

Carl von Ossietzky
Universität Oldenburg

Diplomstudiengang Mathematik

DIPLOMARBEIT

Modellwahl für die Analyse longitudinaler Daten
einer Forschungsstudie des visuellen Systems

vorgelegt von: Veronika Weyer

Betreuende Gutachterin: Prof. Dr. Christine Müller

Zweiter Gutachter: Prof. Dr. Peter Harmand

Oldenburg, 25. November 2008

Inhaltsverzeichnis

| | | |
|----------|--|-----------|
| 1 | Einleitung | 1 |
| 2 | Physiologische Grundlagen und die Forschungsstudie | 3 |
| 2.1 | Die Spezifische Absorbtionsrate | 4 |
| 2.2 | Physiologische Grundlagen | 5 |
| 2.2.1 | Retina und Ganglienzellen | 5 |
| 2.2.2 | Signalaufnahme in der Retina | 8 |
| 2.3 | Der Versuch | 9 |
| 2.3.1 | Zielsetzungen | 10 |
| 2.3.2 | Methoden und Ergebnisse | 12 |
| 3 | Statistische Grundlagen | 17 |
| 3.1 | Verteilungsfunktionen | 17 |
| 3.2 | Wichtige Definitionen | 19 |
| 4 | Lineares Modell und Modelle für longitudinale Daten | 23 |
| 4.1 | Das allgemeine lineare Modell für den multivariaten Fall | 23 |
| 4.2 | Subjektspezifische Modelle | 26 |
| 4.3 | Das lineare gemischte Modell | 27 |
| 4.4 | Das verallgemeinerte lineare Modell (GLM) | 28 |
| 4.5 | Das Marginalmodell | 30 |
| 4.6 | Der normalverteilte Fall | 32 |
| 5 | Das verallgemeinerte lineare gemischte Modell (GLMM) | 35 |
| 5.1 | Maximum-Likelihood-Funktion und deren Annäherung | 41 |
| 5.2 | Penalized-Quasi-Likelihood-Schätzung (PQL) | 48 |
| 5.3 | Marginale-Quasi-Likelihood-Schätzung (MQL) | 56 |
| 5.3.1 | Vergleich von PQL und MQL | 57 |
| 5.4 | Tests beim GLMM (Testen der Varianzkomponente) | 59 |
| 5.5 | Modellwahl | 63 |
| 6 | Schätzmethoden für das GLMM: EBP und GEE | 66 |
| 6.1 | Schätzungen der Zufallseffekte, (EBP) | 66 |
| 6.2 | „Small area estimation“ unter dem GLMM | 67 |
| 6.3 | GEE als Schätzmethode für die festen Effekte | 73 |

| | |
|---|----------------|
| 7 Analyse von longitudinalen Daten | 87 |
| 7.1 Longitudinale Daten | 87 |
| 7.2 Die Analyse und ein Beispiel | 88 |
| 7.3 Verschiedene Ansätze der Analyse longitudinaler Daten | 95 |
| 7.4 Maximum-Likelihood-Schätzung bei longitudinalen Daten | 98 |
| 7.5 Missing Values bei longitudinalen Daten | 107 |
| 8 Auswertung der Forschungsstudie | 111 |
| 8.1 Auswertung anhand der SAS-Prozedur „GLIMMIX“ | 113 |
| 8.1.1 Auswertung der UMTS Standard Versuche | 113 |
| 8.1.2 Auswertung der GSM 900 MHz Versuche | 119 |
| 8.1.3 Auswertung der GSM 1800 MHz Versuche | 122 |
| 8.2 Diskussion der Ergebnisse | 125 |
| 9 Zusammenfassung und Schlussbemerkungen | 128 |
| A Anhang 1: SAS Anweisungen für die MEANS-Prozedur | 129 |
| B Anhang 2: Versuchsdesign Forschungsstudie | 134 |
| C Anhang 3: SAS Output für den Dunnett-Test | 137 |
| D Anhang 4: SAS Output für die GLIMMIX-Prozedur | 142 |

Abbildungsverzeichnis

| | | |
|---|--|-----|
| 1 | Der Aufbau der Retina | 6 |
| 2 | Das Aktionspotential | 9 |
| 3 | Datenkurve von Gruppe 1 | 93 |
| 4 | Datenkurve von Gruppe 1 | 94 |
| 5 | Temperaturkurve bei UMTS Standard | 115 |
| 6 | ON Antwortrate des Quotienten bei UMTS Standard bei 16lx . | 116 |
| 7 | Temperaturkurve bei GSM 900 | 119 |
| 8 | Temperaturkurve bei GSM 1800 | 123 |

