekommt.

Die Negativ-Binomial-Verteilung kann man ähnlich erweitern. Die zu untersuchende Variable Y_i ist negativbinomialverteilt mit den Parametern $t_i\mu_i^*$ und α , d.h.

$$P(Y_i = y_i) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y_i + 1)} \left(\frac{1}{1 + \alpha t_i \mu_i^*} \right)^{\alpha^{-1}} \left(\frac{\alpha t_i \mu_i^*}{1 + \alpha t_i \mu_i^*} \right)^{y_i}$$

für $\alpha > 0, y_i = 0, 1, 2, \dots$

wobei $\mu_i^* = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$ ist. Die Dichte kann in Form der exponentiellen Familie wie folgt geschrieben werden:

$$f(y_i; \theta, \phi) = \exp \left\{ y_i \log \left(\frac{\alpha t_i \mu_i^*}{1 + \alpha t_i \mu_i^*} \right) - \left(-\alpha^{-1} \log \left(\frac{1}{1 + \alpha t_i \mu_i^*} \right) \right) + \log \left(\frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y_i + 1)} \right) \right\}.$$

Es ist $a(\phi) = 1$, $b(\theta_i) = -\alpha^{-1} \log(1 - e^{\theta_i})$ und $c(y_i, \phi) = \log \left(\frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y_i + 1)} \right)$. Der Parameter θ_i ist gegeben durch $\theta_i = \log \left(\frac{\alpha t_i \mu_i^*}{1 + \alpha t_i \mu_i^*} \right)$. Damit ergibt sich folglich für den Erwartungswert und die Varianz

$$\begin{aligned} E(Y_i) &= \mu_i = b'(\theta_i) = e^{\theta_i} = t_i \mu_i^* = t_i \exp(\mathbf{x}_i^T \boldsymbol{\beta}) \\ \text{Var}(Y_i) &= b''(\theta_i) a(\phi) = \alpha^{-1} \frac{e^{\theta_i} (1 - e^{\theta_i}) - e^{\theta_i} (-e^{\theta_i})}{(1 - e^{\theta_i})^2} 1 \\ &= t_i \mu_i^* (1 + \alpha t_i \mu_i^*) \\ &= t_i \exp(\mathbf{x}_i^T \boldsymbol{\beta}) (1 + \alpha t_i \exp(\mathbf{x}_i^T \boldsymbol{\beta})). \end{aligned}$$

Für den linearen Prädiktor gilt dann

$$\eta_i = \mathbf{x}_i^T \boldsymbol{\beta} = \log(\mu_i) - \log(t_i),$$

wobei der letzte Term $\log(t_i)$ den Offset darstellt, welcher als bekannt vorausgesetzt wird. Für $t_i = 1 \forall i$ erhält man wiederum

$$\eta_i = \mathbf{x}_i^T \boldsymbol{\beta} = \log(\mu_i),$$

da $\log(1) = 0$.

8.2 Ein aggregierter KFZ-Haftpflichtdatensatz

8.2.1 Beschreibung des aggregierten KFZ-Haftpflichtdatensatzes

Der hier bearbeitete Datensatz (KFZ-Datensatz) ist auf der Internetseite

<http://www.statsci.org/data/general/carinsca.txt>

zu finden und wird ebenso in Bailey und Simon (1960) verwendet. Wegen der wenigen Datenpunkte wurde der Datensatz in den Anhang F gestellt. Der Datensatz beschreibt die kanadischen Autoversicherungsdaten in den Versicherungsjahren 1957-58. Die Daten wurden vom Statistischen Amt (Canadian Underwriters' Association - Statistical Department) zusammengetragen und enthalten jedes Versicherungsunternehmen in Kanada. Die gegebenen Daten stammen aus der privaten KFZ-Haftpflicht, ausgeschlossen sind Landwirte und der Staat Saskatchewan. Der Datensatz umfasst 6 Variablen. Die erste Spalte *Merit* gibt an, wie lange der Halter des Fahrzeugs unfallfrei war. Dazu wurden 4 Klassen definiert:

- 3 - zugelassen und unfallfrei seit 3 oder mehr Jahren
- 2 - zugelassen und unfallfrei seit 2 Jahren
- 1 - zugelassen und unfallfrei seit 1 Jahr
- 0 - alle anderen

Die zweite Variable *Class* ist ein Verbund aus Geschlecht, Alter, Nutzung und Familienstand. Die Variable wird in 5 Klassen unterteilt:

- 1 - ledig, weiblicher Nutzer unter 25
- 2 - ledig, Hauptnutzer weiblich unter 25
- 3 - gesch"aftliche Nutzung
- 4 - unverheirateter Besitzer oder Hauptnutzer unter 25
- 5 - verheirateter Besitzer oder Hauptnutzer unter 25

Die dritte Variable *Insured* bezeichnet die verdienten Versicherungsjahre und dient hier als Offset, welcher im vorigen Kapitel beschrieben ist. Die jeweiligen verdienten Versicherungsjahre verwendet man somit für eine Gewichtung und entsprechen folglich der Variable t_i . Die Variablen *Premium* und *Cost*, die ebenso im Datensatz aufgeführt sind, lässt man unberücksichtigt. Die letzte aufgeführte Variable *Claims* beschreibt die Anzahl der Schäden. Diese Variable soll als Response dienen. Die Daten treten in aggregierter Form auf, d.h. alle Kovariablenkombinationen treten nur einmal auf, woraus sich eine Länge des Datensatzes von 20 ergibt. Damit wird die Schadenanzahl nicht bezüglich einzelner Individuen modelliert, sondern bezüglich einer Gruppe von Versicherungsnehmern mit identischen Merkmalen. Es kann allerdings vorkommen, dass die Größe der einzelnen Gruppen sehr unterschiedlich ist, weshalb man die Anzahl an verdienten Versicherungsjahren als Gewichtung verwendet, um die Gruppen der verschiedenen Merkmalskombinationen vergleichbar zu machen. Die Kombinationen werden als Beobachtungen bezeichnet, d.h. *Merit 3* mit *Class 1-5* bezeichnen die ersten 5 Beobachtungen, *Merit 2* mit *Class 1-5* die Beobachtungen 6 bis 10, *Merit 1* mit *Class 1-5* die Beobachtungen 11 bis 15 und *Merit 0* mit *Class 1-5* die Beobachtungen 16 bis 20.

Beobachtung 1: Merit 3/ Class 1	Beobachtung 11: Merit 1/ Class 1
Beobachtung 2: Merit 3/ Class 2	Beobachtung 12: Merit 1/ Class 2
Beobachtung 3: Merit 3/ Class 3	Beobachtung 13: Merit 1/ Class 3
Beobachtung 4: Merit 3/ Class 4	Beobachtung 14: Merit 1/ Class 4
Beobachtung 5: Merit 3/ Class 5	Beobachtung 15: Merit 1/ Class 5
Beobachtung 6: Merit 2/ Class 1	Beobachtung 16: Merit 0/ Class 1
Beobachtung 7: Merit 2/ Class 2	Beobachtung 17: Merit 0/ Class 2
Beobachtung 8: Merit 2/ Class 3	Beobachtung 18: Merit 0/ Class 3
Beobachtung 9: Merit 2/ Class 4	Beobachtung 19: Merit 0/ Class 4
Beobachtung 10: Merit 2/ Class 5	Beobachtung 20: Merit 0/ Class 5

8.2.2 Explorative Datenanalyse

Zuerst soll ein Modell angepasst werden, das die Zählvariable *Anzahl der Schäden* erklärt. Es wird zunächst die Poissonverteilung mit der Log-Link-Funktion zur Modellierung der Daten herangezogen. Also erhält man:

$$\begin{aligned}
 \text{Claims} &\sim \text{Poi}(\mu_i), \quad i = 1, \dots, 20, \\
 \log(\mu_i) &= \log(t_i) + \mathbf{x}_i^T \boldsymbol{\beta} \quad \text{bzw.} \quad \mu_i = t_i \exp(\mathbf{x}_i^T \boldsymbol{\beta}).
 \end{aligned}$$

Wenn man nun die jeweilige Kovariable gegen die empirischen Logits anträgt, sollten sich erkennbare Zusammenhänge ergeben. In der Abbildung 8.1 soll dies nun veranschaulicht werden.

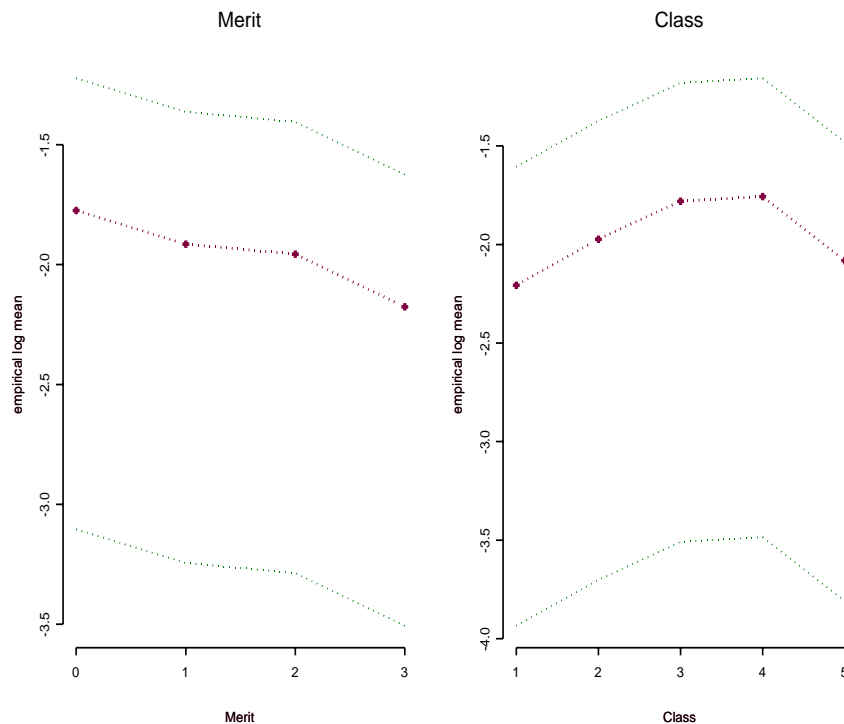


Abbildung 8.1: Explorative Datenanalyse des KFZ-Datensatzes

Die linke Grafik zeigt für die Kovariable *Merit* einen negativen linearen Zusammenhang, also nehmen mit zunehmender unfallfreien Fahrweise die Schäden ab. Dagegen sieht man für die Kovariable *Class* einen eher quadratischen Zusammenhang. Das Modell, das hier betrachtet wird, umfasst also die beiden Haupteffekte der Variablen *Merit* und *Class*. Da die Variablen einen linearen bzw. quadratischen Verlauf aufweisen, besteht die Möglichkeit, diese Variablen nicht kategoriell, sondern metrisch zu betrachten. Da dieses Modell aber eine wesentlich schlechtere Anpassung liefert, was an der großen Anzahl der Beobachtungen pro Zelle liegt, behandelt man die Kovariablen als kategorielle Variablen.

8.2.3 Poissonregression

Mit den entsprechenden Haupteffekten kann man nun ein Poissonmodell in SPlus eingeben, dass wie folgt lautet:

```
cancarpoi_glm(Claims~offset(log(Insured))+Merit+Class,family=poisson,
              link=log,x=T)
```

Mit Hilfe der *summary*-Funktion bekommt man Informationen über die geschätzten Regressionskoeffizienten $\hat{\beta}_j$, $j = 1, \dots, 8$, die in der Spalte *value* angegeben sind.

Der Tabelle 8.1 lässt sich entnehmen, dass alle Kovariablen hoch signifikant sind, da der *t value* für jede Kovariable betragsmäßig viel größer als 2 ist, d.h. die Hypothesen $H : \beta_j = 0$, $j = 1, \dots, 8$ können zu einem kleinen Signifikanzniveau verworfen werden. Jedoch zeigt der Wert der Devianz

```
Call: glm(formula = Claims ~ offset(log(Insured)) + Merit + Class,
family = poisson, x = T, link = log)
```

```
Deviance Residuals:
```

```
      Min       1Q   Median       3Q      Max
-10.79274 -3.007873 -1.575749  2.426679 11.62523
```

```
Coefficients:
```

	Value	Std. Error	t value
(Intercept)	-2.0357359	0.004311305	-472.18556
Merit1	-0.1377590	0.007172219	-19.20730
Merit2	-0.2206796	0.007997189	-27.59465
Merit3	-0.4929506	0.004502371	-109.48689
Class2	0.2998302	0.007258049	41.31003
Class3	0.4690550	0.005039141	93.08233
Class4	0.5258551	0.005364533	98.02439
Class5	0.2155504	0.010734511	20.08013

```
(Dispersion Parameter for Poisson family taken to be 1 )
```

```
Null Deviance: 33854.16 on 19 degrees of freedom
```

```
Residual Deviance: 579.5163 on 12 degrees of freedom
```

```
Number of Fisher Scoring Iterations: 3
```

Tabelle 8.1: *summary* des Poissonmodells für den KFZ-Datensatz

mit 579.52 einen deutlich höheren Wert als die Anzahl der Freiheitsgrade mit 12, was auf eine eher schlechte Anpassung der Daten durch das Poissonmodell schließen lässt.

Nun können ebenso wie für den Patentdatensatz die Devianz- und die Pearson-Residuen berechnet werden. In der Abbildung 8.2 werden die Devianz- und Pearson-Residuen grafisch dargestellt. In den oberen beiden Grafiken lässt sich erkennen, dass die Residuen zufällig um Null streuen, jedoch teilweise sehr hohe Werte annehmen. Die höchsten Werte der Devianz- und der Pearson-Residuen liegen über dem Wert von 10. Die Beobachtungen, die sich um diese Werte bewegen, sind Beobachtung 4, 16 und 19. Die unteren beiden Grafiken zeigen die Quantile der Standardnormalverteilung gegenüber den beiden Residuen. Man sieht, dass sich die Residuen nicht nach der Standardnormalverteilung verhalten, da sie deutlich von der durchgezogenen Linie, der Winkelhalbierenden abweichen.

Mit Hilfe der Funktion *diag.poi* kann man sich die Diagnostiken der Hatmatrix, der $DFFITs_i^1$ und der Cook's Distanz D_i^1 ansehen. In der Tabelle 8.2 sind die Werte der Diagnostiken zu sehen. In der Abbildung 8.3 sind die h_{ii} -Werte mittels eines Stabdiagramms dargestellt.

Die Abbildungen 8.4 und 8.5 zeigen die Werte jeweils für die Ein-Schritt-Approximation und für den exakten Wert durch die jeweiligen Stabdiagramme. Die unteren beiden Grafiken weisen zur besseren Übersicht eine skalierte y-Achse auf.

In der Abbildung 8.3 sind die Beobachtungen, die jeweils die kritische Schranke von $2p/n = 16/20 = 0.8$ überschreiten, angegeben. In diesem Fall überschreiten Beobachtung 1 und 5 den Wert 0.8, d.h. die Gruppen, die seit drei oder mehr Jahren unfallfrei fahren (*Merit 3*) und der

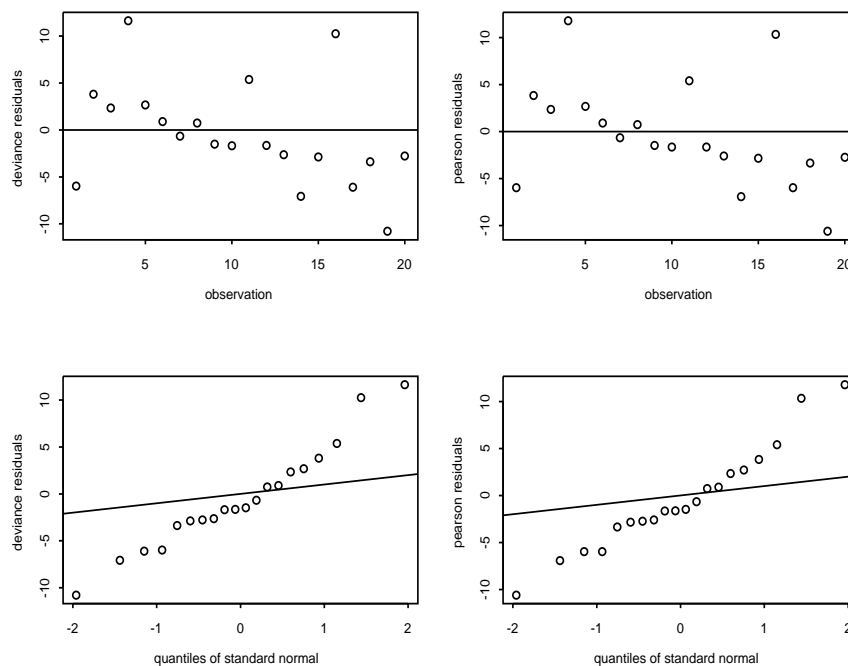


Abbildung 8.2: Devianz- und Pearson-Residuen des Poissonmodells für den KFZ-Datensatz (oben), Plot der Devianz- und Pearson-Residuen gegen die Standardnormalverteilungsquantile (unten). Die durchgezogene Linie stellt die Winkelhalbierende dar.