Tabellenverzeichnis

| | | |
|----|--|-----|
| 1 | Versuchsdesign | 12 |
| 2 | Zwei Gruppen Experiment | 92 |
| 3 | Unsystematisch fehlende Daten | 109 |
| 4 | Temperaturen bei UMTS Standard | 115 |
| 5 | Signifikanzen des Quotienten B1 bei UMTS Standard | 118 |
| 6 | Signifikanzen des Quotienten B2 bei UMTS Standard | 118 |
| 7 | Signifikanzen der Differenzen B1 bei UMTS Standard | 118 |
| 8 | Signifikanzen der Differenzen B2 bei UMTS Standard | 119 |
| 9 | Signifikanzen des Quotienten B1 bei GSM 900 | 121 |
| 10 | Signifikanzen des Quotienten B2 bei GSM 900 | 121 |
| 11 | Signifikanzen der Differenzen B1 bei GSM 900 | 121 |
| 12 | Signifikanzen der Differenzen B2 bei GSM 900 | 121 |
| 13 | Signifikanzen des Quotienten B1 bei GSM 1800 | 123 |
| 14 | Signifikanzen des Quotienten B2 bei GSM 1800 | 124 |
| 15 | Signifikanzen der Differenzen B1 bei GSM 1800 | 124 |
| 16 | Signifikanzen der Differenzen B2 bei GSM 1800 | 124 |

1 Einleitung

In dieser Arbeit wird die Modellwahl für longitudinale Daten thematisiert. In klinischen oder biologischen Studien werden häufig pro Patient, bzw. pro Versuchsperson oder Versuchsobjekt Daten zu verschiedenen Zeitpunkten erhoben. Genau dann ist von longitudinalen Daten oder auch longitudinalen Beobachtungen die Rede. Die zugehörigen Studien werden Longitudinal,- bzw. Längsschnittstudien genannt. Longitudinale Daten, welche korrelierte Daten, also zusammenhängende Daten sind, spielen somit eine wichtige Rolle, wenn zeitliche Entwicklungen und Veränderungen von bestimmten Größen untersucht werden.

Ein Beispiel für longitudinale Daten ist, wenn der Blutdruck bei mehreren Personen wiederholt über der Zeit gemessen wird.

Solche gruppierten Daten benötigen spezielle Analysemethoden, wobei vor allem die Frage nach einem geeigneten statistischen Modell für solch eine Art von Daten im Blickpunkt dieser Arbeit stehen soll. Für solche Wiederholungsmessungen gibt es viele Modellierungsmöglichkeiten und folglich viele Modelle, mit denen longitudinale Daten ausgewertet werden können.

Auf eines dieser Modelle wird in dieser Arbeit ein besonderes Augenmerk gelegt, das GLMM (**G**eneralized **L**inear **M**ixed **M**odel) oder auch das verallgemeinerte lineare gemischte Modell.

Lineare gemischte Modelle, worunter ebenfalls das GLMM fällt, enthalten neben den festen Effekten β zusätzlich Zufallseffekte α . Solche Zufallseffekte sind häufig in longitudinalen Studien vorhanden. In einem derartigen Fall führt das zu dem Schluss, dass das einfache lineare Modell, sowie das verallgemeinerte lineare Modell, welche keine Zufallseffekte enthalten, keine Anwendung finden. Bevor das GLMM in Kapitel (5) weiter vertieft wird, werden in Kapitel (2) zunächst die physiologischen Grundlagen und eine Forschungsstudie aus der Biologie, in welcher longitudinale Daten erhoben werden, vorgestellt. Die Forschungsstudie, in der es um die Auswirkungen von Handystrahlen auf die retinalen Ganglienzellen geht, wird in dieser Arbeit einer besonderen Betrachtung unterliegen, indem anhand dieser Studie eine statistische Analyse durchgeführt wird.

Im darauf folgenden Kapitel (3) werden dann die für den weiteren Verlauf der Arbeit wichtigen statistischen Definitionen gegeben. Wie gerade schon erwähnt, gibt es für longitudinale Daten viele verschiedene Modellierungsmöglichkeiten. Die verschiedenen Modelle, die bei einer Analyse von longitudinalen Studien zu Anwendung kommen können, werden in Kapitel (4) näher erläutert. Zu diesen Modellen zählen vor allem die marginalen,- sowie auch die subjekt-spezifischen Modelle.

In den darauf folgenden Abschnitten der Arbeit wird das GLMM als ein geeignetes Modell für die Analyse einer Längsschnittstudie dargestellt. Dabei geht es in erster Linie darum die Maximum-Likelihood-Schätzungen für das verallgemeinerte lineare gemischte Modell herauszuarbeiten.

Da es in solch einem Modell meist sehr kompliziert ist diese Schätzungen auf einem normalen Weg zu bestimmen, gibt es approximative Methoden, worunter die Quasi-Likelihood-Schätzungen (QL Schätzungen) fallen, um zu konsistenten Schätzungen zu gelangen. Zu den QL Schätzungen zählen die „Penalized-Quasi-Likelihood-Schätzung“ von [(Green 1987)] und die „Marginale-Quasi-Likelihood-Schätzung“.¹

Neben den Maximum-Likelihood-Schätzungen, zum Schätzen der festen Effekte β , wird ebenfalls die Methode der verallgemeinerten Schätzgleichungen im Blickpunkt der Arbeit stehen.

Diese Methode wurde von [(Liang und Zeger 1986)] eingeführt und führt ebenfalls, wie die Maximum-Likelihood-Schätzungen zu Schätzern für die festen Effekte. Neben den festen Effekten ist es natürlich auch von Interesse Schätzungen für die Zufallseffekte α zu finden. Dazu wird die Methode der „small area estimation“ bzw. die „empirical best prediction“ präzisiert.