Nutzer ledig, weiblich und unter 25 (*Class 1*) bzw. der Besitzer verheiratet oder der Hauptnutzer unter 25 ist (*Class 5*). Wegen der Überschreitung der Schranke bezeichnet man diese Gruppen als *x*-Ausreißer. Bereits im theoretischen Teil wurde darauf hingewiesen, dass es bei $p/n > 0.4$ sehr wenige Freiheitsgrade gibt. Deshalb sind Rückschlüsse auch immer unter Berücksichtigung von p und n zu ziehen. In der Abbildung 8.4 sieht man, dass die Beobachtung 1 den höchsten Wert für $DFFITs_i^1$ aufweist. Dieser liegt absolut bei 107.2, wobei der Wert für den exakten Wert bei 88.2 liegt. Auch von den Beobachtungen 2, 3, 4, 5, 6, 11, 14, 16, 17, 18 und 19 wird die Schranke von 2 überschritten. Wenige Werte der Beobachtungen bleiben also unter dem Wert 2 und noch weniger unter der absoluten Schranke $2\sqrt{p/n} = 1.26$. Die Schätzung für die Ein-Schritt-Approximation ist für alle Beobachtungen außer der Beobachtung 1 gut, wie aus der Grafik ersichtlich ist. Aufgrund des niedrigen Wertes für n ergeben sich hier über die Hälfte der Beobachtungen als einflußreich. Man erkennt, dass die Kombinationen von *Merit 3* mit *Class 1-5*, entspricht Beobachtung 1 bis 5, und die Kombinationen von *Class 1* mit *Merit 0-3*, entspricht Beobachtungen 1, 6, 11 und 16, am einflußreichsten sind. Wenn man sich den Datensatz hier genauer ansieht, stellt man fest, dass die entsprechenden Zahlen dieser Beobachtungen in der Variable *Claims* bzw. *Insured* hohe Werte im Vergleich zu den anderen Beobachtungen aufweisen und so den Hauptanteil ausmachen. Die Abbildung 8.5 zeigt die Werte der Cook's Distanz D_i^1 mittels eines Stabdiagramms. Für die Beobachtung 1 ergibt sich der höchste Wert (1512.3). Der exakte Wert ist deutlich geringer, wobei die anderen Werte bis auf die Beobachtung 6 gut approximiert werden. Die Schranke von 1 wird noch von den Werten D_i^1 der Beobachtungen 2, 3, 4, 5, 6, 11, 14, 16, 17 und 19 überschritten. Dies sind hauptsächlich die bereits vorher

\$hat:

```
[1] 0.9471908 0.7681749 0.7492676 0.5832462 0.8234808 0.7580269 0.2160187
[8] 0.2068278 0.2224893 0.2280046 0.7453042 0.2043405 0.2192143 0.2348162
[15] 0.2186635 0.6661830 0.2485535 0.2864491 0.4489382 0.2248099
```

\$dffits:

```
[1] -107.1689641 13.3354739 7.8817959 21.5499571 12.0314996
[6] 2.8295896 -0.1766662 0.3671216 -0.8637527 -0.3663605
[11] 16.7390195 -0.5566489 -1.4191856 -4.2205545 -0.7666600
[16] 25.2661840 -3.7680645 -2.4403133 -12.2670766 -1.3068336
```

\$cook:

```
[1] 1512.27633093 26.17369054 8.21994554 58.23962029 23.90977173
[6] 1.27352750 0.01952230 0.02313873 0.10203115 0.12974835
[11] 42.04462350 0.10900965 0.30518237 2.40926513 0.35916735
[16] 80.02089323 1.97545964 0.79048951 20.83383231 0.35141381
```

Tabelle 8.2: h_{ii} -Werte, $DFFITs_i^1$ und Cook's Distanz D_i^1 des Poissonmodells für den KFZ-Datensatz

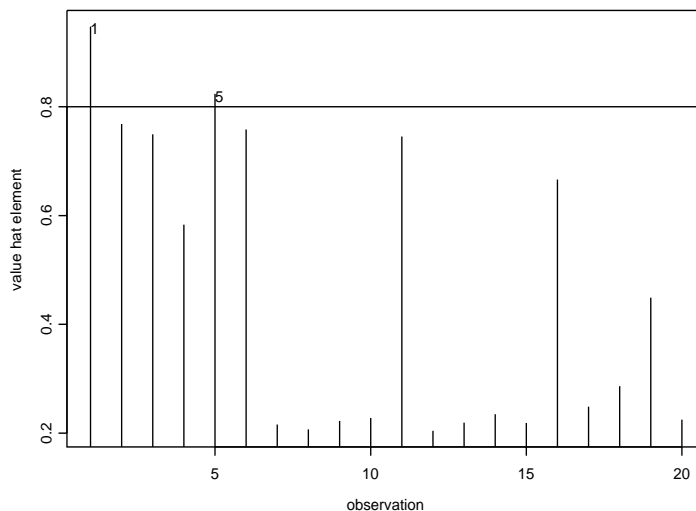


Abbildung 8.3: Stabdiagramm der h_{ii} -Werte des Poissonmodells für den KFZ-Datensatz ohne Skalierung (oben) und mit Skalierung ($|\cdot| \leq 5$) der y-Achse (unten). Die durchgezogene Linie markiert die Schranke $2p/n = 16/20 = 0.8$.

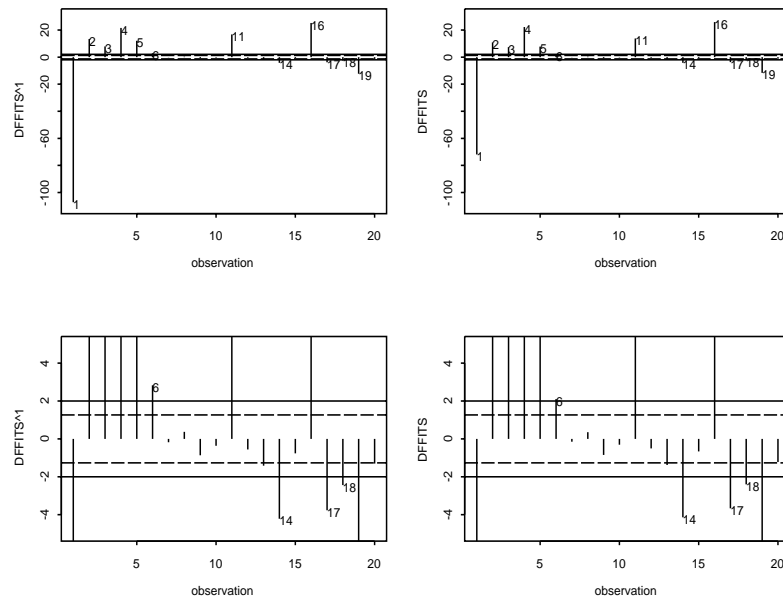


Abbildung 8.4: Stabdiagramm der $DFFITS_i^1$ (links) und der $DFFITS_i$ (rechts) des Poissonmodells für den KFZ-Datensatz ohne Skalierung (oben) und mit Skalierung ($|\cdot| \leq 5$) der y-Achse (unten). Die gestrichelte bzw. durchgezogene Linie markiert die absoluten Schranken $2\sqrt{p/n} = 2\sqrt{8/20} = 1.26$ und 2.

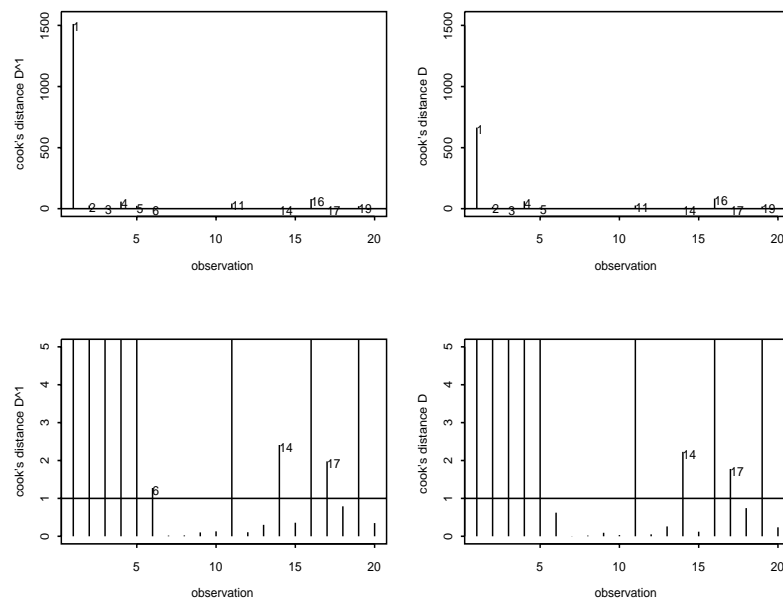


Abbildung 8.5: Stabdiagramm der Cook's Distanz D_i^1 (links) und D_i (rechts) des Poissonmodells für den KFZ-Datensatz ohne Skalierung (oben) und mit Skalierung ($|\cdot| \leq 5$) der y-Achse (unten). Die durchgezogene Linie markiert die Schranke 1.

angesprochenen Beobachtungen.

Nun kann die Funktion *dfbetas.poi* angewendet werden, um die Veränderungen in den Regressionskoeffizienten zu sehen. In den Abbildungen 8.6 und 8.7 sieht man für die einzelnen Kovariablen jeweils die Veränderung nach Löschen der i -ten Beobachtung.

Die Grafiken zeigen jeweils die Werte für die Ein-Schritt-Approximation und die exakten Werte, wobei die Ein-Schritt-Approximationen hier für alle bis auf wenige Ausnahmen recht gut sind. Lediglich die Beobachtung 1 zeigt teilweise größere Unterschiede. Die Schranke von 2 wird von der Beobachtung 16 (Kombination von *Merit 0* (alle anderen) und *Class 1* (Nutzer ledig, weiblich und unter 25)) in allen Kovariablen überschritten. Desweiteren zeigen auch die Beobachtungen 1 und 19 in den meisten Variablen der $DFBETAS_{j,i}$ signifikante Änderungen. Vereinzelt Überschreitungen der Schranke 2 gibt es für die Beobachtungen 2, 3, 4, 6, 11, 14 und 17. Diese Beobachtungen stellen aufgrund der Höhe ihrer Werte für die $DFBETAS_{j,i}$ die einflußreichsten Beobachtungen dar. Hier bleibt zu bemerken, dass in diesem Datensatz eine einzelne Beobachtung aufgrund des niedrigen Wertes für n viel mehr Gewicht hat wie in einem Datensatz, wo der Wert für n größer ist.

Um den $COVRATIO_i$ auszurechnen, wird die Funktion *covratio.poi* verwendet. In Tabelle 8.3 sieht man die geordneten Werte und die Indizes der Beobachtungen und in Abbildung 8.8 die Werte anhand eines Stabdiagramms.

Man sieht die Werte für die Ein-Schritt-Approximation und die exakten Werte. Die Ein-Schritt-Approximation überschätzt für die höchsten Werte den $COVRATIO_i$ deutlich. Für alle Werte ergeben sich sehr hohe Werte, so dass eigentlich alle Beobachtungen als sehr einflußreich deklariert werden können, was aber auch wieder an dem niedrigen Wert von n liegt.

Abschließend wird die Funktion *devred.poi* angewandt, um die Devianzreduktion zu messen. In Tabelle 8.4 sieht man die geordneten Werte mit den Beobachtungsindizes und in Abbildung 8.9 anhand eines Stabdiagramms.

Die Abbildung 8.9 zeigt, dass alle Beobachtungen bis auf wenige Ausnahmen (6 bis 10 (Kombination von *Merit 2* und *Class 1-5*), 12 (Kombination von *Merit 1* und *Class 2*) die Schranke von 4 überschreiten. Der höchste Wert (674.7) ergibt sich für die Beobachtung 1 (Kombination von *Merit 3* und *Class 1*), wobei der exakte Wert bei 453.3 liegt. Für die weiteren Beobachtungen ergeben sich teilweise große Veränderungen in der Devianz. Für die Beobachtungen 4 (Kombination von *Merit 3* und *Class 4*) und 16 (Kombination von *Merit 0* und *Class 1*) wird die Devianz nach Weglassen dieser Beobachtungen fast halbiert. Die Approximationen für diese und die weiteren Beobachtungen sind sehr gut.

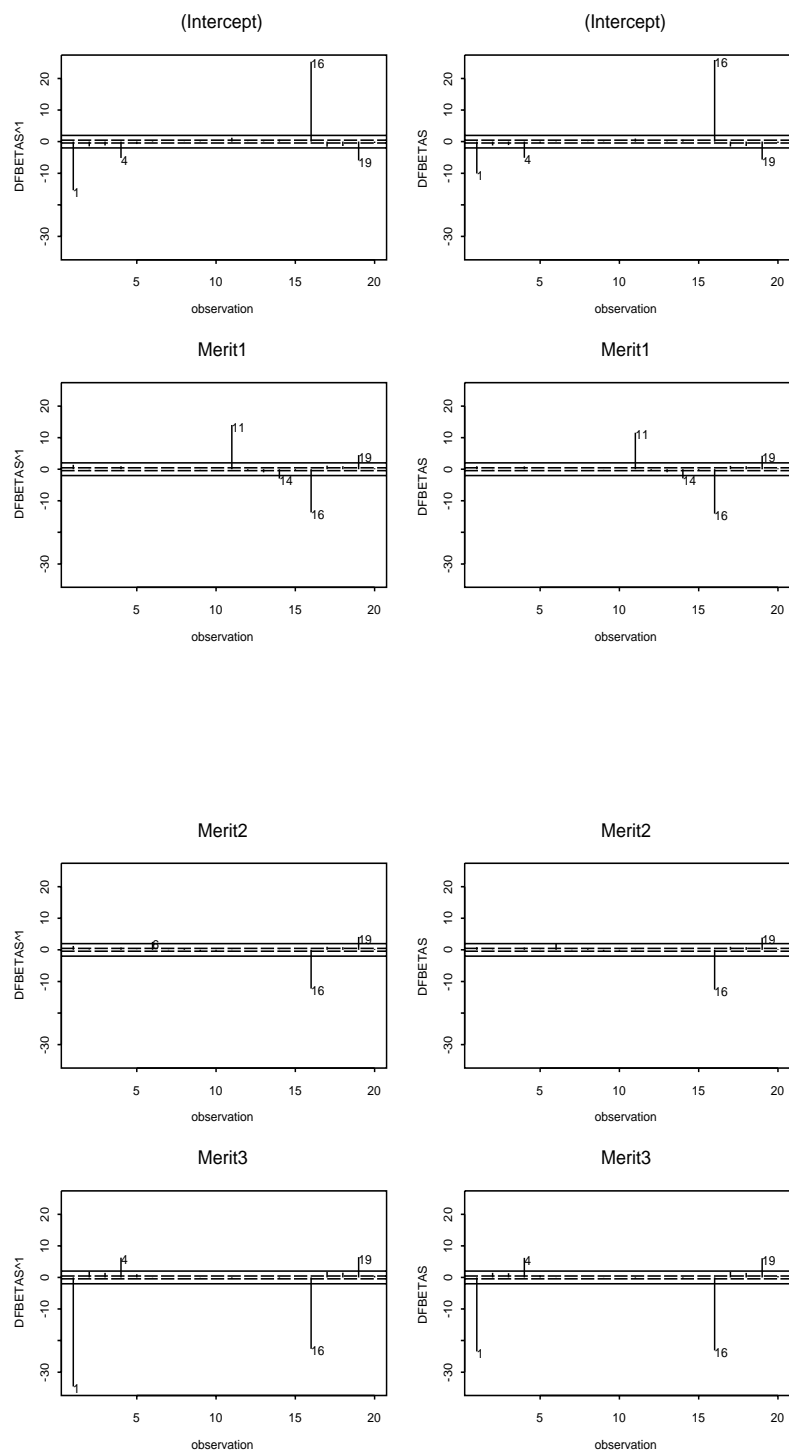


Abbildung 8.6: Stabdiagramm der $DFBETAS^1_{j,i}$ (links) und $DFBETAS_{j,i}$ (rechts) des Poissonmodells für den KFZ-Datensatz. Die gestrichelte bzw. durchgezogene Linie markiert die absoluten Schranken $2\sqrt{1/n} = 2\sqrt{1/20} = 0.45$ und 2.

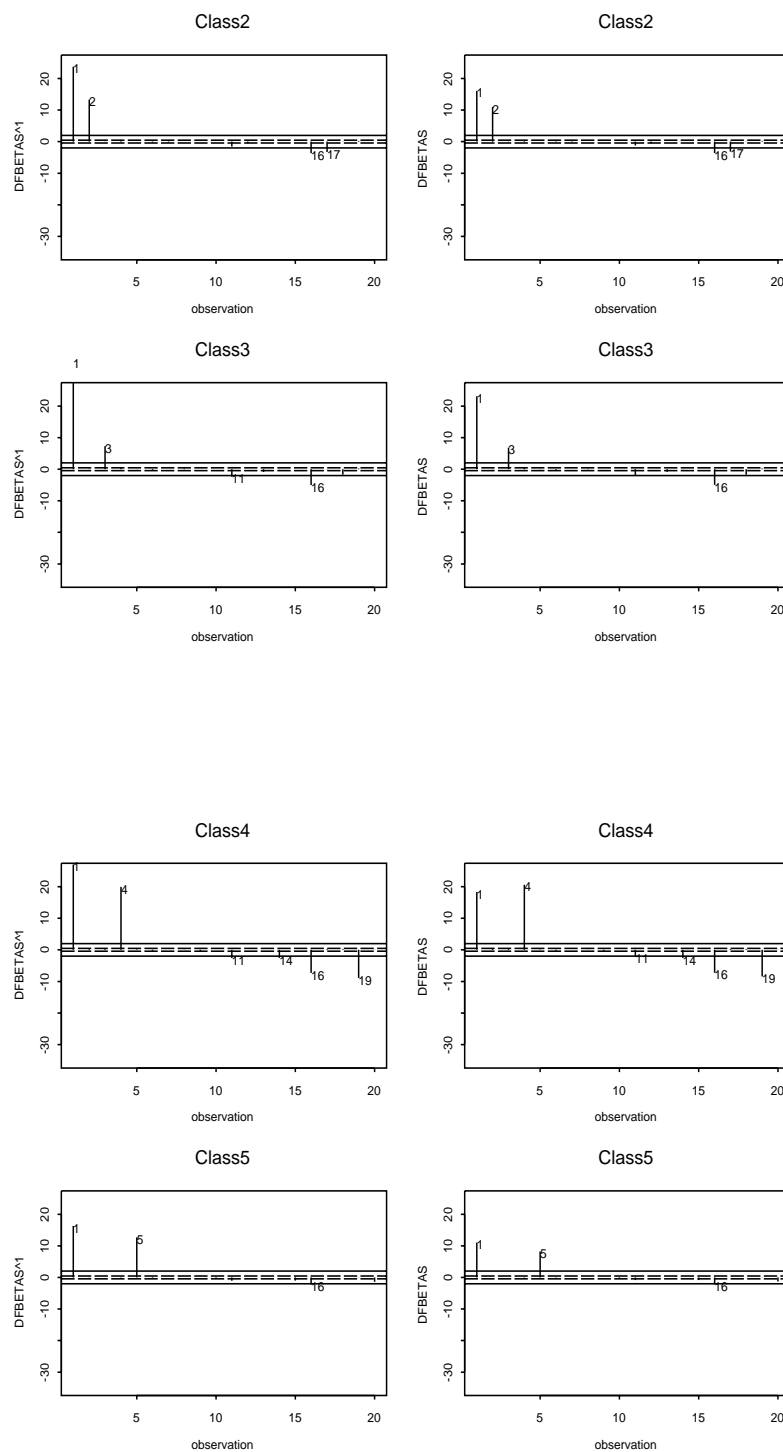


Abbildung 8.7: Stabdiagramm der $DFBETAS^1_{j,i}$ (links) und $DFBETAS_{j,i}$ (rechts) des Poissonmodells für den KFZ-Datensatz. Die gestrichelte bzw. durchgezogene Linie markiert die absoluten Schranken $2\sqrt{1/n} = 2\sqrt{1/20} = 0.45$ und 2.

\$covratio.ord:

[1]	882773.8767	67380.1039	55348.9159	41611.2174	38807.3078
[6]	29317.1927	24346.4754	18021.6716	15259.6423	12724.8280
[11]	6221.4314	5988.3411	5854.3046	5120.2577	4470.7248
[16]	2443.1506	2220.0395	1690.7642	1100.4720	854.6249

\$beob:

[1]	1	16	3	11	4	2	6	19	5	18	14	13	17	9	8	12	20	7	15	10
-----	---	----	---	----	---	---	---	----	---	----	----	----	----	---	---	----	----	---	----	----

Tabelle 8.3: Geordnete Werte des $COVRATIO_i^1$ des Poissonmodells für den KFZ-Datensatz

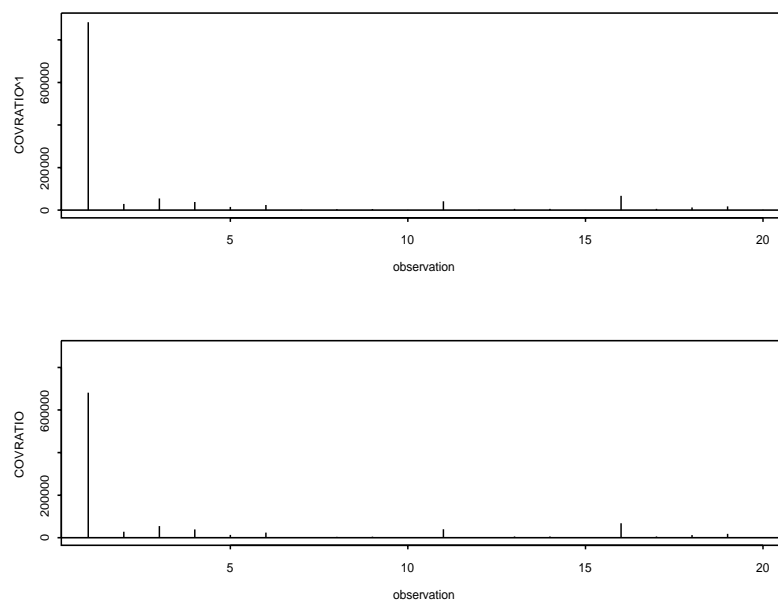


Abbildung 8.8: Stabdiagramm des $COVRATIO_i^1$ (oben) und $COVRATIO_i$ (unten) des Poissonmodells für den KFZ-Datensatz

\$devred.ord:

```
[1] 674.6700417 329.3185606 318.8797518 208.3289784 114.5650750
[6] 64.6525042 63.0358398 49.0040966 40.9217779 21.9813999
[11] 15.8974521 10.5727381 9.8850151 8.7948836 3.5819723
[16] 3.4348345 3.2502446 2.8723987 0.7071722 0.5699280
```

\$beob:

```
[1] 1 4 16 19 11 14 2 17 5 3 18 15 20 13 10 12 6 9 8 7
```

Tabelle 8.4: Geordnete Werte der Devianzreduktion ΔD_i^1 des Poissonmodells für den KFZ-Datensatz

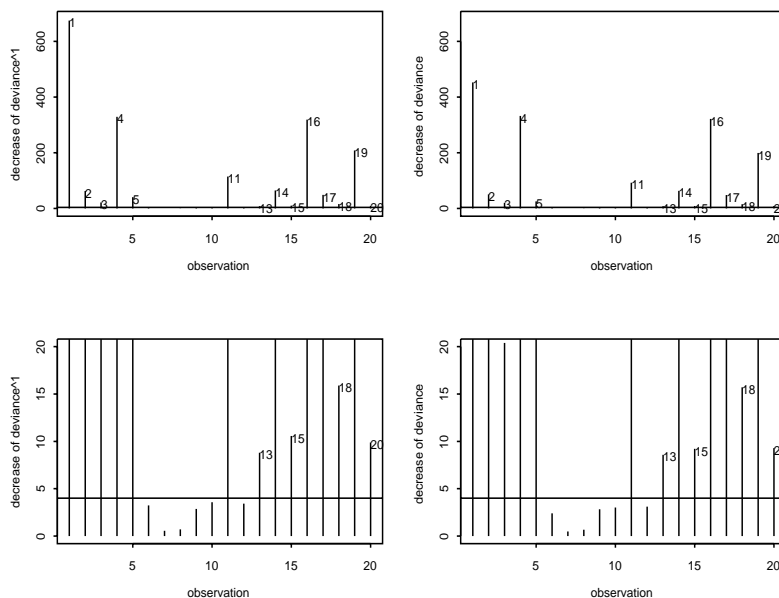


Abbildung 8.9: Stabdiagramm der Werte der Devianzreduktion ΔD_i^1 (links) und ΔD_i (rechts) des Poissonmodells für den KFZ-Datensatz ohne Skalierung (oben) und mit Skalierung ($|\cdot| \leq 20$) der y-Achse (unten). Die durchgezogene Linie markiert die Schranke 4.

8.3 Negativ-Binomial-Regression mit bekanntem α

Nun will man ein Modell mit Hilfe der Negativ-Binomial-Verteilung mit bekanntem α anpassen. Dazu wird wieder das Programm *negbin* von B. Venables verwendet. Um einen angemessenen Wertebereich für α bzw. θ zu bekommen, nimmt man wiederum den Signal-to-Noise-Quotient zu Hilfe. Für den Fall der Negativ-Binomial-Verteilung ergibt sich

$$SN_{NB}(\alpha, \mu_i) = \frac{E(Y_i)}{\sqrt{Var(Y_i)}} = \sqrt{\frac{\mu_i}{1 + \alpha\mu_i}}$$

Da der Mittelwert \bar{y} für die KFZ-Daten bei 20199.95 liegt, wird auch für den Signal-to-Noise-Quotienten ein entsprechend höherer Wert benutzt. Für diese Zwecke hier wird ein Wert von 18 festgelegt. Also ergibt sich eine obere Schranke für α aus

$$\sqrt{\frac{\bar{y}}{1 + \alpha\bar{y}}} = 18 \Leftrightarrow \alpha \approx 0.003$$

Somit ergibt sich für θ ein Wert von $1/0.003 = 333.3$. Ausgehend von $\alpha = 0$ bis $\alpha = 0.003$ wird dieser Bereich in äquidistante Intervalle der Länge 0.0005 zerlegt. Es ergeben sich sieben Werte und damit sieben verschiedene Modelle. Für $\alpha = 0$ wird wieder der Wert $\alpha = 10^{-9}$ verwendet. Es ergeben sich also folgende Werte:

$$\begin{aligned} \alpha &= (10^{-9}, 0.0005, 0.001, 0.0015, 0.002, 0.0025, 0.003) \\ \theta &= (10^9, 2000, 1000, 666.67, 500, 400, 333.33) \end{aligned}$$

Nun können die sieben Modelle in SPlus eingegeben werden. Es gelten dieselben Effekte wie im Poissonfall.

```
cancarnba1_glm(Claims~offset(log(Insured))+Merit+Class,family=negative.binomial
               (theta=theta[1]),x=T)
cancarnba2_glm(Claims~offset(log(Insured))+Merit+Class,family=negative.binomial
               (theta=theta[2]),x=T)
cancarnba3_glm(Claims~offset(log(Insured))+Merit+Class,family=negative.binomial
               (theta=theta[3]),x=T)
cancarnba4_glm(Claims~offset(log(Insured))+Merit+Class,family=negative.binomial
               (theta=theta[4]),x=T)
cancarnba5_glm(Claims~offset(log(Insured))+Merit+Class,family=negative.binomial
               (theta=theta[5]),x=T)
cancarnba6_glm(Claims~offset(log(Insured))+Merit+Class,family=negative.binomial
               (theta=theta[6]),x=T)
cancarnba7_glm(Claims~offset(log(Insured))+Merit+Class,family=negative.binomial
               (theta=theta[7]),x=T)
```

Mit Hilfe der in Anhang E definierten Funktionen kann nun der Datensatz auf Ausreißer und einflußreiche Beobachtungen überprüft werden. Mittels der Funktion *diag.nba* können die h_{ii} -Werte, die $DFFITs_i^1$ und die Cook's Distanz D_i^1 berechnet werden. Die folgenden Abbildungen 8.10, 8.11 und 8.12 zeigen die h_{ii} -Werte, $DFFITs_i^1$ und die Cook's Distanz D_i^1 für verschiedene Werte von α . Für die $DFFITs_i$ und die Cook's Distanz D_i wird jeweils die Approximation und der exakte Wert angezeigt. Wegen der zum Teil hohen Werte werden die z-Achsen in den Grafiken skaliert, was in den unteren beiden Grafiken jeweils sichtbar wird.

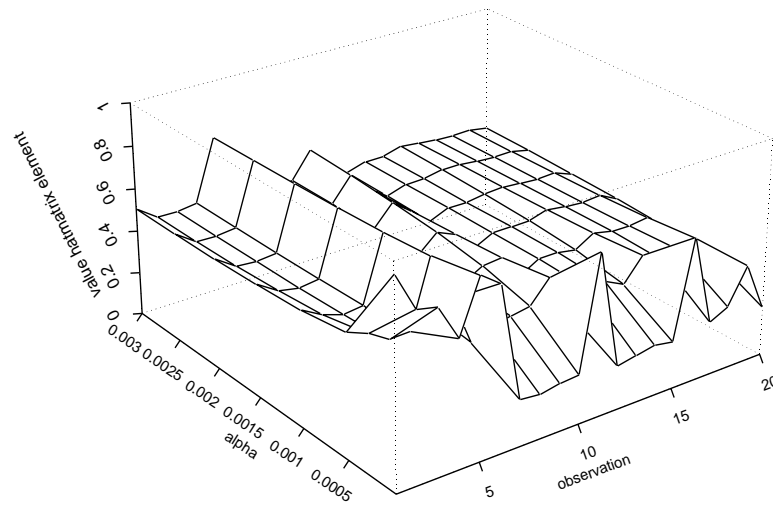


Abbildung 8.10: h_{ii} -Werte des Negativ-Binomial-Modells mit festem α für den KFZ-Datensatz für verschiedene Werte von α

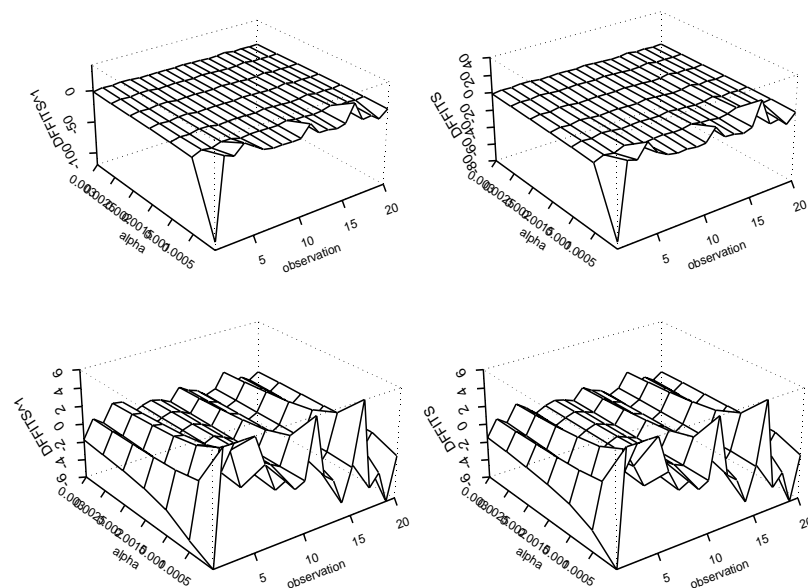


Abbildung 8.11: $DFITS_i^1$ (links) und $DFITS_i$ (rechts) des Negativ-Binomial-Modells mit festem α für den KFZ-Datensatz für verschiedene Werte von α ohne Skalierung (oben) und mit Skalierung ($|\cdot| \leq 6$) der z-Achse (unten)

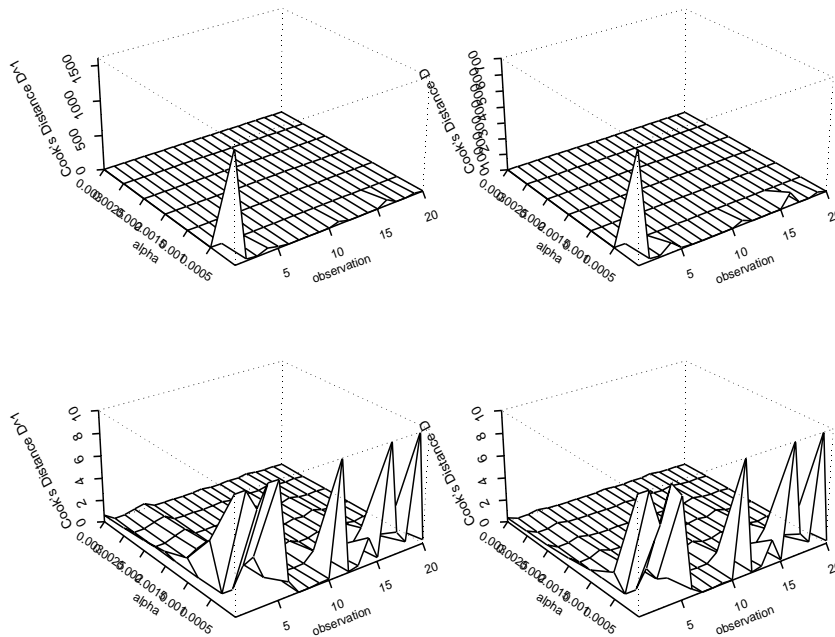


Abbildung 8.12: Cook's Distanz D_i^1 (links) und D_i (rechts) des Negativ-Binomial-Modells mit festem α für den KFZ-Datensatz für verschiedene Werte von α ohne Skalierung (oben) und mit Skalierung ($|\cdot| \leq 10$) der z-Achse (unten)

In Abbildung 8.10 kann man erkennen, dass sich die h_{ii} -Werte für die verschiedenen Werte von α ändern. Der Wert für die Beobachtungen 1, 2, 3, 6, 11 und 16 nimmt schon für den α -Wert 0.005 deutlich ab und bleibt dann auf einem konstantem Niveau für die weiteren Werte von α . Dagegen nehmen die h_{ii} -Werte der Beobachtungen 7, 8, 9, 10, 12, 13, 14, 15, 17, 18 und 20 zu und erreichen bei einem Wert von $\alpha = 0.002$ einen konstanten Wert. Für die Beobachtungen 5 und 19 ändern sich die Werte kaum. Anzumerken bleibt lediglich, dass der Wert der Beobachtung 5 ebenso wie der Wert der Beobachtung 1 unter die Schranke 0.8 fällt, so dass dieser Wert nicht mehr überschritten wird.

Für die $DFFITs_i^1$ zeigt Abbildung 8.11, dass die Werte absolut gesehen abnehmen, am deutlichsten ist es an der Beobachtung 1 zu sehen, deren absoluter Wert mit zunehmendem α abnimmt. Auch die absoluten Werte für die Beobachtungen 2, 3, 4, 11, 14, 16 und 19 nehmen für $\alpha = 0.005$ schon deutlich ab. Der Einfluß wird also mit zunehmendem α geringer. Die Ein-Schritt-Approximationen sind bis auf den hohen Wert von $DFFITs_1^1$ für $\alpha = 0$ akzeptabel.

Für die Cook's Distanz D_i^1 in Abbildung 8.12 verhalten sich die Werte ähnlich. Die Werte der Beobachtungen 1 bis 5, 11, 16 und 19 nehmen für einen α -Wert von 0.0005 schon deutlich ab. Mit zunehmendem α geht dieser Trend weiter.

Mit der Funktion *covratio.nba* kann der $COVRATIO_i^1$ berechnet werden und mittels eines Stabdiagramms dargestellt werden. In der Abbildung 8.13 sind die $COVRATIO_i^1$ für verschiedene Werte von α zu sehen. Es werden ebenso die exakten Werte aufgezeigt und in den unteren beiden Grafiken, welche eine z-Achsenkalierung beinhalten, sieht man die niedrigeren Werte besser.

Aus der Grafik wird sichtbar, dass die Werte des $COVRATIO_i^1$ mit steigendem α deutlich sinken, also in einem Negativ-Binomial-Modell für den KFZ-Datensatz einzelne Beobachtungen weitaus weniger Einfluß haben, als dies im Poissonmodell der Fall war. Die Approximationen werden für die niedrigeren Werte relativ gesehen wieder besser und sind annehmbar.

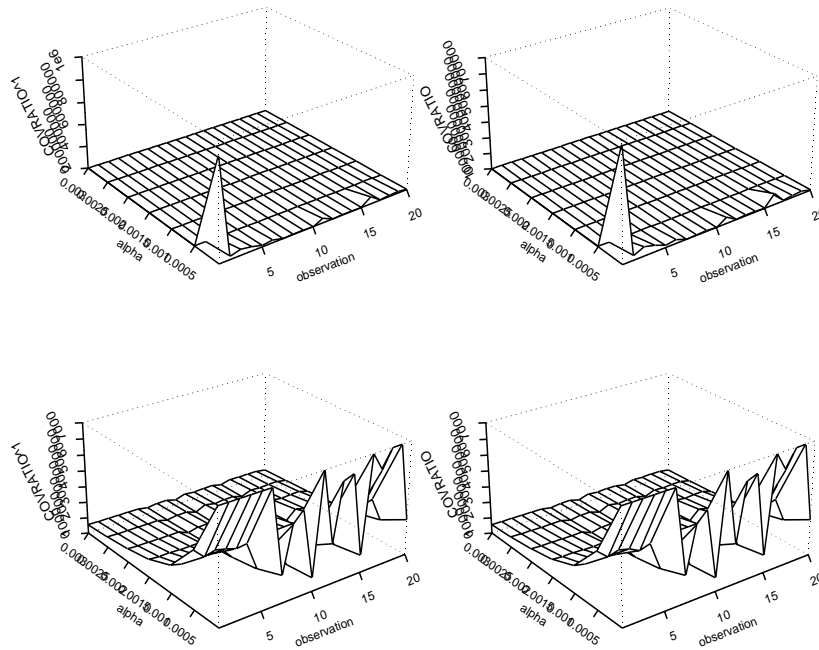


Abbildung 8.13: $COVRATIO_i^1$ (links) und $COVRATIO_i$ (rechts) des Negativ-Binomial-Modells mit festem α für den KFZ-Datensatz für verschiedene Werte von α ohne Skalierung (oben) und mit Skalierung ($|\cdot| \leq 7000$) der z-Achse (unten)

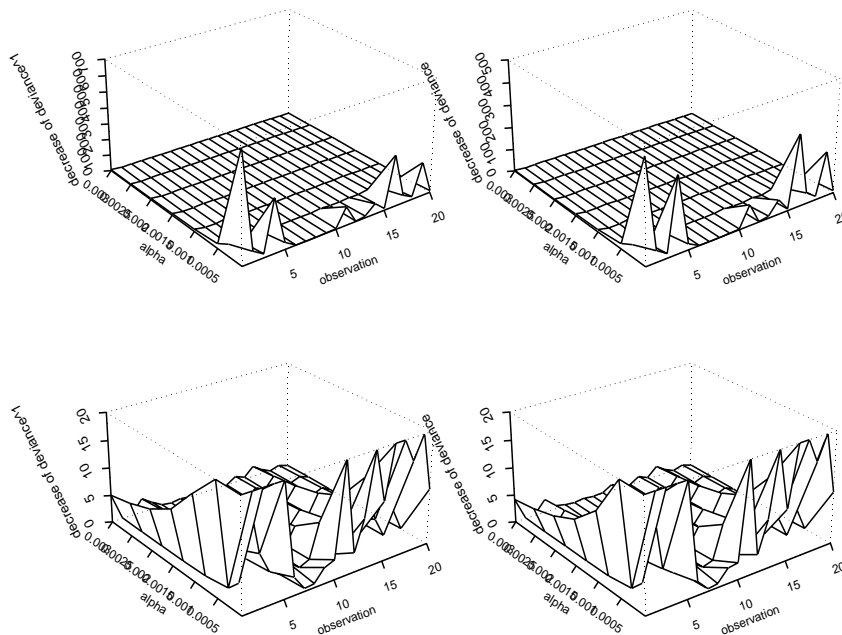


Abbildung 8.14: Werte der Devianzreduktion ΔD_i^1 (links) und ΔD_i (rechts) des Negativ-Binomial-Modells mit festem α für den KFZ-Datensatz für verschiedene Werte von α ohne Skalierung (oben) und mit Skalierung ($|\cdot| \leq 20$) der z-Achse (unten)

Die Funktion *devred.nba* berechnet nun die Devianzreduktion. In der Abbildung 8.14 sind die Werte der Devianzreduktion für die verschiedenen Werte von α zu sehen, zum einen die Approximationen und zum anderen die exakten Werte. Die unteren Grafiken sind wieder skaliert. Die Werte für die Devianzreduktion reduzieren sich auch hier für zunehmendes α , deutlich sichtbar für die Beobachtungen 1, 4, 16 und 19. Die Approximationen für die Devianzreduktion sind gut. Lediglich die Devianzreduktion ΔD_i der Beobachtung 6 für $\alpha = 0$ überschätzt den wahren Wert deutlich.