Nachdem im Anschluss in Kapitel (7) die Analyse von longitudinalen Daten dargestellt wurde, werden im folgenden und letzten Kapitel (8) dieser Arbeit die Daten der Forschungsstudie ausgewertet. Die Auswertung wird anhand der Statistiksoftware SAS² in erster Linie mit der zugehörigen „GLIMMIX“-Prozedur zum Auswerten von verallgemeinerten linearen gemischten Modellen durchgeführt.

Abschließend wird noch eine kurze Diskussion der Ergebnisse der Auswertung und eine Zusammenfassung der in der Arbeit vorgestellten Methoden folgen.

¹siehe: [(Molenberghs und Verbeke 2005)]

²SAS=Statistical Analysis System

2 Physiologische Grundlagen und die Forschungsstudie

Heutzutage sind elektromagnetische Felder aus unserem Alltag nicht mehr weg zu denken und man kann sagen, dass sich ein Großteil der Bevölkerung an die Kenntnis des Vorhandenseins von elektromagnetischer Strahlung, welche von Radio- und Fernsehfrequenzen ausgehen, gewöhnt hat. Die technische Entwicklung der letzten Jahre bei der Mobilfunkkommunikation hat allerdings zu einer breiten Sensibilisierung bezüglich eventueller gesundheitlicher Gefahren durch Mobilfunkfelder geführt.

Bei den quantitativ häufigsten Strahlenquellen handelt es sich um hochfrequente elektromagnetische Felder der Mobilfunkbasisstationen und Mobilfunktelefone.

Da heutzutage nahezu alle Menschen ein Handy nutzen, ist eine sofort folgende Konsequenz, dass auch die Anzahl der Basisstationen erhöht werden muss. Dieses führt wiederum dazu, dass wir Menschen natürlich einer immer größeren Strahlung und einer wachsenden Angst vor den Auswirkungen solcher Strahlungen ausgesetzt werden. In den Medien werden meistens nur die Gefahren, welche von den Mobilfunkbasisstationen ausgehen, dargestellt. Was allerdings vom wissenschaftlichen Standpunkt genauso wichtig ist, ist die Belastung durch die Mobilstationen, bzw. durch Handys.

Die Aussendung (Emmission) der Strahlen während des Betriebs findet meist in der Nähe des Kopfes und somit in der Nähe der wichtigen Sinnesorgane Auge und Ohr statt. Eine Konsequenz davon ist, dass an solchen Stellen starke hochfrequente elektromagnetische Felder (HF-EMF) und SAR-Werte (siehe nächster Abschnitt) aufgebaut werden.

Genau aus diesem Grund ist es von großer Bedeutung weitere Untersuchungen des Einflusses hochfrequenter Felder der Mobilfunktechnologie auf das Nervensystem, und hierbei insbesondere auf die Sinnesorgane, zu machen, um Antworten auf die Fragen zu finden, ob elektromagnetische Felder beispielsweise Ursache für

- Kanzerogenese (Krebs),

- subjektiv erfahrene Beschwerden (z.B. Kopfschmerzen, Müdigkeit und Schlafstörungen),
- Beeinflussung kognitiver Leistungen (z.B. Wahrnehmungsstörungen),
- erhöhten Blutdruck oder Funktionsstörungen elektrischer Implantate

sind.

2.1 Die Spezifische Absorptionsrate

Die gesundheitsschädigende Wirkung von Handystrahlen verunsichert viele Verbraucher und es stellt sich zunehmend die Frage, was man dagegen tun kann, bzw. was man beim Kauf eines Handys beachten muss. Ein wichtiger Anhaltspunkt bei dem Kauf eines Handys mit nicht allzu gesundheitsschädlicher Wirkung ist die spezifische Absorptionsrate (bzw. der SAR-Wert), welche mittlerweile von Handy anbietern bei jedem Handy angegeben wird und einen Grenzwert von 2 Watt pro Kilogramm nicht überschreiten darf, um keine größeren gesundheitlichen Schäden hervorzurufen.

Zunächst wird kurz erklärt, was ein SAR-Wert ist, um einen kleinen Überblick in die Materie zu bekommen:

SAR ist die Abkürzung für die spezifische Absorptionsrate. Die zugehörige Einheit ist $\frac{W}{Kg} = \frac{Leistung}{Masse}$. Die spezifische Absorptionsrate ist eine physikalische Größe und gibt das Maß für die Absorption von elektromagnetischen Feldern in biologischem Gewebe an. Die Absorption der elektromagnetischen Felder führt in den meisten Fällen zu einer Erwärmung, also zu einem thermischen Effekt, des jeweiligen Gewebes. In manchen Fällen ist es je nach Feldstärke und Frequenz sogar möglich, dass Effekte, wie beispielsweise das Auslösen von Aktionspotentialen, auftreten. Es kann zwar vorkommen, dass sich die Effekte unterschiedlich bemerkbar machen, jedoch bleibt die Ursache im entferntesten Sinne immer die Gleiche: Die Erwärmung.