In den folgenden Tabellen 8.5, 8.6, 8.7 und 8.8, werden nochmals die auffälligsten Beobachtungen für die h_{ii} -Werte, die $DFFITS_i^1$, die Cook's Distanz D_i^1 und die Devianzreduktion ΔD_i^1 für die verschiedenen Werte von α aufgelistet. Wie man erkennt, nimmt der Einfluß der einzelnen Beobachtungen deutlich ab, je höher der Wert α wird.

h_{ii} -Werte	
α	Beobachtungen
0	1, 5
0.0005	-
0.001	-
0.0015	-
0.002	-
0.0025	-
0.003	-

Tabelle 8.5: Beobachtungen mit $h_{ii} \geq 2p/n$ des Negativ-Binomial-Modells mit festem α für den KFZ-Datensatz

$DFFITS_i$		
α	Beobachtungen für Ein-Schritt-Approximation und exakten Wert	
0	1, 16, 4, 11, 2, 19, 5, 3, 14, 17, 6, 18	1, 16, 4, 11, 2, 19, 5, 3, 14, 17, 18, 6
0.0005	1, 4, 16, 5, 11	1, 4, 16, 5, 11
0.001	1, 4, 16, 5	1, 4, 16
0.0015	1,4	1
0.002	1	-
0.0025	-	-
0.003	-	-

Tabelle 8.6: Beobachtungen mit $DFFITS_i^1 \geq 2$ bzw. $DFFITS_i \geq 2$ des Negativ-Binomial-Modells mit festem α für den KFZ-Datensatz

Cook's Distanz		
α	Beobachtungen für Ein-Schritt-Approximation und exakten Wert	
0	1, 16, 4, 11, 2, 5, 19, 3, 14, 17, 6	1, 16, 4, 11, 2, 5, 19, 3, 14, 17
0.0005	1, 5, 4, 16	1, 4, 16
0.001	1, 5	1
0.0015	1	-
0.002	-	-
0.0025	-	-
0.003	-	-

Tabelle 8.7: Beobachtungen mit $D_i^1 > 1$ bzw. $D_i > 1$ des Negativ-Binomial-Modells mit festem α für den KFZ-Datensatz

Devianzreduktion		
α	Beobachtungen für Ein-Schritt-Approximation und exakten Wert	
0	1, 4, 16, 19, 11, 14 (64.7)	1, 4, 16, 19, 11, 14 (63.8)
0.0005	1, 4, 16, 11, 14, 5 (5.2)	1, 4, 16, 11, 14, 19 (4.1)
0.001	1, 4, 16	1, 4, 16
0.0015	1, 4, 16	1, 4, 16
0.002	1	1
0.0025	1	1
0.003	1	1

Tabelle 8.8: Beobachtungen mit $\Delta D_i^1 \geq 4$ bzw. $\Delta D_i \geq 4$ des Negativ-Binomial-Modells mit festem α für den KFZ-Datensatz. Es sind bis $\alpha = 0.25$ jeweils die 6 höchsten Werte aufgeführt. Die Zahl in runden Klammern gibt jeweils die Devianzreduktion der 6-höchsten Beobachtung an.

8.4 Negativ-Binomial-Regression mit geschätztem α

In diesem Kapitel sollen die KFZ-Daten mit Hilfe der Negativ-Binomial-Verteilung mit unbekanntem α modelliert werden. Hierzu wird wiederum das Programm *negbin* von B. Venables benötigt. Das Modell, welches in SPlus eingegeben wird, lautet wie folgt:

```
cancarnb_glm.nb(Claims~offset(log(Insured))+Merit+Class,link=log,x=T)
```

Mit Hilfe der *summary*-Funktion erhält man Informationen über die geschätzten Regressionskoeffizienten $\hat{\beta}_j$, $j = 1, \dots, 8$, und dem geschätztem $\hat{\alpha}$ bzw. $\hat{\theta} = 1/\hat{\alpha}$.

```
Call: glm.nb(formula = Claims ~ offset(log(Insured)) + Merit + Class, x = T,
link = log, init.theta = 708.212336442048)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.295403	-0.6352415	0.06536695	0.4684358	1.70963

Coefficients:

	Value	Std. Error	t value
(Intercept)	-2.0406606	0.02473653	-82.495823
Merit1	-0.1383142	0.02649014	-5.221347
Merit2	-0.1882696	0.02695719	-6.984020
Merit3	-0.4131854	0.02495160	-16.559474
Class2	0.2421677	0.02881989	8.402800
Class3	0.4312753	0.02765678	15.593837
Class4	0.4642906	0.02755253	16.851105
Class5	0.1430389	0.03082644	4.640137

(Dispersion Parameter for Negative Binomial family taken to be 1)

Null Deviance: 671.294 on 19 degrees of freedom

Residual Deviance: 17.67152 on 12 degrees of freedom

Number of Fisher Scoring Iterations: 1

Theta:	708.21234
Std. Err.:	243.26643

2 x log-likelihood: 8163155.1053

Tabelle 8.9: *summary* des Negativ-Binomial-Modells mit geschätztem α für den KFZ-Datensatz

Aus der Tabelle 8.4 kann man erkennen, dass die Kovariablen einen deutlichen Einfluß haben, da der Wert für den *t value* oberhalb von 2 liegt. Der Wert der Devianz hat mit 17.67 einen deutlich geringeren Wert, als dies im Poissonmodell (579.52) der Fall war. Das Verhältnis von Devianz und der Anzahl der Freiheitsgrade mit 12 ist nun deutlich besser, was auf eine bessere Anpassung der Daten durch das Negativ-Binomial-Modell schließen lässt. Schließlich ist der geschätzte Wert $\hat{\theta}$ für θ mit 708.2 aufgeführt und die geschätzte Standardabweichung beträgt

243.3. Es ergibt sich also für $\hat{\alpha}$ ein Wert von $\hat{\alpha} = 1/\hat{\theta} = 1/708.2 = 0.0014$, welcher relativ klein ist. Die geschätzte Standardabweichung von $\hat{\alpha}$, welche mit Hilfe der Delta-Methode berechnet werden kann, ist 0.0005.

Es werden die Devianz- und Pearson-Residuen berechnet. In der Abbildung 8.15 werden die Devianz- und Pearson-Residuen grafisch dargestellt.

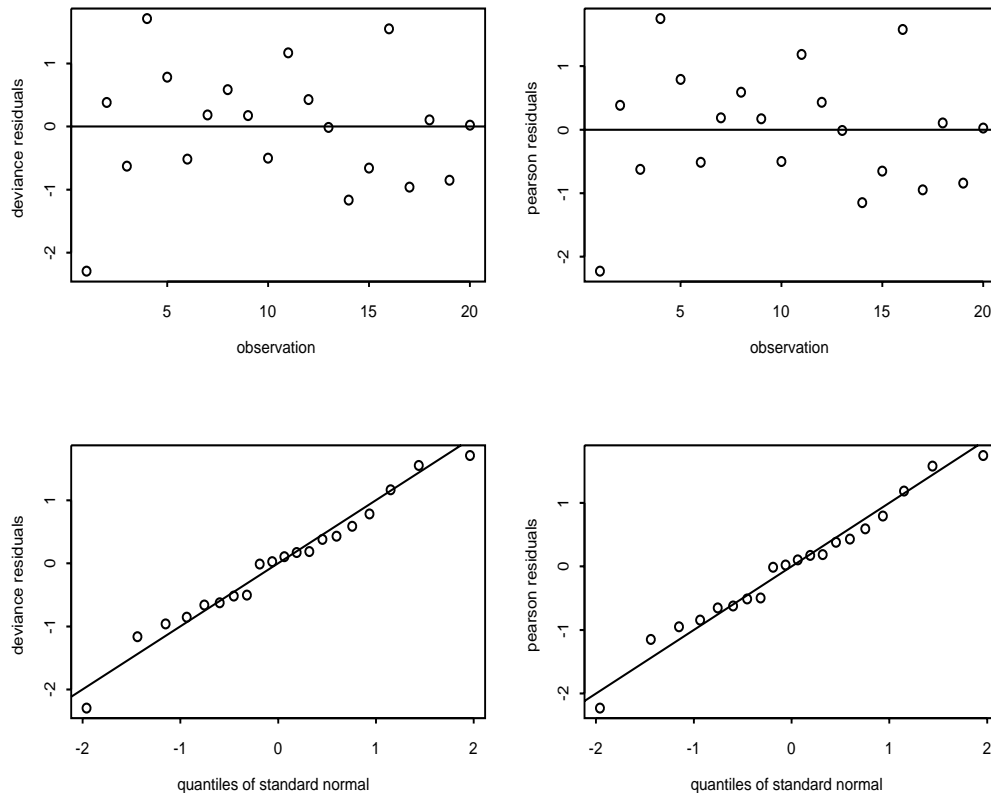


Abbildung 8.15: Devianz- und Pearson-Residuen Negativ-Binomial-Modells mit geschätztem α für den KFZ-Datensatz (oben), Plot der Devianz- und Pearson-Residuen gegen die Standardnormalverteilungsquantile (unten). Die durchgezogene Linie stellt die Winkelhalbierende dar.

In den oberen Grafiken streuen die Residuen um Null, lediglich die Beobachtung 1 überschreitet absolut den Wert von 2. Die Residuen der anderen Beobachtungen sind alle absolut kleiner als 2. Die unteren beiden Grafiken zeigen die Standardnormalverteilungsquantile gegen die Devianz- bzw. Pearson-Residuen. Man sieht, dass die Residuen sich nun deutlich entlang der Winkelhalbierenden bewegen.

Mit der Funktion `diag.nb` lassen sich die h_{ii} -Werte, die $DFFITs_i$ und die Cook's Distanz D_i berechnen. In Tabelle 8.10 stehen die einzelnen Werte und in Abbildung 8.16 werden sie mittels Stabdiagrammen dargestellt.

Wie aus der Grafik und der Tabelle 8.10 ersichtlich, hat nun die Beobachtung 5 (Kombination von *Merit 3* und *Class 5*) den höchsten Wert für das h_{ii} -Element, überschreitet aber nicht die Schranke von $2p/n = 0.8$. Der Wert ist 0.75. Die nächsthöheren Werte bewegen sich um den Wert 0.5. Dies sind die Beobachtungen 1 (Kombination von *Merit 3* und *Class 1*), 6 (Kombination von

\$hat:

```
[1] 0.5245487 0.4672855 0.4473486 0.4585271 0.7528586 0.5184710 0.3769528
[8] 0.4041988 0.4174793 0.4935997 0.4614502 0.3656911 0.4104051 0.4107780
[15] 0.4096386 0.4364737 0.4237319 0.4180313 0.4250115 0.3775184
```

\$dffits:

```
[1] -3.098963464 0.439935815 -0.688380635 2.418296264 0.846274557
[6] -0.547137058 0.147588708 0.609097347 0.178579372 -0.265494527
[11] 1.447829048 0.401110755 -0.009991505 -1.262559186 -0.436255154
[16] 2.103710457 -1.020626120 0.112623721 -0.956113779 0.037553481
```

\$cook:

```
[1] 0.77606435538 0.02592117552 0.06453073103 0.54216030592 0.13785426449
[6] 0.04608143715 0.00317944145 0.04746564984 0.00431796560 0.01601539394
[11] 0.24601021713 0.02081431466 0.00003873383 0.19242417284 0.03259038500
[16] 0.42762202880 0.13335077054 0.00169465915 0.11396025220 0.00018893088
```

Tabelle 8.10: h_{ii} -Werte, $DFFITs_i$ und Cook's Distanz D_i des Negativ-Binomial-Modells mit geschätztem α für den KFZ-Datensatz

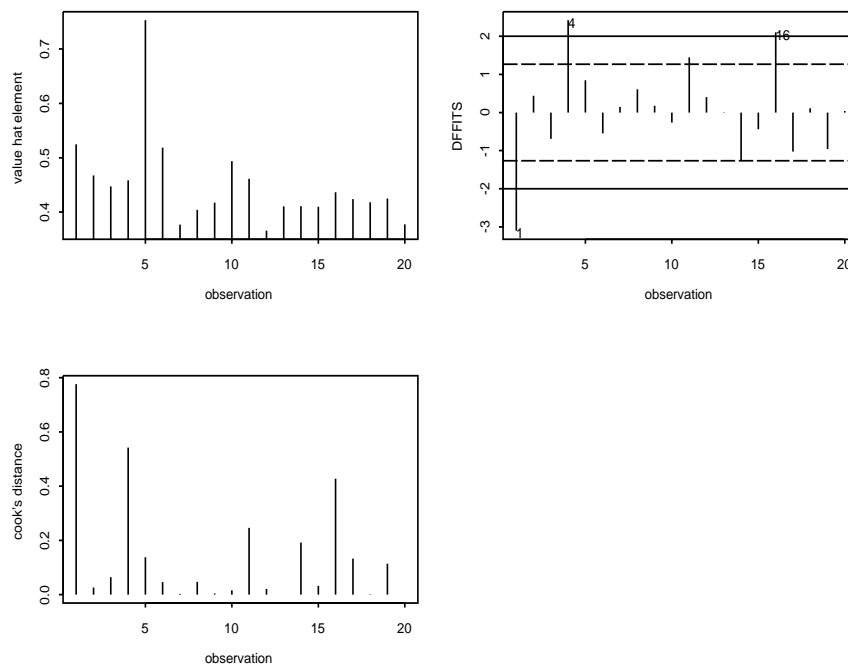


Abbildung 8.16: Stabdiagramm der h_{ii} -Werte (links oben), $DFFITs_i$ (rechts oben) und Cook's Distanz D_i (links unten) des Negativ-Binomial-Modells mit geschätztem α für den KFZ-Datensatz. Die gestrichelte bzw. durchgezogene Linie markiert die absoluten Schranken $2\sqrt{p/n} = 2\sqrt{8/20} = 1.26$ und 2.

Merit 2 und *Class 1*) und 10 (Kombination von *Merit 2* und *Class 5*). Für die Beobachtungen 1, 4 (Kombination von *Merit 3* und *Class 4*) und 16 (Kombination von *Merit 0* und *Class 1*) übersteigen die Werte der $DFFITs_i$ die Schranke von 2, dagegen bleiben die Werte für die Cook's Distanz D_i der 20 Beobachtungen alle unter dem Wert von 1. Der höchste Wert wird hier mit einem Wert von 0.78 von der Beobachtung 1 eingenommen.

Nun werden die $DFBETAS_{j,i}$ mittels der Funktion `dfbetas.nb` berechnet. In der Abbildung 8.17 sind die $DFBETAS_{j,i}$ für jede Kovariable zu sehen, welche mit Hilfe der Funktion `dfbetas.nb` berechnet werden können.

Keine der $DFBETAS_{j,i}$ überschreitet den Wert von 2. Lediglich die Schranke von $2\sqrt{1/n} = 0.45$ wird von einigen Beobachtungen überschritten. Hier ist anzumerken, dass die $DFBETAS_{j,i}$ der Beobachtung 16 für jede Kovariable den Wert $2\sqrt{1/n} = 0.45$ übersteigen. Auch die Werte der Beobachtungen 1 und 11 überschreiten in 6 der 8 Fälle die Schranke von 0.45.

Um den $COVRATIO_i$ auszurechnen, wird die Funktion `covratio.nb` verwendet. In Tabelle 8.11 sieht man die geordneten Werte mit den Beobachtungsindizes und in Abbildung 8.18 anhand eines Stabdiagramms.

Die Werte des $COVRATIO_i$ sind sehr hoch, wie man aus Tabelle 8.11 erkennen kann, lediglich der Wert für Beobachtung 1 liegt mit 0.03 deutlich unter dem Wert von 1. Somit liegt kein Wert annähernd um 1, so dass sich für alle Beobachtungen große Veränderungen nach Weglassen einer Beobachtung ergeben würden. Dies resultiert aus der geringen Anzahl der Beobachtungen von 20.

Abschließend wird die Funktion `devred.nb` angewandt, um die Devianzreduktion zu messen. In Tabelle 8.12 sieht man die geordneten Werte mit den Beobachtungsindizes und in Abbildung 8.19 anhand eines Stabdiagramms.

Die Schranke von 4 wird hier von keinem Wert erreicht. Die höchste Devianzreduktion ergibt sich für die Beobachtung 1 mit einem Wert von 1.2. Die Werte der Beobachtungen 3, 5, 11 und 17 sind ähnlich hoch und verringern nach deren Weglassen die Devianz am stärksten.

Nun werden anhand von Tabelle 8.13 und 8.14 die Schäden aus dem Datensatz (*Claims*) mit den geschätzten Erwartungswerten des Poissonmodells und des Negativ-Binomial-Modells mit geschätztem α verglichen. In Tabelle 8.13 sind die Daten (*Claims*) und die geschätzten Erwartungswerte für das Poissonmodell und das Negativ-Binomial-Modell mit geschätztem α aufgeführt und in Tabelle 8.14 sind diese Werte noch durch die Versicherungsjahre (*Insured*) t_i gewichtet. Aus den Tabellen 8.13 und 8.14 erkennt man, dass die meisten geschätzten Erwartungswerte für das Negativ-Binomial-Modell mit geschätztem α eher den Daten gleichen als das für die geschätzten Erwartungswerte für das Poissonmodell der Fall ist. In 6 der 20 Fälle ist die Schätzung des Poissonmodells besser, nämlich für die Kombinationen von *Class 1* und *Merit 0-3* und für die Kombinationen *Class 3* und *Merit 2-3*. Auffallend dabei ist, dass es gerade die hohen Werte sind, die besser durch das Poissonmodell geschätzt werden. In den Abbildungen 8.20 und 8.21 werden jeweils die Differenzen von wahren Datenwerten und den geschätzten Erwartungswerten des entsprechenden Modells geplottet. Um eine bessere Vergleichbarkeit zu erreichen, sind die unteren Grafiken in den beiden Abbildungen gleich skaliert. Die Grafiken spiegeln die vorherigen Aussagen wieder. Vor allem für die gewichteten Erwartungswerte in Abbildung 8.21 erkennt man, dass die Differenzen zwischen Daten und geschätzten, gewichteten Erwartungswerten aus dem Negativ-Binomial-Modell mit geschätztem α relativ konstante abweichende Werte aufweisen, was aus dem flacheren Verlauf der Oberfläche ersichtlich ist. Dagegen unterscheiden sich die Differenzen aus Daten und geschätzten, gewichteten Erwartungswerten aus dem Poissonmodell deutlich, was aus dem steigendem Verlauf der Oberfläche sichtbar wird. Für *Class 4* und *5* kombiniert mit *Merit 0* und *1* ergeben sich die höchsten Abweichungen.

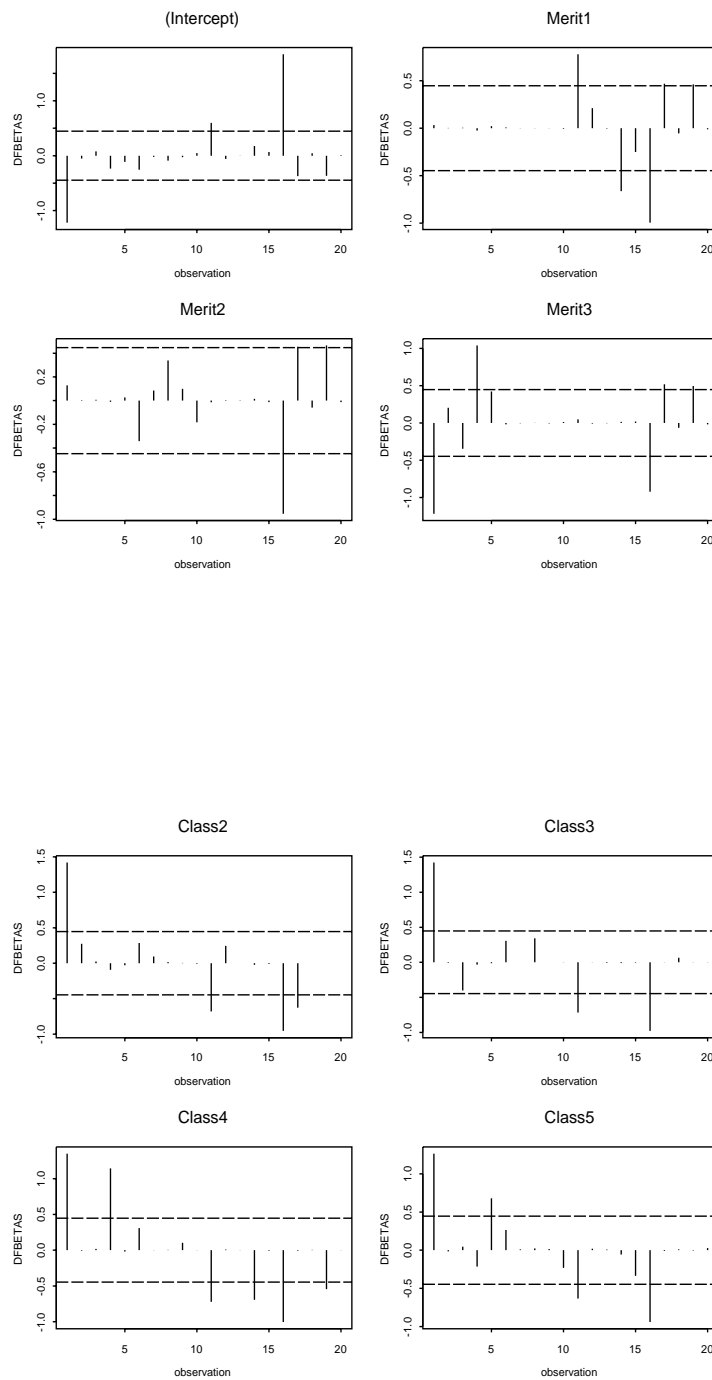


Abbildung 8.17: Stabdiagramm der $DFBETAS_{j,i}$ des Negativ-Binomial-Modells mit geschätztem α für den KFZ-Datensatz. Die gestrichelte Linie markiert die absolute Schranke $2\sqrt{1/n} = 2\sqrt{1/20} = 0,45$.

```
$covratio.ord:
```

```
[1] 2525.5263007 2187.1422173 2042.0524185 2007.6012619 1831.4616702
[6] 1780.8424536 1443.2245469 1320.2608661 1231.8329330 1179.6343547
[11] 1068.3731068 725.5634232 609.8901585 576.1404879 538.6218900
[16] 284.6537324 244.1875207 26.6361993 6.1593592 0.0324231
```

```
$beob:
```

```
[1] 18 13 9 2 6 3 20 12 8 7 5 19 17 10 15 11 14 16 4 1
```

Tabelle 8.11: Geordnete Werte des $COVRATIO_i$ des Negativ-Binomial-Modells mit geschätztem α für den KFZ-Datensatz

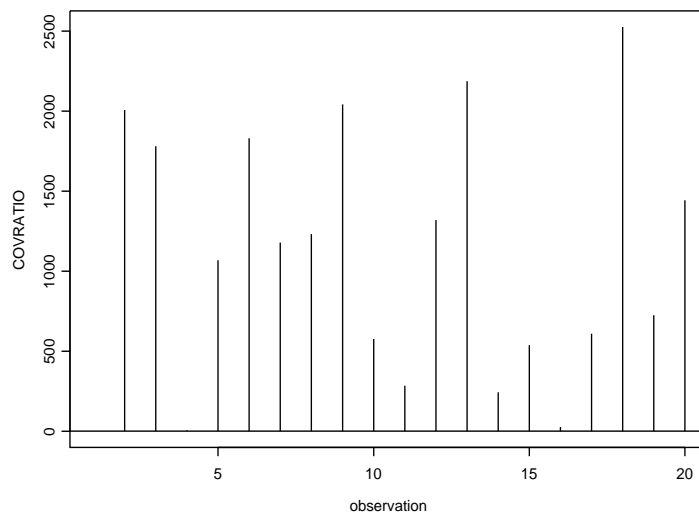


Abbildung 8.18: Stabdiagramm der Werte des $COVRATIO_i$ des Negativ-Binomial-Modells mit geschätztem α für den KFZ-Datensatz

```

$devred.ord:
[1] 1.1949938 1.1717836 1.1283231 1.1197826 1.0896266 0.9752102 0.9305105
[8] 0.9203602 0.8927752 0.8505613 0.8503775 0.8265635 0.8227866 0.8193791
[15] 0.8112499 0.7748868 0.6517496 0.5957814 0.5923764 0.5194916
    