Physikalisch kann die Spezifische Absorptionsrate folgendermaßen dargestellt werden:

1. der Feldstärke im Gewebe:

$$SAR = \frac{\sigma |\vec{E}|^2}{\rho}$$

2. der Stromdichte im Gewebe:

$$SAR = \frac{J^2}{\rho\sigma}$$

3. und der Temperaturerhöhung im Gewebe:

$$SAR = c_i \frac{dT}{dt}$$

dabei sind:

- E = Elektrische Feldstärke
- J = Stromdichte
- ρ = Dichte des Gewebes
- σ = Elektrische Leitfähigkeit des Gewebes
- c_i = Wärmekapazität des Gewebes
- $\frac{dT}{dt}$ = Zeitliche Ableitung der Gewebetemperatur

Der SAR-Wert wird im Folgenden, vor allem in der Forschungsstudie und deren Auswertung, von großer Bedeutung sein.

2.2 Physiologische Grundlagen

In diesem Kapitel werden einige neurobiologische Grundlagen näher erläutert, welche für den weiteren Verlauf dieser Arbeit wichtig sind, vor allem, um verschiedene Aspekte der Forschungsstudie zu verstehen.

2.2.1 Retina und Ganglienzellen

Von großer Bedeutung für die Forschungsstudie sind die retinalen Ganglienzellen. Das sind Neurone, also Nervenzellen, welche in der Retina, der Netzhaut des Auges, liegen.

Die Netzhaut, auch Retina genannt, (lateinisch: rete=netz) ist eine Schicht von spezialisiertem Netzgewebe an der hinteren Innenseite des Auges.

Abbildung (1) stellt den Aufbau einer Retina dar. Es sind die verschiedenen Schichten und Zelltypen einer Säugetierretina zu sehen, dabei sind: RPE=

retinales Pigmentepithel, OS= Außensegmente der Photorezeptorzellen, IS= Innensegmente der Photorezeptorzellen, ONL= äußere nukleäre Schicht, OPL= äußere plexiforme Schicht, INL= innere nukleäre Schicht, IPL= innere plexiforme Schicht, GC= Ganglienzellschicht, P= Pigmentepithelzelle, R= Stäbchen, C= Zapfen, H= Horizontalzelle, B= Bipolarzelle, M= Müller-Zelle, A= Amakrine-Zelle, G= Ganglienzelle, AX= Axone; und der Pfeil zeigt die Membrane limitans externa an.

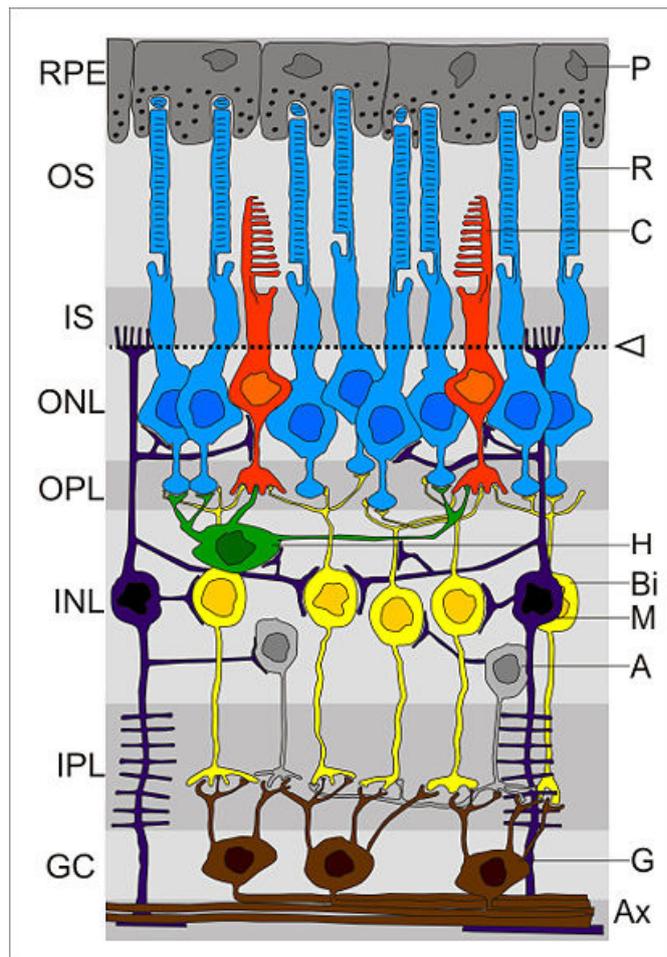


Abbildung 1: Der Aufbau der Retina

Nachdem das auftreffende Licht Hornhaut, Linse und Glaskörper durchquert hat, wird es in der Retina in Nervenimpulse umgewandelt. Die Netzhaut besteht, neben dem lichtempfindlichen Gewebsanteil, aus Nervenzellen zur Verarbeitung und Weiterleitung der erzeugten Impulse, sowie aus verschiedenen

Unterstützungsstrukturen zur Aufrechterhaltung der Funktion der reizerzeugenden, und -verarbeitenden Zellen.

Die Nervenzellen der Netzhaut lassen sich in drei Gruppen einteilen:

1. Die lichtempfindlichen oder fotorezeptiven Zellen, welche das eintreffende Licht in Nervenimpulse umwandeln. Dazu gehören die *Stäbchen* und *Zapfen* und eventuell weitere Zelltypen, welche nicht weiter erklärt werden, da sie hier nicht von Bedeutung sind.
2. Die zwischengeschalteten Zellen oder auch Interneurone, welche eine erste Verarbeitung an den erzeugten Impulsen innerhalb der Netzhaut vornehmen. Zu dieser Gruppe gehören die *Horizontalzellen*, die *Bipolarzellen* und die *Amakrinen Zellen*. (Auch diese werden nicht weiter erläutert.)
3. Die *retinalen Ganglienzellen*, welche die verarbeiteten Informationen an die nächste Schaltstelle außerhalb der Netzhaut weiterleiten. (Diese werden, wie eben schon erwähnt, im Folgenden von großer Bedeutung sein.)

Auf die retinalen Ganglienzellen, wird nun etwas detaillierter eingegangen:

Die Ganglienzellen sind den Bipolar- und Amakrinzellen nachgeschaltet. Sie leiten die visuelle Information über den Sehnerv, den „*Nervus opticus*“, zur nächsten Schaltstation im Gehirn, dem „*Ganglion geniculatum laterale*“, weiter und sind somit die Ausgangsneuronen der Netzhaut. Insgesamt lassen sich etwa 20 Ganglienzelltypen unterscheiden. Wie auch Bipolar- und Amakrinzellen lassen sich die Ganglienzellen der Netzhaut in *ON- und OFF-Zentrum-Zellen* unterscheiden.

Bevor die beiden Typen von Ganglienzellen weiter erklärt werden, muss zuvor erläutert werden, was ein rezeptives Feld ist.

Ein *rezeptives Feld (RF)* eines visuellen Neurons wird derjenige Teil des Gesichtsfeldes genannt, der auf Grund von einer adäquaten Stimulation zu einer Aktivitätsänderung des Neurons führt.³

Die *ON-Zentrum-Ganglienzellen* reagieren auf Belichtung des rezeptiven Feldes mit einer Aktivierung, auf Verdunklung hingegen mit einer Hemmung. Wenn nicht das RF- Zentrum, sondern die RF-Peripherie betrachtet wird, bemerkt man, dass die Antwort eines Reizes auf die Peripherie genau spiegelbildlich zur Antwort des Zentrums verläuft.

³siehe [(Schmidt,Lang,Thews 2005)]

Die *OFF-Zentrum-Ganglienzellen* verhalten sich genau entgegengesetzt der ON-Zentrum-Ganglienzellen, was bedeutet, dass eine Belichtung des rezeptiven Feldes eine Hemmung bewirkt, sowie eine Verdunklung eine Aktivierung. Nebeneinander liegende retinale Ganglienzellen besitzen benachbarte oder sogar überlappende rezeptive Felder und projizieren immer zu benachbarten Neuronen der nächst höheren Stufe. Diese Ordnung bleibt von der Rezeptorebene bis hin in die höheren Verarbeitungsebenen im *Kortex*, also in die Großhirnrinde (äußere Rindenschicht des Gehirns, in der sich die Nervenzellen befinden), so weit es geht erhalten.

2.2.2 Signalaufnahme in der Retina

Die elektrische Informationsverarbeitung im Nervensystem, also auch in der Retina, findet mit *graduierten Potentialen* und mit *Aktionspotentialen*, welche auch *Spikes* genannt werden, statt. Für die Übermittlung von weiten Strecken sind vor allem die Aktionspotentiale verantwortlich.

Um den Vorgang eines Aktionspotential besser verstehen zu können, muss man wissen, dass der *Intra- und Extrazellulärraum* (Innen- und Außenmedium) eine elektrische Spannung aufweist, das sogenannte Membranpotential, welches durch die unterschiedlichen Ionenkonzentrationen innerhalb und außerhalb der Zelle zustande kommt.

Das Ruhemembranpotential, also das Membranpotential bei einer nicht erregten Zelle, liegt bei ungefähr -70 mV. Bei Erregung eines Neurons werden ausgehend vom Ruhemembranpotential schnelle Potentialänderungen, eben die Aktionspotentiale, ausgelöst. An dieser Stelle ist es wichtig zu erwähnen, dass ein Aktionspotential erst dann ausgelöst werden kann sofern ein bestimmter Schwellenwert, der von Zelle zu Zelle unterschiedlich ausfallen kann, überschritten wird. (*Alles oder Nichts Gesetz*)

Ein Aktionspotential ist somit eine transiente Änderung der Membranspannung auf Werte bis zu 40 mV und kann in genau vier Phasen eingeteilt werden, die hier allerdings nicht ausführlich erläutert werden. Die folgende Abbildung (2) zeigt allerdings den Ablauf eines Aktionspotentials mit seinen vier Phasen.⁴

Im Folgenden Kapitel wird die Forschungsstudie mit ihren Zielsetzungen, Methoden und Ergebnissen kurz vorgestellt.