```

```

$beob:
[1] 1 5 11 3 17 6 16 14 18 2 8 19 9 13 12 4 20 7 15 10
    
```

Tabelle 8.12: Geordnete Werte der Devianzreduktion ΔD_i des Negativ-Binomial-Modells mit geschätztem α für den KFZ-Datensatz

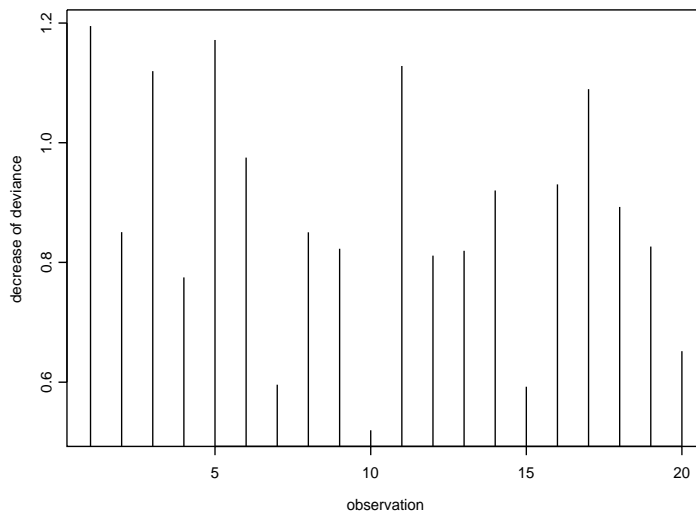


Abbildung 8.19: Stabdiagramm der Werte der Devianzreduktion ΔD_i des Negativ-Binomial-Modells mit geschätztem α für den KFZ-Datensatz

Vergleich der Daten mit geschätzten Erwartungswerten						
	Class 1	Class 2	Class 3	Class 4	Class5	
Merit 0	37730	3421	7565	11345	1291	y_i
	35772	3789	7862	12533	1393	$\hat{\mu}_i(Poi)$
	35597	3559	7533	11727	1289	$\hat{\mu}_i(NB\hat{\alpha})$
Merit 1	19346	1430	3546	3618	613	y_i
	18607	1493	3704	4059	687	$\hat{\mu}_i(Poi)$
	18506	1402	3547	3796	635	$\hat{\mu}_i(NB\hat{\alpha})$
Merit 2	13792	1001	2695	3054	487	y_i
	13688	1022	2656	3137	524	$\hat{\mu}_i(Poi)$
	14069	991	2629	3032	501	$\hat{\mu}_i(NB\hat{\alpha})$
Merit 3	217151	14506	31964	22884	6560	y_i
	219950	14052	31546	21170	6345	$\hat{\mu}_i(Poi)$
	237042	14295	32737	21453	6360	$\hat{\mu}_i(NB\hat{\alpha})$

Tabelle 8.13: Vergleich der Schadenzenahlen mit den geschätzten Erwartungswerten des Poissonmodells und des Negativ-Binomial-Modells mit geschätztem α . Die obere Zahl ist jeweils die Schadenzenahl aus dem Datensatz, die mittlere der geschätzte Erwartungswert für das Poissonmodell und die untere für das Negativ-Binomial-Modell mit geschätztem α .

Vergleich der Daten mit geschätzten Erwartungswerten						
	Class 1	Class 2	Class 3	Class 4	Class5	
Merit 0	0.138	0.159	0.201	0.200	0.150	y_i/t_i
	0.131	0.176	0.209	0.221	0.162	$\hat{\mu}_i(Poi)/t_i$
	0.130	0.166	0.200	0.207	0.150	$\hat{\mu}_i(NB\hat{\alpha})/t_i$
Merit 1	0.118	0.147	0.174	0.172	0.126	y_i/t_i
	0.114	0.154	0.182	0.193	0.141	$\hat{\mu}_i(Poi)/t_i$
	0.113	0.144	0.174	0.180	0.131	$\hat{\mu}_i(NB\hat{\alpha})/t_i$
Merit 2	0.106	0.138	0.170	0.172	0.121	y_i/t_i
	0.105	0.141	0.167	0.177	0.130	$\hat{\mu}_i(Poi)/t_i$
	0.108	0.137	0.166	0.171	0.124	$\hat{\mu}_i(NB\hat{\alpha})/t_i$
Merit 3	0.079	0.111	0.129	0.146	0.102	y_i/t_i
	0.080	0.108	0.128	0.135	0.099	$\hat{\mu}_i(Poi)/t_i$
	0.086	0.110	0.132	0.137	0.099	$\hat{\mu}_i(NB\hat{\alpha})/t_i$

Tabelle 8.14: Vergleich der Schadenzenahlen mit den geschätzten, gewichteten Erwartungswerten des Poissonmodells und des Negativ-Binomial-Modells mit geschätztem α . Die obere Zahl ist jeweils die Schadenzenahl aus dem Datensatz gewichtet mit t_i , die mittlere der geschätzte Erwartungswert gewichtet mit t_i für das Poissonmodell und die untere für das Negativ-Binomial-Modell mit geschätztem α .

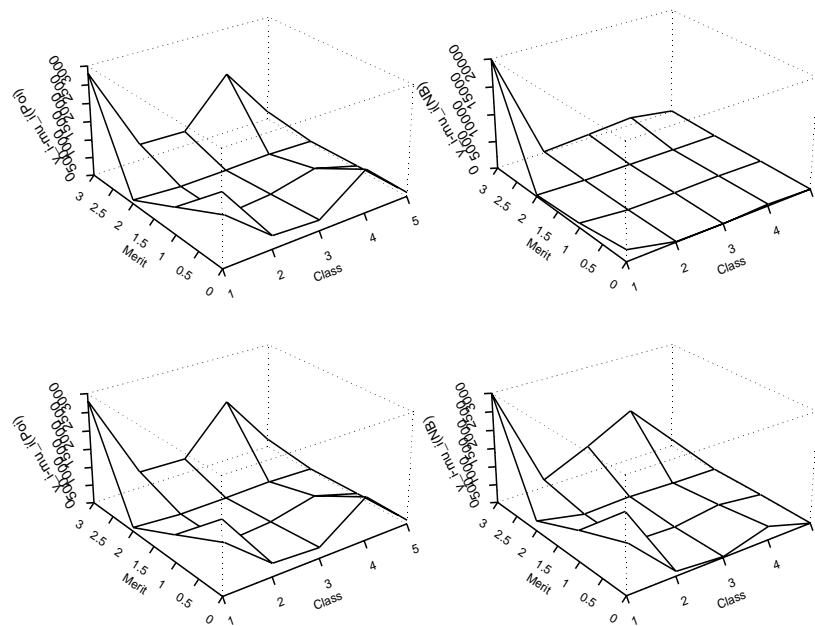


Abbildung 8.20: Differenz der y_i und der $\hat{\mu}_i(Poi)$ (links) bzw. der y_i und der $\hat{\mu}_i(NB\hat{\alpha})$ (rechts) ohne Skalierung (oben) und mit Skalierung ($|\cdot| \leq 3000$) der z-Achse (unten)

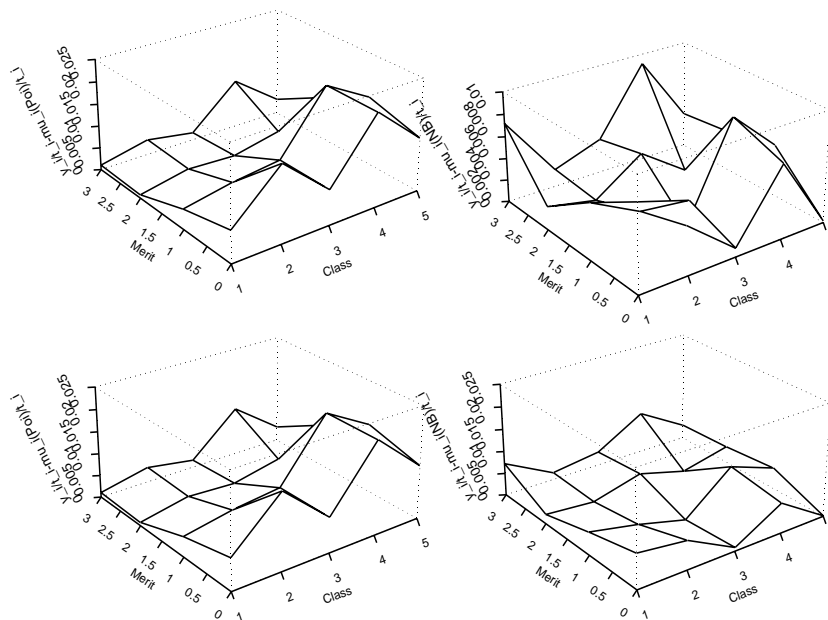


Abbildung 8.21: Differenz der y_i/t_i und der $\hat{\mu}_i(Poi)/t_i$ (links) bzw. der y_i/t_i und der $\hat{\mu}_i(NB\hat{\alpha})/t_i$ (rechts) ohne Skalierung (oben) und mit Skalierung ($|\cdot| \leq 0.025$) der z-Achse (unten)

Der KFZ-Datensatz wurde nach demselben Schema wie der Patentdatensatz analysiert. Die deutlich geringere Anzahl an Beobachtungen ($n = 20$) hatte zur Folge, dass eigentlich fast jeder einzelne Punkt deutlichen Einfluß auf die geschätzten Werte und das Modell hatte. Die geringe Anzahl an Beobachtungen war auch ein Grund für teilweise schlechtere Schätzungen, was sich wiederum in den zumeist hohen Überschätzungen spiegelte. Generell wurde mit dem Negativ-Binomial-Modell eine bessere Anpassung erreicht.

Kapitel 9

Zusammenfassung

Abschließend sollen in dieser Arbeit die wichtigsten Erkenntnisse und Ergebnisse, sowie mögliche Erweiterungen zusammengefasst werden. Anfangs wurde die Problematik der Regressionsdiagnostiken anhand der Linearen Modelle erklärt. Deshalb war es zunächst wichtig, die Linearen Modelle und die Parameterschätzung darin vorzustellen. Die Definition von einem Ausreißer und einer einflußreichen Beobachtung folgte. Mit Hilfe der elementarsten Diagnostiken, den Residuen und der Hatmatrix wurden komplexere Diagnostiken hergeleitet. Wesentlicher Punkt bei der Herleitung der Regressionsdiagnostiken war, dass man die geschätzten Werte basierend auf dem vollständigen Datensatz und dem Datensatz nach Weglassen der i -ten Beobachtung betrachtete. Das vorrangige Ziel war nun das Entwickeln von Regressionsdiagnostiken für Zähldaten. Da die Modellierung von Zähldaten hauptsächlich durch das Poissonmodell durchgeführt wird, wurde der Übergang zu den Generalisierten Linearen Modellen nötig, welche es erlauben, dass alle Verteilungen für die Zielvariable zugelassen werden, die die exponentielle Familie umfassen. Wegen der Eigenschaft der Gleichheit von Erwartungswert und Varianz des Poissonmodells, wurde eine weitere Verteilung, die Negativ-Binomial-Verteilung, zur Modellierung herangezogen. Zunächst wurden die Generalisierten Linearen Modelle definiert und die Parameterschätzung basierend auf dem IWLS, welcher einen iterativen Algorithmus darstellt, vorgestellt. Ebenso wie in den Linearen Modellen definierte man die einfachsten Diagnostiken, nämlich die Residuen und die Hatmatrix. Um nun analog zu den Linearen Modellen die komplexeren Regressionsdiagnostiken zu berechnen, war eine Approximation, welche auf Taylor basierte, nötig. Für die geschätzten Regressionskoeffizienten nach Weglassen einer Beobachtung wurde eine Ein-Schritt-Approximation hergeleitet, welche von dem Wert für die geschätzten Regressionskoeffizienten basierend auf dem vollständigen Datensatz abhängte. In der praktischen Anwendung zeigte sich, dass diese Ein-Schritt-Approximation relativ gut den exakten Wert widerspiegelte. Für manche x - bzw. y -Ausreißer waren die größten Differenzen festzustellen. Nun konnten weitere Diagnostiken basierend auf dieser Ein-Schritt-Approximation angegeben werden und mit Hilfe dieser Diagnostiken sollten Ausreißer und einflußreiche Beobachtungen identifiziert werden. Für die Poissonverteilung und die Negativ-Binomial-Verteilung mit festem α konnten nun die Regressionsdiagnostiken berechnet werden. In einigen Fällen wurde der wahre Wert überschätzt. Vor allem für hohe Werte der Regressionsdiagnostiken war diese Überschätzung besonders hoch. Beobachtungen, die in der praktischen Anwendung auffallende Werte hatten, wurden demnach als Ausreißer oder als einflußreiche Beobachtung deklariert. Festzustellen war, dass aufgrund der verschiedenen Modellierung mit dem Poissonmodell bzw. dem Negativ-Binomial-Modell mit festem α auch teilweise Unterschiede in der Identifizierung der einflußreichen Beobachtungen bestanden. Dazu sollte man den Datensatz in einer explorativen Datenanalyse genau untersuchen und danach festlegen, welche Verteilung zur Modellierung am besten geeignet ist. Hierbei ist festzustellen, ob diskrete

oder stetige Daten analysiert werden und wie sich die Zielvariable gegenüber den Kovariablen verhält. Nach dieser Entscheidung berechnet man die Regressionsdiagnostiken und findet so die einflußreichsten Beobachtungen. Ob man nun eine Schätzung ohne eine Beobachtung durchführt oder nicht bleibt danach zu entscheiden. Wenn man feststellt, dass der Datenpunkt ein Meßfehler oder ein Fehler der Datenübertragung war, wird man ihn sicherlich löschen oder korrigieren. Kann dies ausgeschlossen werden, so wird man die Beobachtung im Datensatz belassen oder eventuell niedriger gewichten. Außerdem können auffällige Beobachtungen immer wieder neue Informationen beinhalten, die eine nochmalige Modellanalyse rechtfertigen.

Da das Modell auch mit Hilfe der Negativ-Binomial-Verteilung mit geschätztem α angegeben werden kann, wurden die Regressionsdiagnostiken für dieses Modell ebenso berechnet. Hier machte man vom Profile-Likelihood Gebrauch und die Schätzungen der Regressionskoeffizienten und des Wertes von α wurden exakt berechnet. Eine Anwendung einer Ein-Schritt-Approximation hätte zu einer ungenauen Schätzung geführt, weshalb die wahren Werte verwendet wurden. Generell sollte man Datensätze, die ein hohes Maß an Überdispersion aufweisen, mit dem Negativ-Binomial-Modell modellieren. Die praktische Anwendung zeigte, dass die Daten zumeist viel besser angepasst wurden, als dies mit dem Poissonmodell möglich war. Wie gesagt, gab es zwischen den Modellen bezüglich einflußreicher Beobachtungen Unterschiede. So beeinflussten einzelne Beobachtungen in einem Negativ-Binomial-Modell mit geschätztem α die Schätzung der Werte generell nicht so stark wie im Poissonmodell. Natürlich spielt hier die Modellanpassung der Daten auch eine große Rolle.

In dieser Arbeit wurden die wichtigsten Regressionsdiagnostiken dargestellt und jeweils für die Linearen Modelle und die Generalisierten Linearen Modelle hergeleitet. Weitere Diagnostiken basierend auf den Linearen Modellen sind im Artikel von Chatterjee und Hadi (1986) zu finden. Hier alle aufzuführen, hätte sicherlich den Rahmen einer Diplomarbeit gesprengt. Desweiteren ist es möglich, nicht nur das Weglassen einer Beobachtung zu betrachten, sondern die Löschung von mehreren Beobachtungen. Im Buch von Belsley et al. (1980) wird darauf für den Fall der Linearen Modelle eingegangen. Darüberhinaus beschäftigt sich dieses Buch mit der Multikollinearität, was ein weiterführendes Thema darstellt. Desweiteren besteht die Möglichkeit eine Simulation durchzuführen, die die Genauigkeit der Approximation untersucht und offen legt, was passiert, wenn man einen negativbinomialverteilten Datensatz durch das Poissonmodell anpasst und umgekehrt.

Anhang A

Die Poissonverteilung

Definition und Eigenschaften der Poissonverteilung

Sei Y eine Zufallsvariable mit einer diskreten Verteilung, die definiert ist auf $\mathbb{N} \cup \{0\} = \{0, 1, 2, \dots\}$. Y ist dann poissonverteilt mit Parameter μ , kurz $Y \sim Poi(\mu)$, falls die Wahrscheinlichkeitsfunktion folgende Form hat:

$$P(Y = y) = e^{-\mu} \frac{\mu^y}{y!} \quad \text{für} \quad \mu > 0 \quad \text{und} \quad y = 0, 1, 2, \dots$$

Der Erwartungswert und die Varianz der Poissonverteilung sind gleich, d.h.

$$E(Y) = Var(Y) = \mu.$$

Diese Gleichheit von Erwartungswert und Varianz wird als *Äquidispersion* bezeichnet. Falls die Varianz größer als der Erwartungswert ist, spricht man von *Überdispersion*. Umgekehrt spricht man von *Unterdispersion*, falls die Varianz kleiner als der Erwartungswert ist.

Die momenterzeugende Funktion ist definiert durch

$$M_Y(s) = E(e^{sY}) = \exp(\mu(e^s - 1)).$$

Durch k -maliges Differenzieren erhält man die ersten k Momente, d.h.

$$E(Y^k) = M_Y^{(k)}(s) = \left. \frac{\partial^k M_Y(s)}{\partial s^k} \right|_{s=0} \quad \text{für} \quad k = 0, 1, 2, \dots$$

Desweiteren gilt die Additivität der Poissonverteilung, d.h. seien $Y_i \sim Poi(\mu_i)$, $i = 1, 2, \dots$, unabhängige Zufallsvariablen und es gelte $\sum_{i=1}^{\infty} \mu_i < \infty$, dann gilt:

$$Z = \sum Y_i \sim Poi\left(\sum \mu_i\right).$$

Weitere Besonderheiten der Poissonverteilung sind die Approximation der Binomialverteilung für genügend kleine p und grosse n , d.h. die Poissonverteilung stellt eine Grenzverteilung für die Binomialverteilung dar. Außerdem ist es möglich die Poissonverteilung durch die Normalverteilung zu approximieren.

Die Poissonverteilung ist ein Spezialfall der Negativ-Binomial-Verteilung, was in Cameron und Trivedi (1998) (Seite 75) kurz gezeigt wird. Wird nämlich in der Negativ-Binomial-Verteilung

$\alpha = 0$ gewählt, so ergibt sich daraus die Poissonverteilung. Ist also Y eine negativbinomialverteilte Zufallsvariable mit Parametern μ und $\alpha = 0$, so kann gezeigt werden, dass Y eine poissonverteilte Zufallsvariable mit Parameter μ ist. Es ist nach Anhang B

$$P(Y = y) = \frac{\Gamma(y + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y + 1)} \left(\frac{1}{1 + \alpha\mu}\right)^{\alpha^{-1}} \left(\frac{\alpha\mu}{1 + \alpha\mu}\right)^y$$

$$\stackrel{a := \alpha^{-1}}{=} \frac{\Gamma(y + a)}{\Gamma(a)\Gamma(y + 1)} \left(\frac{1}{1 + \frac{\mu}{a}}\right)^a \left(\frac{\frac{\mu}{a}}{1 + \frac{\mu}{a}}\right)^y$$

Es gilt folgendes:

$$\frac{\Gamma(y + a)}{\Gamma(a)} = \frac{(y + a - 1)!}{(a - 1)!} = \prod_{j=0}^{y-1} (j + a), \quad (\text{A.1})$$

$$\left(\frac{1}{a + \mu}\right)^y = \prod_{j=0}^{y-1} \left(\frac{1}{a + \mu}\right), \quad (\text{A.2})$$

$$\left(\prod_{j=0}^{y-1} \frac{\frac{j}{a} + 1}{1 + \frac{\mu}{a}}\right) \xrightarrow{a \rightarrow \infty} 1, \quad (\text{A.3})$$

$$\left(\frac{1}{1 + \frac{\mu}{a}}\right)^a = \left(1 + \frac{\mu}{a}\right)^{-a} \xrightarrow{a \rightarrow \infty} e^{-\mu}. \quad (\text{A.4})$$

Damit gilt weiter

$$P(Y = y) \stackrel{(\text{A.1})}{=} \left(\prod_{j=0}^{y-1} (j + a)\right) \frac{1}{y!} \left(\frac{1}{1 + \frac{\mu}{a}}\right)^a \left(\frac{\mu}{a + \mu}\right)^y$$

$$\stackrel{(\text{A.2})}{=} \left(\prod_{j=0}^{y-1} \frac{j + a}{a + \mu}\right) \left(\frac{1}{1 + \frac{\mu}{a}}\right)^a \frac{\mu^y}{y!}$$

$$= \left(\prod_{j=0}^{y-1} \frac{\frac{j}{a} + 1}{1 + \frac{\mu}{a}}\right) \frac{1}{y!} \left(\frac{1}{1 + \frac{\mu}{a}}\right)^a \mu^y$$

Nun gilt mit (A.3) und (A.4)

$$\lim_{\alpha \rightarrow 0} P(Y = y) = \lim_{a \rightarrow \infty} P(Y = y) = 1e^{-\mu} \frac{\mu^y}{y!}.$$

Somit erhält man die Wahrscheinlichkeitsfunktion der Poissonverteilung mit Parameter μ .

Anhang B

Die Negativ-Binomial-Verteilung

Definition und Eigenschaften der Negativ-Binomial-Verteilung

Sei Y eine Zufallsvariable mit einer diskreten Verteilung, die definiert ist auf $\mathbb{N} \cup \{0\} = \{0, 1, 2, \dots\}$. Y ist dann negativbinomialverteilt mit Parameter μ und α , kurz $Y \sim NB(\mu, \alpha)$, falls die Wahrscheinlichkeitsfunktion folgende Form hat:

$$P(Y = y) = \frac{\Gamma(y + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y + 1)} \left(\frac{1}{1 + \alpha\mu} \right)^{\alpha^{-1}} \left(\frac{\alpha\mu}{1 + \alpha\mu} \right)^y$$

für $\alpha \geq 0$ und $y = 0, 1, 2, \dots$

Der Erwartungswert und die Varianz berechnen sich wie folgt:

$$E(Y) = \mu, \quad \text{Var}(Y) = \mu + \alpha\mu^2.$$

Wie man sieht, ist die Varianz größer als der Erwartungswert, falls $\alpha > 0$ ist. Dadurch lassen sich Daten mit Überdispersion besser modellieren.

Die momenterzeugende Funktion ist gegeben durch

$$M_Y(s) = E(e^{sY}) = \left(\frac{\frac{1}{1+\alpha\mu}}{1 - \left(1 - \frac{1}{1+\alpha\mu}\right) e^s} \right)^{\alpha^{-1}}.$$

Wie in Anhang A bei der Poissonverteilung definiert, lassen sich wiederum die ersten k Momente ausrechnen.

Ebenso wie die Poissonverteilung ist die Negativ-Binomial-Verteilung additiv, d.h. seien $Y_i \sim NB(\mu_i, \alpha)$, $i = 1, 2, \dots$, unabhängige Zufallsvariablen und es gelte $\sum_{i=1}^{\infty} \mu_i < \infty$, dann gilt:

$$Z = \sum Y_i \sim NB\left(\sum \mu_i, \alpha\right).$$

Darüberhinaus enthält die Negativ-Binomial-Verteilung die Poissonverteilung als Spezialfall. Diese ergibt sich, wenn man die Parameter der Negativ-Binomial-Verteilung $\alpha = 0$ und μ beliebig wählt.

Anhang C

Maximum-Likelihood-Schätzung

Es sei $f(\mathbf{y}; \boldsymbol{\theta})$ die Wahrscheinlichkeits- bzw. Dichtefunktion eines Zufallsvariablenvektors $\mathbf{Y} = (Y_1, \dots, Y_n)^T$, wobei der Parametervektor $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)^T$ in dem zulässigen Parameterraum Θ liegt. Betrachtet man nun $f(\mathbf{y}; \boldsymbol{\theta})$ als Funktion von $\boldsymbol{\theta}$ für feste Realisierungen \mathbf{y} , so bekommt man die *Likelihood-Funktion* $l(\boldsymbol{\theta})$. Für den Fall der Unabhängigkeit der Zufallsvariablen Y_i , $i = 1, \dots, n$, gegeben den Kovariablenvektoren \mathbf{x}_i , $i = 1, \dots, n$ mit Wahrscheinlichkeits- bzw. Dichtefunktion $f(y_i | \mathbf{x}_i; \boldsymbol{\theta})$ ergibt sich wegen der Unabhängigkeit der Y_i die Likelihood-Funktion als

$$l(\boldsymbol{\theta}) := \prod_{i=1}^n f(y_i | \mathbf{x}_i; \boldsymbol{\theta}).$$

Um das Maximieren zu erleichtern, geht man häufig zum Log-Likelihood über, den man durch Logarithmieren erhält:

$$\mathcal{L}(\boldsymbol{\theta}) := \log l(\boldsymbol{\theta}) = \sum_{i=1}^n \log f(y_i | \mathbf{x}_i; \boldsymbol{\theta}) = \sum_{i=1}^n \mathcal{L}_i(\boldsymbol{\theta}).$$

Da die Logarithmusfunktion streng monoton ist, stimmen die Maximalstellen von \mathcal{L} und l überein. Meistens ist es leichter mit dem Log-Likelihood zu arbeiten. Somit wird ein Maximum-Likelihood-Schätzer wie folgt definiert:

Die Zufallsvariablen Y_i , $i = 1, \dots, n$, seien unabhängig verteilt mit Wahrscheinlichkeits- bzw. Dichtefunktion $f(y_i | \mathbf{x}_i; \boldsymbol{\theta})$, $i = 1, \dots, n$. Sei $l(\boldsymbol{\theta}) = \prod_{i=1}^n f(y_i | \mathbf{x}_i; \boldsymbol{\theta})$ die Likelihood-Funktion und $\mathcal{L}(\boldsymbol{\theta}) = \log l(\boldsymbol{\theta})$ die Log-Likelihood-Funktion. Dann heißt jede Maximalstelle

$$\hat{\boldsymbol{\theta}}_{ML} = \operatorname{argmax}\{l(\hat{\boldsymbol{\theta}}); \hat{\boldsymbol{\theta}} \in \Theta\} = \operatorname{argmax}\{\mathcal{L}(\hat{\boldsymbol{\theta}}); \hat{\boldsymbol{\theta}} \in \Theta\}$$

Maximum-Likelihood-Schätzer (ML-Schätzer) für $\boldsymbol{\theta}$. $\hat{\boldsymbol{\theta}}_{ML}$ muss nicht immer existieren und im Falle der Existenz nicht immer eindeutig sein.

Eine wichtige Eigenschaft der ML-Schätzer ist, wenn $d(\boldsymbol{\theta})$ eine Funktion des Parameters $\boldsymbol{\theta}$ ist, dann ist $d(\hat{\boldsymbol{\theta}})$ der ML-Schätzer von $d(\boldsymbol{\theta})$. Dies folgt aus der Definition von $\hat{\boldsymbol{\theta}}$ und wird als *Invarianzeigenschaft* bezeichnet.

Anhang D

Newton-Raphson- und Fisher-Scoring-Methode

Hier wird die *Newton-Raphson-Methode* zur Lösung der nichtlinearen Gleichungen aus Kapitel 4.2 vorgestellt. Es gelten die Definitionen und Bezeichnungen aus Kapitel 4.2. Um diese Methode zur Lösung von

$$S(\boldsymbol{\beta}) := \begin{pmatrix} s_1(\boldsymbol{\beta}) \\ \vdots \\ s_p(\boldsymbol{\beta}) \end{pmatrix} = \mathbf{0}$$

einzusetzen, benötigt man eine Taylorentwicklung erster Ordnung, um die nichtlinearen Gleichungen zu linearisieren:

$$S(\hat{\boldsymbol{\beta}}^{t+1}) \approx S(\hat{\boldsymbol{\beta}}^t) + S'(\hat{\boldsymbol{\beta}}^t)(\hat{\boldsymbol{\beta}}^{t+1} - \hat{\boldsymbol{\beta}}^t) \stackrel{!}{=} \mathbf{0}, \quad (\text{D.1})$$

wobei $\hat{\boldsymbol{\beta}}^t$ momentaner Schätzer ist, $\hat{\boldsymbol{\beta}}^{t+1}$ der neu zu berechnende und $S'(\hat{\boldsymbol{\beta}}^t) =: H(\hat{\boldsymbol{\beta}}^t)$ die Hessematrix der momentanen Schätzung $\hat{\boldsymbol{\beta}}^t$ ist. Man benötigt also die Hessematrix $H(\boldsymbol{\beta})$ an der Stelle $\hat{\boldsymbol{\beta}}^t$. Die Hessematrix ist definiert als das Negative der beobachteten Fisher-Informationsmatrix $FI_{obs}(\boldsymbol{\beta})$. Die Hessematrix ist gegeben durch

$$H(\boldsymbol{\beta}) := \left(\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right) = \begin{bmatrix} \frac{\partial s_1(\boldsymbol{\beta})}{\partial \beta_1} & \cdots & \frac{\partial s_1(\boldsymbol{\beta})}{\partial \beta_p} \\ \vdots & & \vdots \\ \frac{\partial s_p(\boldsymbol{\beta})}{\partial \beta_1} & \cdots & \frac{\partial s_p(\boldsymbol{\beta})}{\partial \beta_p} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 \mathcal{L}}{\partial \beta_1 \partial \beta_1} & \cdots & \frac{\partial^2 \mathcal{L}}{\partial \beta_1 \partial \beta_p} \\ \vdots & & \vdots \\ \frac{\partial^2 \mathcal{L}}{\partial \beta_p \partial \beta_1} & \cdots & \frac{\partial^2 \mathcal{L}}{\partial \beta_p \partial \beta_p} \end{bmatrix}.$$

Also bezeichnet $FI_{obs}(\boldsymbol{\beta}) = - \left(\frac{\partial^2 \mathcal{L}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right)$ die beobachtete Fisher-Informationsmatrix. Es gilt:

$$\begin{aligned} \frac{\partial^2 \mathcal{L}}{\partial \beta_r \partial \beta_s} &= \frac{\partial}{\partial \beta_r} \left[\sum_{i=1}^n \frac{(y_i - \mu_i)}{v_i} \frac{d\mu_i}{d\eta_i} x_{is} \right] \\ &= \sum_{i=1}^n (y_i - \mu_i) \frac{\partial}{\partial \beta_r} \left[v_i^{-1} \frac{d\mu_i}{d\eta_i} x_{is} \right] + \sum_{i=1}^n v_i^{-1} \frac{d\mu_i}{d\eta_i} x_{is} \underbrace{\frac{\partial}{\partial \beta_r} (y_i - \mu_i)}_{= -\frac{d\mu_i}{d\eta_i} x_{ir}}. \end{aligned} \quad (\text{D.2})$$

Im allgemeinen hängt (D.2) von den Daten ab. Deshalb modifiziert man die Newton-Raphson-Methode, d.h. man benutzt anstelle der beobachteten Fisher-Informationsmatrix die erwartete Fisher-Informationsmatrix $FI(\boldsymbol{\beta})$. Diese erhält man, indem man den Erwartungswert der beobachteten Fisher-Informationsmatrix $FI_{obs}(\boldsymbol{\beta})$ bildet. Die Komponenten der erwarteten Fisher-Informationsmatrix lauten

$$\begin{aligned} -E\left(\frac{\partial^2 \mathcal{L}}{\partial \beta_r \partial \beta_s}\right) &= -E\left(\underbrace{\sum_{i=1}^n (y_i - \mu_i) \frac{\partial}{\partial \beta_r} \left[v_i^{-1} \frac{d\mu_i}{d\eta_i} x_{is} \right]}_{=0}\right) - E\left(-\sum_{i=1}^n \underbrace{v_i^{-1} \left(\frac{d\mu_i}{d\eta_i}\right)^2}_{w_i} x_{is} x_{ir}\right) \\ &= \sum_{i=1}^n w_i x_{ir} x_{is}. \end{aligned}$$

Die erwartete Fisher-Informationsmatrix kann dann wie folgt dargestellt werden:

$$FI(\boldsymbol{\beta}) := \begin{bmatrix} -E\left(\frac{\partial^2 \mathcal{L}}{\partial \beta_r \partial \beta_s}\right) \\ \vdots \\ -E\left(\frac{\partial^2 \mathcal{L}}{\partial \beta_r \partial \beta_s}\right) \end{bmatrix}_{\substack{r=1, \dots, p \\ s=1, \dots, p}} = \mathbf{X}^T \mathbf{W} \mathbf{X}, \quad (\text{D.3})$$

wobei $\mathbf{W} = \text{diag}(w_1, \dots, w_n)$. Wenn man die erwartete Fisher-Informationsmatrix anstelle der beobachteten benützt, bezeichnet man dies als die *Fisher-Scoring-Methode*.

Die Fisher-Scoring-Methode ist ein iterativer Algorithmus zur Bestimmung des ML-Schätzers $\hat{\boldsymbol{\beta}}$. Nun kann ausgehend von Gleichung (D.1) in der $(t+1)$ -ten Iteration ein Schätzwert $\hat{\boldsymbol{\beta}}^{t+1}$ berechnet werden. Da nach Annahme \mathbf{X} vollen Rang besitzt, existiert die Inverse von $FI(\hat{\boldsymbol{\beta}}^t)$ und man erhält somit:

$$S(\hat{\boldsymbol{\beta}}^t) - FI(\hat{\boldsymbol{\beta}}^t)(\hat{\boldsymbol{\beta}}^{t+1} - \hat{\boldsymbol{\beta}}^t) = \mathbf{0} \quad \Leftrightarrow \quad \hat{\boldsymbol{\beta}}^{t+1} - \hat{\boldsymbol{\beta}}^t = FI^{-1}(\hat{\boldsymbol{\beta}}^t)S(\hat{\boldsymbol{\beta}}^t).$$

Es sei also $\hat{\boldsymbol{\beta}}^t$ die momentane Schätzung der Lösung zu $S(\boldsymbol{\beta}) = \mathbf{0}$, dann ist die neue Schätzung $\hat{\boldsymbol{\beta}}^{t+1}$ durch

$$\hat{\boldsymbol{\beta}}^{t+1} = \hat{\boldsymbol{\beta}}^t + FI^{-1}(\hat{\boldsymbol{\beta}}^t)S(\hat{\boldsymbol{\beta}}^t)$$

definiert.

Anhang E

Benutzerdefinierte SPlus-Funktionen

Die Funktion *diag.poi* dient zur Berechnung der Werte der h_{ii} -Werte, der $DFFITs_i^1$ und der Cook's Distanz D_i^1 . Außerdem kann man sich die Werte jeweils in einer Grafik durch *plot=T* veranschaulichen lassen. Die Argumente *r* und *weights* bezeichnen das Modell bzw. bei Vorhandensein eines Offsets die Gewichtung.

Funktion *diag.poi*