⁴Die vier Phasen eines Aktionspotentials sind: Initiationsphase, Depolarisation, Repolarisation und Nachhyperpolarisation; siehe: [(Schmidt,Lang,Thews 2005)] : „Physiologie des Menschen“

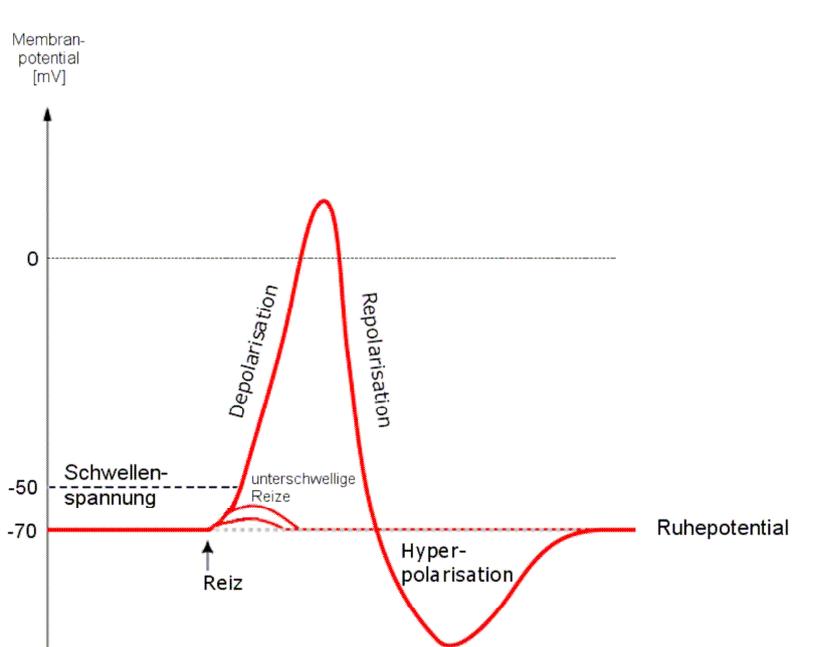


Abbildung 2: Das Aktionspotential

2.3 Der Versuch

In Kapitel (2) wurde erwähnt, dass wegen der zunehmenden Angst der Bevölkerung vor Handystrahlungen, weitere Untersuchungen des Einflusses hochfrequenter Felder der Mobilfunktechnologie auf das Nervensystem, und hier speziell auf die Sinnesorgane, nötig sind. In dem Versuch, der in dieser Arbeit näher betrachtet und ausgewertet wird, geht es um die Auswirkungen auf das Sinnesorgan Auge, also auf das visuelle System.

Die Forschungsstudie, welche von dem Institut für Biologie und Umweltwissenschaften der Universität Oldenburg unter der Leitung von Herrn Apl. Prof. Dr. Josef Ammermüller im Auftrag des Bundesamtes für Strahlenschutz durchgeführt wurde, befasst sich mit dem Einfluss hochfrequenter elektromagnetischer Felder der Mobilfunkkommunikation auf das visuelle System.⁵

Es wurden in der Vergangenheit zwar schon einige wenige Studien in dieser Richtung gemacht, jedoch zeigen die nur auf, dass thermische Effekte⁶ vorliegen, sie berücksichtigen allerdings keine weiteren nicht-thermischen Effekte auf das Auge.

In der Studie der Universität Oldenburg wurde der Einfluss hochfrequenter

⁵siehe: [(Abschlussbericht 2007)]

⁶wird im Folgenden noch näher drauf eingegangen

elektromagnetischer Felder auf die neuronale Aktivität retinaler Ganglienzellen in einer isolierten Retina einer Maus bei verschiedenen SAR-Werten untersucht, sowie bei unterschiedlichen Frequenzen des Mobilfunkstandards. Die Temperaturunterschiede wurden bei diesem Versuch möglichst minimiert. Dieser Punkt wird im folgenden Abschnitt noch näher behandelt.

Die Messergebnisse des Versuches wurden im Anschluss statistisch mit dem GLMM (siehe: Kapitel(5)) auf signifikante Haupteffekte und Zwei-Wege Interaktion getestet. In den Ganglienzellen wurden die Spikeraten und Latenzen der Antworten auf die Stimulation mit unterschiedlichen Lichtintensitäten gemessen.

Für die statistische Analyse wurden ebenfalls die Quotienten und die Differenzen der Antwortraten und Latenzen zu einer Kontrollgruppe berechnet. In erster Linie wurden Lichtintensität, Messreihenfolge⁷, Temperatur und natürlich der SAR-Wert als Effekte der Änderung von Antwortrate und Latenz (bzw. dessen Quotienten und Differenzen) untersucht. Das verallgemeinerte lineare gemischte Modell für diese Daten kann dargestellt werden durch die folgende Gleichung:

$$Y = X_i\beta + Z_i\alpha + \epsilon \quad (1)$$

wobei X_i und Z_i die Designmatrizen darstellen, β sind die sogenannten festen Effekte, die in dieser Studie die Lichtintensität, die Messreihenfolge, die Temperatur und den SAR-Wert beinhalten und α steht für den zufälligen Effekt, der hier die gemessene Ganglienzelle ist. Y ist die sogenannte Ziel-, bzw. Beobachtungsvariable und besteht hier aus allen gemessenen Antwort-, und Latenzraten, bzw. dessen Quotienten und Differenzen (zur einer Kontrollgruppe C^8). ϵ ist der Fehler. Wie genau das GLMM mit all seinen Einzelheiten aussieht, wird erst in Kapitel 5) erläutert.

2.3.1 Zielsetzungen

Für die Zielsetzungen des Versuches stellt die Retina ein optimales Testsystem dar, da sie ein neuronales Übertragungssystem ist, welches Lichtreize in elektrische Aktivität der retinalen Ganglienzellen umwandelt.⁹ Für das System der Retina gibt es eine oder mehrere räumlich-zeitliche Übertragungsfunktionen,

⁷Die Messreihenfolge wird in dieser Arbeit zur Vereinfachung weggelassen

⁸siehe nächster Abschnitt

⁹siehe: (Abschlussbericht 2007)

welche sich wie bei anderen Sinnessystemen aus der neuronalen Verarbeitung, hier innerhalb der Retina, ergeben. Diese Übertragungsfunktionen beziehen sich auf die Aktivität der Sehzellen und auf die Verarbeitung in den nachgeschalteten Interneuronen.

Sobald sich in einer Zelle etwas ändert, ändert sich ebenfalls die Übertragungsfunktion und folgedessen das Ausgangssignal der retinalen Ganglienzellen. Damit ist nun jede Änderung messbar egal wo die Änderung in der Netzhaut entsteht.

In den folgenden drei Punkten werden die Zielsetzungen des Forschungsprojekts erläutert:

- Die Forschung soll als eine bewertende Literaturstudie zu den möglichen Einflüssen von Hochfrequenzfeldern des Mobilfunks auf das Auge von Säugetieren wie auch von Menschen angesehen werden und deren Wirkungsmechanismus darstellen.
- Es ging darum die Frequenzen des Mobilfunkstandards nach
 - GSM 900 MHz (GSM = Global System for Mobile Communications)
 - GSM 1800 MHz
 - UMTS Standard (UMTS = Universal Mobile Telecommunication System)

auf die neuronale Aktivität der retinalen Ganglienzellen bei folgenden SAR-Werten:

- 0,02 W/kg
- 0,2 W/kg
- 2 W/kg
- 20 W/kg
- 0 W/kg (Scheinbefeldung)

zu testen, um im Folgenden die Ergebnisse festzuhalten und mögliche Auswirkungen zu entdecken.

- Als dritte Zielsetzung ging es in der Studie darum, Temperatureffekte, soweit wie möglich, auszuschließen, damit die Effekte der hochfrequenten Felder getrennt voneinander untersucht werden konnten.

2.3.2 Methoden und Ergebnisse

Die experimentellen Methoden werden im Folgenden kurz erläutert. Genauer kann im Abschlussbericht der Studie ([Abschlussbericht 2007]) nachgelesen werden. Die Ergebnisse der Forschungsstudie werden zu diesem Zeitpunkt zunächst sehr minimalistisch dargestellt. Die genaue Datenauswertung wird mit Hilfe des Statistikprogramms SAS erst im Hauptteil (Kapitel (8)) dieser Arbeit dargelegt.

In der Studie wurden die experimentellen Ableitungen und somit die elektrische Aktivität, also die Aktionspotentiale der retinalen Ganglienzellen, als Antwort auf eine Lichtreizung direkt anhand von Glasmikroelektroden gemessen. Die Ganglienzellen stammen von C57 BL/6N Mäusen, welche mit CO_2 betäubt und anschließend getötet wurden.

Zur Lichtstimulation und Befeldung wurden die Lichtantworten der Ganglienzellen vor, während und nach der Befeldung gemessen und im Anschluss mit den Lichtantworten der Scheinbefeldung verglichen.

Die Lichtreize erfolgten über eine Leutdiode und unterschieden sich in zwei Gruppen:

1. die konstanten Reize (K)
2. die Testreize (T)

Die konstanten Lichtblitze (K) waren dazu da, dass sich der Adaptationszustand der Retina beim laufenden Versuch nicht verändert.

Das Versuchsdesign zeigt Tabelle (1). Der Versuch ist also in folgende drei

| | | |
|-------------|---------------------------------|------------------|
| Messreihe C | Messreihen B1 und B2 und Pausen | Messreihen N1-N4 |
|-------------|---------------------------------|------------------|

Tabelle 1: Versuchsdesign

Phasen aufgeteilt:

1. Messreihe C vor Befeldung (Kontrollmessreihe)= 15 Minuten Ausgangszustand = 20K+T3+20K+T6 mit K= konstante Reize mit 40 ms Lichtblitzen mit 0,5 Hz und 2,66 lx retinaler Beleuchtungsstärke und den T= Testreizen mit 200 ms Lichtblitzen ebenfalls mit 0,5 Hz und mit aufsteigenden Intensitäten der Art, dass T3= Lichtblitze mit den 3 aufsteigenden Intensitäten: 0,5 lx; 16 lx; 445 lx und zehnmaliger Wiederholung und

T6= Lichtblitzen mit den 6 aufsteigenden Intensitäten:

0,07 lx; 0,5 lx; 2,66 lx; 16 lx; 83 lx; 445 lx mit 20-maliger Wiederholung.