```
diag.poi_function(r,weights,plot=T)
{X_r$x
beta_r$coef
eta_ X%*%beta
mu_weights*exp(eta)
W_diag(c(mu))
h_hat(W^(1/2)%*%X)
n_dim(X)[1]
p_dim(X)[2]
esa_matrix(0,nrow=p,ncol=n)
dffits_rep(0,n)
cook_rep(0,n)
pres_resid(r,type="pearson")
presstand_resid(r,type="pearson")/sqrt(1-h)
for (i in 1:n)
{esa[,i]_beta-(W[i,i]^(1/2)*solve(t(X)%*%W%*%X)%*%X[i,]*pres[i])/(1-h[i])
dffits[i]_(mu[i]-weights[i]*exp(t(X[i,])%*%esa[,i]))/sqrt(weights[i]*
exp(t(X[i,])%*%esa[,i])*h[i])
cook[i]_(1/p)*(h[i]/(1-h[i]))*presstand[i]^2}
if (plot)
{par(mfrow=c(2,2))
par(ask=T)
plot(h,type="h",xlab="observation",ylab="value hat element")
abline(2*p/n,0)
eb_rev(sort(h[h>2*p/n]))
b_rev(order(h))[1:length(eb)]
for (i in 1:length(eb))
{t_paste(b[i])
```

```

text(b[i],eb[i],t,adj=0)}
plot(dffits,type="h",xlab="observation",ylab="DFFITs^1")
abline(2*sqrt(p/n),0,lty=4)
abline(-2*sqrt(p/n),0,lty=4)
abline(2,0)
abline(-2,0)
eb_rev(sort(abs(dffits)[abs(dffits)>2]))
b_rev(order(abs(dffits)))[1:length(eb)]
for (i in 1:length(eb))
{t_paste(b[i])
text(b[i],dffits[b[i]],t,adj=0)}
plot(cook,type="h",xlab="observation",ylab="cook's distance^1")
abline(1,0)
eb_rev(sort(cook[cook>1]))
b_rev(order(cook))[1:length(eb)]
for (i in 1:length(eb))
{t_paste(b[i])
text(b[i],eb[i],t,adj=0)}}
return(list(hat=h,dffits=dffits,cook=cook,p=p))}

```

Die Funktion *dfbetas.poi* dient zur Berechnung der Werte der $DFBETAS_{j,i}^1$. Außerdem kann man sich die Werte jeweils in einer Grafik veranschaulichen.

Funktion *dfbetas.poi*

```

dfbetas.poi_function(r,weights,plot=T)
{X_r$x
beta_r$coef
eta_ X%*%beta
mu_weights*exp(eta)
W_diag(c(mu))
h_hat(W^(1/2)%*%X)
n_dim(X)[1]
p_dim(X)[2]
pres_resid(r,type="pearson")
obs_matrix(0,nrow=p,ncol=n)
R_solve(t(X)%*%W%*%X)%*%t(X)%*%W^(1/2)
for (j in 1:p)
{for (i in 1:n)
{obs[j,i]_(R[j,i]/sqrt(t(R[j,])%*%R[j,]))*(pres[i]/(1-h[i]))}}
dfbetas_data.frame(obs)
s_summary(r,cor=F)
dimnames(dfbetas)[[1]]_dimnames(s$coef)[[1]]
if(plot)
{par(mfrow=c(2,2))
par(ask=T)
for (j in 1:p)
{kovar_paste(dimnames(s$coef)[[1]][j])
plot(obs[j,],type="h",xlab="observation",ylab="DFBETAS^1",main=kovar)

```



```

abline(2*sqrt(1/n),0,lty=4)
abline(-2*sqrt(1/n),0,lty=4)
abline(2,0)
abline(-2,0)
eb_rev(sort(abs(obs[j,])[abs(obs[j,])>2]))
b_rev(order(abs(obs[j,]))[1:length(eb)])
for (i in 1:length(eb))
{t_paste(b[i])
text(b[i],obs[j,b[i]],t,adj=0)}}}
return(list(dfbetas=dfbetas))}

```

Die Funktion *covratio.poi* dient zur Berechnung der Werte des $COVRATIO_i^1$. Außerdem kann man sich die Werte in einer Grafik veranschaulichen lassen.

Funktion *covratio.poi*

```

covratio.poi_function(r,weights,plot=T)
{X_r$x
beta_r$coef
eta_ X%%beta
mu_weights*exp(eta)
W_diag(c(mu))
h_hat(W^(1/2)%%X)
h.lm_hat(X)
n_dim(X)[1]
p_dim(X)[2]
esa_matrix(0,nrow=p,ncol=n)
pres_resid(r,type="pearson")
covratio_rep(0,n)
for (i in 1:n)
{esa[,i]_beta-(W[i,i]^(1/2)*solve(t(X)%%W%%X)%%X[i,]*pres[i])/(1-h[i])
covratio[i]_(1/(1-h.lm[i]))*((prod(mu)*weights[i]*exp(X[i,]%%esa[,i]))/
prod(weights*exp(X%%esa[,i])))}
if (plot)
{plot(covratio,type="h",xlab="observation",ylab="COVRATIO^1")}
beob_rev(order(covratio))
covratio.ord_rev(sort(covratio))
return(list(covratio=covratio,covratio.ord=covratio.ord,beob=beob))}

```

Die Funktion *devred.poi* dient zur Berechnung der Werte für die Devianzreduktion für die Ein-Schritt-Approximation. Außerdem kann man sich die Werte in einer Grafik veranschaulichen lassen.

Funktion *devred.poi*

```

devred.poi_function(r,weights,plot=T)
{X_r$x
beta_r$coef
eta_ X%%beta
mu_weights*exp(eta)

```

```

W_diag(c(mu))
h_hat(W^(1/2)%*%X)
n_dim(X)[1]
presstand_resid(r,type="pearson")/sqrt(1-h)
dres_resid(r,type="deviance")
devred_rep(0,n)
for (i in 1:n)
{devred[i]_dres[i]^2+h[i]*presstand[i]^2}
if (plot)
{plot(devred,type="h",xlab="observation",ylab="decrease of deviance^1")
abline(4,0)
eb_rev(sort(devred[devred>4]))
b_rev(order(devred))[1:length(eb)]
for (i in 1:length(eb))
{t_paste(b[i])
text(b[i],eb[i],t,adj=0)}}
beob_rev(order(devred))
devred.ord_rev(sort(devred))
return(list(devred=devred,devred.ord=devred.ord,beob=beob))}

```

Die Funktionen *einflbeob.hat*, *einflbeob.dffits* und *einflbeob.cook* geben die Werte für die entsprechenden Diagnostiken aus. Zudem werden die Beobachtungen mit den zugehörigen Werten, die die kritischen Schranken überschreiten, ausgegeben.

Funktion *einflbeob.hat*

```

einflbeob.hat_function(diag.out)
{h_diag.out$hat
n_length(h)
p_diag.out$p
eb_rev(sort(h[h>2*p/n]))
b_rev(order(h))[1:length(eb)]
return(list(hat=h,eb=eb,b=b))}

```

Funktion *einflbeob.dffits*

```

einflbeob.dffits_function(diag.out)
{dffits_diag.out$dffits
dffits1_abs(diag.out$dffits)
n_length(dffits1)
p_diag.out$p
eb_rev(sort(dffits1[dffits1>2]))
b_rev(order(dffits1))[1:length(eb)]
return(list(dffits=dffits,eb=eb,b=b))}

```

Funktion *einflbeob.cook*

```

einflbeob.cook_function(diag.out)

```

```
{cook_diag.out$cook
n_length(cook)
p_diag.out$p
eb_rev(sort(cook[cook>1]))
b_rev(order(cook))[1:length(eb)]
return(list(cook=cook,eb=eb,b=b))}
```

Die Funktionen *diag.nba*, *dfbetas.nba*, *covratio.nba* und *devred.nba* sind genauso aufgebaut wie für den Poissonfall. Hier spielt nun der Parameter α bzw. θ noch eine wesentliche Rolle, der durch das Argument *theta* übergeben wird.

Funktion *diag.nba*

```
diag.nba_function(r,theta,weights,plot=T)
{X_r$x
beta_r$coef
eta_ X%*%beta
mu_weights*exp(eta)
W_diag(c(mu/(1+mu/theta)))
h_hat(W^(1/2)%*%X)
n_dim(X)[1]
p_dim(X)[2]
esa_matrix(0,nrow=p,ncol=n)
dffits_rep(0,n)
cook_rep(0,n)
pres_resid(r,type="pearson")
presstand_resid(r,type="pearson")/sqrt(1-h)
for (i in 1:n)
{esa[,i]_beta-(W[i,i]^(1/2)*solve(t(X)%*%W%*%X)%*%X[i,]*pres[i])/(1-h[i])
dffits[i]_(mu[i]-weights[i]*exp(t(X[i,])%*%esa[,i]))/sqrt(weights[i]*
exp(t(X[i,])%*%esa[,i])*(1+weights[i]*exp(t(X[i,])%*%esa[,i])/
theta)*h[i])
cook[i]_(1/p)*(h[i]/(1-h[i]))*presstand[i]^2}
if (plot)
{par(mfrow=c(2,2))
par(ask=T)
plot(h,type="h",xlab="observation",ylab="value hat element")
abline(2*p/n,0)
eb_rev(sort(h[h>2*p/n]))
b_rev(order(h))[1:length(eb)]
for (i in 1:length(eb))
{t_paste(b[i])
text(b[i],eb[i],t,adj=0)}
plot(dffits,type="h",xlab="observation",ylab="DFFITS^1")
abline(2*sqrt(p/n),0,lty=4)
abline(-2*sqrt(p/n),0,lty=4)
abline(2,0)
abline(-2,0)
eb_rev(sort(abs(dffits)[abs(dffits)>2]))}
```

```

b_rev(order(abs(dffits)))[1:length(eb)]
for (i in 1:length(eb))
{t_paste(b[i])
text(b[i],dffits[b[i]],t,adj=0)}
plot(cook,type="h",xlab="observation",ylab="cook's distance^1")
abline(1,0)
eb_rev(sort(cook[cook>1]))
b_rev(order(cook))[1:length(eb)]
for (i in 1:length(eb))
{t_paste(b[i])
text(b[i],eb[i],t,adj=0)}}
return(list(hat=h,dffits=dffits,cook=cook,p=p))}

```

Funktion *dfbetas.nba*

```

dfbetas.nba_function(r,theta,weights,plot=T)
{X_r$x
beta_r$coef
eta_ X%*%beta
mu_weights*exp(eta)
W_diag(c(mu/(1+mu/theta)))
h_hat(W^(1/2)%*%X)
n_dim(X)[1]
p_dim(X)[2]
pres_resid(r,type="pearson")/(1-h)
obs_matrix(0,nrow=p,ncol=n)
R_solve(t(X)%*%W%*%X)%*%t(X)%*%W^(1/2)
for (j in 1:p)
{for (i in 1:n)
{obs[j,i]_(R[j,i]/sqrt(t(R[j,])%*%R[j,]))*(pres[i]/(1-h[i]))}}
dfbetas_data.frame(obs)
s_summary(r,cor=F)
dimnames(dfbetas)[[1]]_dimnames(s$coef)[[1]]
if(plot)
{par(mfrow=c(2,2))
par(ask=T)
for (j in 1:p)
{kovar_paste(dimnames(s$coef)[[1]][j])
plot(obs[j,],type="h",xlab="observation",ylab="DFBETAS^1",main=kovar)
abline(2*sqrt(1/n),0,lty=4)
abline(-2*sqrt(1/n),0,lty=4)
abline(2,0)
abline(-2,0)
eb_rev(sort(abs(obs[j,])[abs(obs[j,])>2]))
b_rev(order(abs(obs[j,])))[1:length(eb)]
for (i in 1:length(eb))
{t_paste(b[i])
text(b[i],obs[j,b[i]],t,adj=0)}}}

```

```
return(list(dfbetas=dfbetas))}
```

Funktion *covratio.nba*

```
covratio.nba_function(r,theta,weights,plot=T)
{X_r$x
beta_r$coef
eta_ X%*%beta
mu_weights*exp(eta)
W_diag(c(mu/(1+mu/theta)))
h_hat(W^(1/2)%*%X)
h.lm_hat(X)
n_dim(X)[1]
p_dim(X)[2]
esa_matrix(0,nrow=p,ncol=n)
pres_resid(r,type="pearson")
covratio_rep(0,n)
for (i in 1:n)
{esa[,i]_beta-(W[i,i]^(1/2)*solve(t(X)%*%W%*%X)%*%X[i,]*pres[i])/(1-h[i])
covratio[i]_(1/(1-h.lm[i]))*((prod(mu/(1+mu/theta))*(weights[i]*
exp(t(X[i,])%*%esa[,i])/(1+weights[i]*exp(t(X[i,])%*%
esa[,i])/theta)))/prod(weights*exp(X%*%esa[,i])/(1+weights*
exp(X%*%esa[,i])/theta)))}
if (plot)
{plot(covratio,type="h",xlab="observation",ylab="COVRATIO^1")}
beob_rev(order(covratio))
covratio.ord_rev(sort(covratio))
return(list(covratio=covratio,covratio.ord=covratio.ord,beob=beob))}
```

Funktion *devred.nba*

```
devred.nba_function(r,theta,weights,plot=T)
{X_r$x
beta_r$coef
eta_ X%*%beta
mu_weights*exp(eta)
W_diag(c(mu/(1+mu/theta)))
h_hat(W^(1/2)%*%X)
n_dim(X)[1]
presstand_resid(r,type="pearson")/sqrt(1-h)
dres_resid(r,type="deviance")
devred_rep(0,n)
for (i in 1:n)
{devred[i]_dres[i]^2+h[i]*presstand[i]^2}
if (plot)
{plot(devred,type="h",xlab="observation",ylab="decrease of deviance^1")
abline(4,0)}
```

```

eb_rev(sort(devred[devred>4]))
b_rev(order(devred)[1:length(eb)])
for (i in 1:length(eb))
{t_paste(b[i])
text(b[i],eb[i],t,adj=0)}
beob_rev(order(devred))
devred.ord_rev(sort(devred))
return(list(devred=devred,devred.ord=devred.ord,beob=beob))}

```

Die Funktionen *diag.nb*, *dfbetas.nb*, *covratio.nb* und *devred.nb* dienen zur Schätzung im Negativ-Binomial-Modell mit geschätztem α und sind genauso aufgebaut wie für den Negativ-Binomial-Fall mit festem α . Hier wird nun für $\hat{\beta}_{-i}$ und $\hat{\alpha}_{-i}$ bzw. $\hat{\theta}_{-i}$ der exakte Wert verwendet. Diese werden durch die Argumente *koef.true* und *theta* übergeben.

Funktion *diag.nb*

```

diag.nb_function(r,koef.true,theta,weights,plot=T)
{X_r$x
beta_r$coef
eta_ X%%beta
mu_weights*exp(eta)
W_diag(c(mu/(1+mu/r$theta)))
h_hat(W^(1/2)%%X)
n_dim(X)[1]
p_dim(X)[2]
dffits_rep(0,n)
cook_rep(0,n)
for (i in 1:n)
{dffits[i]_(mu[i]-weights[i]*exp(t(X[i,])%%koef.true[,i]))/sqrt(weights[i]
*exp(t(X[i,])%%koef.true[,i])*(1+weights[i]*exp(t(X[i,])%%
koef.true[,i])/theta[i])*h[i])
cook[i]_(1/p)*t(beta-koef.true[,i])%%t(X)%%W%%X%%(beta-koef.true[,i])}
if (plot)
{par(mfrow=c(2,2))
par(ask=T)
plot(h,type="h",xlab="observation",ylab="value hat element")
abline(2*p/n,0)
eb_rev(sort(h[h>2*p/n]))
b_rev(order(h)[1:length(eb)])
for (i in 1:length(eb))
{t_paste(b[i])
text(b[i],eb[i],t,adj=0)}
plot(dffits,type="h",xlab="observation",ylab="DFFITS")
abline(2*sqrt(p/n),0,lty=4)
abline(-2*sqrt(p/n),0,lty=4)
abline(2,0)
abline(-2,0)
eb_rev(sort(abs(dffits)[abs(dffits)>2]))
b_rev(order(abs(dffits)))[0:length(eb)]

```

```

for (i in 1:length(eb))
{t_paste(b[i])
text(b[i],dffits[b[i]],t,adj=0)}
plot(cook,type="h",xlab="observation",ylab="cook's distance")
abline(1,0)
eb_rev(sort(cook[cook>1]))
b_rev(order(cook))[1:length(eb)]
for (i in 1:length(eb))
{t_paste(b[i])
text(b[i],eb[i],t,adj=0)}}
return(list(hat=h,dffits=dffits,cook=cook,p=p))}

```

Funktion *dfbetas.nb*

```

dfbetas.nb_function(r,koef.true,weights,plot=T)
{X_r$x
beta_r$coef
eta_ X%%beta
mu_weights*exp(eta)
W_diag(c(mu/(1+mu/r$theta)))
n_dim(X)[1]
p_dim(X)[2]
obs_matrix(0,nrow=p,ncol=n)
C_solve(t(X)%*%W%*%X)
for (j in 1:p)
{for (i in 1:n)
{obs[j,i]_(beta[j]-koef.true[j,i])/sqrt(C[j,j])}}
dfbetas_data.frame(obs)
s_summary(r,cor=F)
dimnames(dfbetas)[[1]]_dimnames(s$coef)[[1]]
if(plot)
{par(mfrow=c(2,2))
par(ask=T)
for (j in 1:p)
{kovar_paste(dimnames(s$coef)[[1]][j])
plot(obs[j,],type="h",xlab="observation",ylab="DFBETAS",main=kovar)
abline(2*sqrt(1/n),0,lty=4)
abline(-2*sqrt(1/n),0,lty=4)
abline(2,0)
abline(-2,0)
eb_rev(sort(abs(obs[j,])[abs(obs[j,])>2]))
b_rev(order(abs(obs[j,]))) [0:length(eb)]
for (i in 1:length(eb))
{t_paste(b[i])
text(b[i],obs[j,b[i]],t,adj=0)}}}
return(list(dfbetas=dfbetas))}

```

Funktion *covratio.nb*

```

covratio.nb_function(r,koef.true,theta,weights,plot=T)
{X_r$x
beta_r$coef
eta_ X%%beta
mu_weights*exp(eta)
W_diag(c(mu/(1+mu/r$theta)))
h.lm_hat(X)
n_dim(X)[1]
covratio_rep(0,n)
for (i in 1:n)
{covratio[i]_(1/(1-h.lm[i]))*((prod(mu/(1+mu/r$theta))*(weights[i]*
exp(t(X[i,])%%koef.true[,i])/(1+weights[i]*exp(t(X[i,])%%
koef.true[,i])/theta[i])))/prod(weights*exp(X%%
koef.true[,i])/(1+weights*exp(X%%koef.true[,i])/theta[i])))}
if (plot)
{plot(covratio,type="h",xlab="observation",ylab="COVRATIO")
abline(1,0)}
beob_rev(order(covratio))
covratio.ord_rev(sort(covratio))
return(list(covratio=covratio,covratio.ord=covratio.ord,beob=beob))}

```

Funktion *devred.nb*

```

devred.nb_function(r,dev,plot=T)
{X_r$x
n_dim(X)[1]
devred_rep(0,n)
for (i in 1:n)
{devred[i]_r$deviance-dev[i]}
if (plot)
{plot(devred,type="h",xlab="observation",ylab="decrease of deviance")
abline(4,0)
eb_rev(sort(devred[devred>4]))
b_rev(order(devred))[1:length(eb)]
for (i in 1:length(eb))
{t_paste(b[i])
text(b[i],eb[i],t,adj=0)}}
beob_rev(order(devred))
devred.ord_rev(sort(devred))
return(list(devred=devred,devred.ord=devred.ord,beob=beob))}

```


Beschreibung benutzerdefinierter Funktionen		
Bezeichnung Funktion	Zweck	(Formelnr.)
<i>diag.poi</i> <i>diag.nba</i> <i>diag.nb</i>	Die h_{ii} -Werte, die $DFFITS_i^1$ und die Cook's Distanz D_i^1 werden berechnet und können mittels eines Stabdiagramms veranschaulicht werden.	(5.3), (5.16), (5.23)
<i>dfbetas.poi</i> <i>dfbetas.nba</i> <i>dfbetas.nb</i>	Die $DFBETAS_{j,i}^1$ werden berechnet und können mittels eines Stabdiagramms veranschaulicht werden.	(5.17)
<i>covratio.poi</i> <i>covratio.nba</i> <i>covratio.nb</i>	Der $COVRATIO_i^1$ wird berechnet und kann mittels eines Stabdiagramms veranschaulicht werden.	(5.27)
<i>devred.poi</i> <i>devred.nba</i> <i>devred.nb</i>	Die Devianzreduktion ΔD_i wird berechnet und kann mittels eines Stabdiagramms veranschaulicht werden.	(5.33)
<i>inflbeob.hat</i> <i>inflbeob.dffits</i> <i>inflbeob.cook</i>	Die h_{ii} -Werte, die $DFFITS_i^1$ und die Cook's Distanz D_i^1 , welche die Schranken von $2p/n$, 2 und 1 überschreiten, werden von den Funktionen ausgegeben.	-

Tabelle E.1: Beschreibung der benutzerdefinierten Funktionen. Die Spalte Bezeichnung Funktion gibt jeweils den Funktionsnamen der Diagnostiken für die Poissonverteilung (.poi), Negativ-Binomial-Verteilung mit bekanntem α (.nba) und mit geschätztem α (.nb) an. Einzig die Funktionen in Zeile 5 sind für jede der drei Verteilungen zulässig. In Spalte Zweck wird beschrieben, was die Funktion ausgibt. In der Spalte Formelnr. werden die Formelnummern der Diagnostiken, die von der Funktion ausgerechnet werden, angegeben.

Anhang F

Aggregierter KFZ-Haftpflichtdatensatz

Hier ist der vollständige KFZ-Haftpflichtdatensatz, der in Kapitel 8 beschrieben und bearbeitet wird, aufgelistet.

Merit	Class	Insured	Premium	Claims	Cost
3	1	2757520	159108	217151	63191
3	2	130535	7175	14506	4598
3	3	247424	15663	31964	9589
3	4	156871	7694	22884	7964
3	5	64130	3241	6560	1752
2	1	130706	7910	13792	4055
2	2	7233	431	1001	380
2	3	15868	1080	2695	701
2	4	17707	888	3054	983
2	5	4039	209	487	114
1	1	163544	9862	19346	5552
1	2	9726	572	1430	439
1	3	20369	1382	3546	1011
1	4	21089	1052	3618	1281
1	5	4869	250	613	178
0	1	273944	17226	37730	11809
0	2	21504	1207	3421	1088
0	3	37666	2502	7565	2383
0	4	56730	2756	11345	3971
0	5	8601	461	1291	382

Literatur

- Bailey und Simon (1960). Two studies in automobile insurance ratemaking. *ASTIN Bulletin*, 192–217.
- Belsley, Kuh, und Welsch (1980). *Regression Diagnostics - Identifying Influential Data and Sources of Collinearity*. New York: Wiley.
- Bickel und Doksum (1977). *Mathematical Statistics - Basic Ideas and Selected Topics*. New Jersey: Prentice Hall.
- Cameron und Trivedi (1998). *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.
- Chatterjee und Hadi (1986). Influential observations, high leverage points and outliers in linear regression. *Statistical Science* 1(3), 379–416.
- Cook (1977). Detection of influential observations in linear regression. *Technometrics* 19, 15–18.
- Cook und Prescott (1981). On accuracy of bonferroni significance levels for detecting outliers in linear models. *Technometrics* 23, 59–63.
- Cook und Weisberg (1982). *Residuals and Influence in Regression*. New York: Chapman and Hall.
- Dobson (2002). *An Introduction to Generalized Linear Models* (2nd ed.). New York: Chapman and Hall.
- Draper und Smith (1998). *Applied Regression Analysis* (3rd ed.). New York: Wiley.
- Fahrmeir, Hamerle, und Tutz (1996). *Multivariate statistische Verfahren* (2nd ed.). Berlin: de Gruyter.
- Fahrmeir und Kaufmann (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Annals of Statistics* 13, 342–368.
- Fahrmeir, Kuenstler, Pigeot, und Tutz (2000). *Statistik - Der Weg zur Datenanalyse* (3rd ed.). Berlin: Springer.
- Fox (1984). *Linear Statistical Models and Related Methods with Applications to Social Research*. New York: Wiley.
- Hinkley (1977). Jackknifing in unbalanced situations. *Technometrics* 19, 285–292.
- Hinkley, Reid, und Snell (1991). *Statistical Theory and Modelling*. London: Chapman and Hall.
- Lawless (1987). Negative binomial and mixed regression. *The Canadian Journal of Statistics* 15(3), 209–225.

- McCullagh und Nelder (1989). *Generalized Linear Models* (2nd ed.). London: Chapman und Hall.
- Myers (1990). *Classical and Modern Regression with Applications* (2nd ed.). Belmont: Duxbury Press.
- Pregibon (1981). Logistic regression diagnostics. *The Annals of Statistic* 9, 705–724.
- Rade und Westergren (1997). *Springers Mathematische Formeln* (2nd ed.). Berlin: Springer.
- Rao (1973). *Linear Statistical Inference and Its Applications* (2nd ed.). New York: Wiley.
- Resnick (1992). *Adventures in Stochastic Processes with Illustrations*. Boston: Birkhaeuser.
- Seber (1977). *Linear Regression Analysis*. New York: Wiley.
- Williams (1987). Generalized linear model diagnostics using the deviance and single case deletions. *Applied Statistics* 36(2), 181–191.

Abbildungsverzeichnis

3.1	Grafiken von Ausreißern und einflußreichen Beobachtungen	14
5.1	Explorative Datenanalyse des Patentdatensatzes	52
5.2	Devianz- und Pearson-Residuen des Poissonmodells für den Patentdatensatz . . .	57
5.3	Stabdiagramm der h_{ii} -Werte des Poissonmodells für den Patentdatensatz	59
5.4	Stabdiagramm der Differenz von Ein-Schritt-Approximation $\hat{\beta}_{-i}^1$ und exaktem Wert $\hat{\beta}_{-i}$ des Poissonmodells für den Patentdatensatz	62
5.5	Stabdiagramm der $DFFITS_i^1$ und der $DFFITS_i$ des Poissonmodells für den Patentdatensatz	66
5.6	Stabdiagramme der $DFBETAS_{j,i}^1$ und der $DFBETAS_{j,i}$ des Poissonmodells für den Patentdatensatz	67
5.7	Stabdiagramm der Cook's Distanz D_i^1 und D_i des Poissonmodells für den Patentdatensatz	70
5.8	Stabdiagramm des $COVRATIO_i^1$ und $COVRATIO_i$ des Poissonmodells für den Patentdatensatz	73
5.9	Stabdiagramm der Werte der Devianzreduktion ΔD_i^1 und ΔD_i des Poissonmodells für den Patentdatensatz	77
5.10	Influenzkurven des Poissonmodells für den Patentdatensatz	82
5.11	Influenzkurven des Poissonmodells für den Patentdatensatz mit Skalierung ($ \cdot \leq 10$) der z-Achse	82
7.1	h_{ii} -Werte des Negativ-Binomial-Modells mit festem α für den Patentdatensatz für verschiedene Werte von α	96
7.2	$DFFITS_i^1$ und $DFFITS_i$ des Negativ-Binomial-Modells mit festem α für den Patentdatensatz für verschiedene Werte von α	97
7.3	Cook's Distanz D_i^1 und D_i des Negativ-Binomial-Modells mit festem α für den Patentdatensatz für verschiedene Werte von α	97
7.4	$COVRATIO_i^1$ und $COVRATIO_i$ des Negativ-Binomial-Modells mit festem α für den Patentdatensatz für verschiedene Werte von α	99
7.5	Werte der Devianzreduktion ΔD_i^1 und ΔD_i des Negativ-Binomial-Modells mit festem α für den Patentdatensatz für verschiedene Werte von α	99
7.6	Devianz- und Pearson-Residuen des Negativ-Binomial-Modells mit geschätztem α für den Patentdatensatz	103
7.7	Stabdiagramme der h_{ii} -Werte, der $DFFITS_i$ und der Cook's Distanz D_i des Negativ-Binomial-Modells mit geschätztem α für den Patentdatensatz	104
7.8	Stabdiagramme der $DFBETAS_{j,i}$ des Negativ-Binomial-Modells mit geschätztem α für den Patentdatensatz	106

7.9 Stabdiagramm des $COVRATIO_i$ des Negativ-Binomial-Modells mit geschätztem α für den Patentdatensatz 107

7.10 Stabdiagramm der Werte der Devianzreduktion des Negativ-Binomial-Modells mit geschätztem α für den Patentdatensatz 109

8.1 Explorative Datenanalyse des KFZ-Datensatzes 113

8.2 Devianz- und Pearson-Residuen des Poissonmodells für den KFZ-Datensatz . . . 115

8.3 Stabdiagramm der h_{ii} -Werte des Poissonmodells für den KFZ-Datensatz 116

8.4 Stabdiagramm der $DFFITS_i^1$ und der $DFFITS_i$ des Poissonmodells für den KFZ-Datensatz 117

8.5 Stabdiagramm der Cook's Distanz D_i^1 und D_i des Poissonmodells für den KFZ-Datensatz 117

8.6 Stabdiagramm der $DFBETAS_{j,i}^1$ und $DFBETAS_{j,i}$ des Poissonmodells für den KFZ-Datensatz 119

8.7 Stabdiagramm der $DFBETAS_{j,i}^1$ und $DFBETAS_{j,i}$ des Poissonmodells für den KFZ-Datensatz 120

8.8 Stabdiagramm des $COVRATIO_i^1$ und $COVRATIO_i$ des Poissonmodells für den KFZ-Datensatz 121

8.9 Stabdiagramm der Werte der Devianzreduktion ΔD_i^1 und ΔD_i des Poissonmodells für den KFZ-Datensatz 122

8.10 h_{ii} -Werte des Negativ-Binomial-Modells mit festem α für den KFZ-Datensatz für verschiedene Werte von α 124

8.11 $DFFITS_i^1$ und $DFFITS_i$ des Negativ-Binomial-Modells mit festem α für den KFZ-Datensatz für verschiedene Werte von α 124

8.12 Cook's Distanz D_i^1 und D_i des Negativ-Binomial-Modells mit festem α für den KFZ-Datensatz für verschiedene Werte von α 125

8.13 $COVRATIO_i^1$ und $COVRATIO_i$ des Negativ-Binomial-Modells mit festem α für den KFZ-Datensatz für verschiedene Werte von α 126

8.14 Werte der Devianzreduktion ΔD_i^1 und ΔD_i des Negativ-Binomial-Modells mit festem α für den KFZ-Datensatz für verschiedene Werte von α 126

8.15 Devianz- und Pearson-Residuen des Negativ-Binomial-Modells mit geschätztem α für den KFZ-Datensatz 130

8.16 Stabdiagramm der h_{ii} -Werte, $DFFITS_i$ und Cook's Distanz D_i des Negativ-Binomial-Modells mit geschätztem α für den KFZ-Datensatz 131

8.17 Stabdiagramm der $DFBETAS_{j,i}$ des Negativ-Binomial-Modells mit geschätztem α für den KFZ-Datensatz 133

8.18 Stabdiagramm der Werte des $COVRATIO_i$ des Negativ-Binomial-Modells mit geschätztem α für den KFZ-Datensatz 134

8.19 Stabdiagramm der Werte der Devianzreduktion ΔD_i des Negativ-Binomial-Modells mit geschätztem α für den KFZ-Datensatz 135

8.20 Differenz der y_i und der $\hat{\mu}_i(Poi)$ bzw. der y_i und der $\hat{\mu}_i(NB\hat{\alpha})$ 137

8.21 Differenz der y_i/t_i und der $\hat{\mu}_i(Poi)/t_i$ bzw. der y_i/t_i und der $\hat{\mu}_i(NB\hat{\alpha})/t_i$ 137

Tabellenverzeichnis

3.1	Übersicht der betrachteten Diagnostiken für LM's	29
4.1	Charakteristiken der Poisson- und Negativ-Binomial-Verteilung in der exponen- tiellen Familie	39
5.1	<i>summary</i> des Poissonmodells für den Patentdatensatz	53
5.2	h_{ii} -Werte des Poissonmodells für den Patentdatensatzes	58
5.3	h_{ii} -Werte mit $h_{ii} \geq 2p/n$ des Poissonmodells für den Patentdatensatz	59
5.4	$DFFITs_i^1$ des Poissonmodells für den Patentdatensatz	66
5.5	$DFFITs_i^1$ mit $DFFITs_i^1 \geq 2$ des Poissonmodells für den Patentdatensatz	67
5.6	Cook's Distanz D_i^1 des Poissonmodells für den Patentdatensatz	70
5.7	Cook's Distanz D_i^1 mit $D_i^1 > 1$ des Poissonmodells für den Patentdatensatz	70
5.8	Geordnete Werte des $COVRATIO_i^1$ des Poissonmodells für den Patentdatensatz	73
5.9	Geordnete Werte der Devianzreduktion ΔD_i^1 des Poissonmodells für den Patent- datensatz	77
5.10	Übersicht der betrachteten Diagnostiken für GLM's	78
7.1	Beobachtungen mit $h_{ii} \geq 2p/n$ des Negativ-Binomial-Modells mit festem α für den Patentdatensatz	100
7.2	Beobachtungen mit $DFFITs_i^1 \geq 2$ bzw. $DFFITs_i \geq 2$ des Negativ-Binomial- Modells mit festem α für den Patentdatensatz	100
7.3	Beobachtungen mit $D_i^1 > 1$ bzw. $D_i > 1$ des Negativ-Binomial-Modells mit festem α für den Patentdatensatz	101
7.4	Beobachtungen mit $\Delta D_i^1 \geq 4$ bzw. $\Delta D_i \geq 4$ des Negativ-Binomial-Modells mit festem α für den Patentdatensatz	101
7.5	<i>summary</i> des Negativ-Binomial-Modells mit geschätztem α für den Patentdatensatz	102
7.6	h_{ii} -Werte, $DFFITs_i$ und Cook's Distanz D_i des Negativ-Binomial-Modells mit geschätztem α für den Patentdatensatz	105
7.7	Geordnete Werte des $COVRATIO_i$ des Negativ-Binomial-Modells mit geschätz- tem α für den Patentdatensatz	107
7.8	Geordnete Werte der Devianzreduktion des Negativ-Binomial-Modells mit geschätz- tem α für den Patentdatensatz	109
8.1	<i>summary</i> des Poissonmodells für den KFZ-Datensatz	114
8.2	h_{ii} -Werte, $DFFITs_i^1$ und Cook's Distanz D_i^1 des Poissonmodells für den KFZ- Datensatz	116
8.3	Geordnete Werte des $COVRATIO_i^1$ des Poissonmodells für den KFZ-Datensatz	121

8.4	Geordnete Werte der Devianzreduktion ΔD_i^1 des Poissonmodells für den KFZ-Datensatz	122
8.5	Beobachtungen mit $h_{ii} \geq 2p/n$ des Negativ-Binomial-Modells mit festem α für den KFZ-Datensatz	127
8.6	Beobachtungen mit $DFFITs_i^1 \geq 2$ bzw. $DFFITs_i \geq 2$ des Negativ-Binomial-Modells mit festem α für den KFZ-Datensatz	127
8.7	Beobachtungen mit $D_i^1 > 1$ bzw. $D_i > 1$ des Negativ-Binomial-Modells mit festem α für den KFZ-Datensatz	128
8.8	Beobachtungen mit $\Delta D_i^1 \geq 4$ bzw. $\Delta D_i \geq 4$ des Negativ-Binomial-Modells mit festem α für den KFZ-Datensatz	128
8.9	<i>summary</i> des Negativ-Binomial-Modells mit geschätztem α für den KFZ-Datensatz	129
8.10	h_{ii} -Werte, $DFFITs_i$ und Cook's Distanz D_i des Negativ-Binomial-Modells mit geschätztem α für den KFZ-Datensatz	131
8.11	Geordnete Werte des $COVRATIO_i$ des Negativ-Binomial-Modells mit geschätztem α für den KFZ-Datensatz	134
8.12	Geordnete Werte der Devianzreduktion ΔD_i des Negativ-Binomial-Modells mit geschätztem α für den KFZ-Datensatz	135
8.13	Vergleich der Schadenzahlen mit den geschätzten Erwartungswerten des Poissonmodells und des Negativ-Binomial-Modells mit geschätztem α	136
8.14	Vergleich der Schadenzahlen mit den geschätzten, gewichteten Erwartungswerten des Poissonmodells und des Negativ-Binomial-Modells mit geschätztem α	136
E.1	Beschreibung der benutzerdefinierten Funktionen	157