2. Start der Befeldung: Dann 30 min Befeldung mit jeweils 2 Befeldungspausen von je 1 Minute; Messreihen B1 und B2 in den Befeldungspausen sind gleich 300K, T3, 300K, T3,300K, T und K wie eben. Das heisst also, ab Beginn der Befeldung erfolgen zunächst 300 konstante Reize, dann folgt in der Befeldungspause (1 min) die Testreihe T3 gefolgt von Befeldung mit wiederum 300 konstanten Reizen, dann kommt wieder eine Befeldungspause (1 min) mit T3 und zum Schluss wieder eine Befeldung mit 300K
3. Messreihen N1-N4 nach Befeldung (30 min)
=T3+20K+T6+130K+T3+20K+T6 zur Messung eventueller Nacheffekte und zur Kontrolle.

Der temporäre Ablauf des Experimentes kann somit in drei Phasen eingeteilt werden. Die erste Phase bestand aus 15 Minuten Ausgangszustand vor der Befeldung, in der die Kontrollmessreihe C durchgeführt wurde. Phase zwei bestand aus der 30 Minuten langen Befeldung zwischen denen an zwei Stellen jeweils eine Minute Expositionspause gemacht wurde und in Phase drei kam es anschließend zu einer Kontrollmessung von 30 Minuten, in der eventuelle Nacheffekte aufgezeigt werden konnten.

Es wurden für alle drei Versuchsreihen, also GSM 900, GSM 1800 und UMTS Standard, jeweils mehrere Versuche gemacht, von denen allerdings viele nicht mit in die statistische Auswertung eingehen konnten und verworfen werden mussten¹⁰. In den verschiedenen Versuchen wurden jeweils 5 retinale Ganglienzellen untersucht. Die verschiedenen Ganglienzellen (1-5) wurden jeweils bei den verschiedenen SAR-Werten gemessen. Bei jeder Ganglienzelle wurde jeweils eine komplette Intensitätsreihe untersucht, also 20K+T3+20K+T6, siehe oben. In Anhang (B) wird das gerade Erklärte anhand von einer Tabelle näher veranschaulicht.

Zur besseren Veranschaulichung siehe auch: [(Abschlussbericht 2007)]

Bei der Forschungsstudie war es von großer Bedeutung eine passende Temperaturregulation zu finden und somit wurde zur Vermeidung von eventuellen

¹⁰dazu mehr in Abschnitt 8.1

Temperaturveränderungen jeweils während der Befeldung über eine Temperaturregelung das Retinapräparat auf einer möglichst konstanten Temperatur gehalten. Eine gute Vermeidung von Änderungen der Temperatur wurde während des Experimentes dadurch erreicht, dass es außerhalb des Resonators zusätzlich Heizwiderstände gab, die dann den Wasserkreislauf, der durch den Probenhalter führte, entsprechend dem Sollwert, jeweils mehr oder weniger aufheizen konnten.¹¹ Zum Messen der Temperatur wurde ein sogenanntes Luxtron, ein fluoro-optischer Messfühler¹² verwendet, der in unmittelbarer Nähe der Retina plaziert wurde.

In der obigen kurzen Beschreibung des Versuches wurde bereits erwähnt, dass die Latenzen und Antwortraten gemessen wurden.

Um diese Latenzen und Antwortraten zu bestimmen, wurde ein Programm namens „*Wave-cluster*“¹³ verwendet. Bei diesem Verfahren wurde ein Algorithmus verwendet, welcher die gemessenen Aktionspotentiale, die eine ähnliche Form aufwiesen, in bestimmte Cluster einteilt, also verschiedenen Gruppen zuordnet. Nach dem Clustering wurden nur noch die Zeiten des Schwellendurchtritts betrachtet. Dabei wurden für jede einzelne Zelle die Aktionspotentiale, welche nach einem Lichtblitz aufgetreten sind, als sogenannte „*Peristimulus-Zeit-Histogramme (PSTH)*“ über die Zeit aufgetragen. Die Wiederholungen¹⁴ der verschiedenen Intensitäten wurden anschließend gemittelt und aus den eben erwähnten PSTH wurden dann die Latenzen und Antwortraten berechnet. Für die statistische Analyse wurden im Anschluss noch die Quotienten (relative Änderung) und die Differenzen (absolute Änderung) der verschiedenen Messreihen zur Kontrollmessreihe C berechnet. Dies war wichtig, da die Standardabweichungen ansonsten sehr große Werte, mit welchen schlecht gerechnet werden kann, angenommen hätten. Für genauere Erklärungen siehe [(Abschlussbericht 2007)].

Als nächstes wird kurz auf die Ergebnisse des Forschungsprojekts eingegangen.

¹¹siehe: [(Abschlussbericht 2007)] S.8

¹²Ein Luxtron, oder der sogenannte fluoro-optische Messfühler, ist bei der „berührenden“ Temperaturmessung von Bedeutung. Dieses Prinzip ist für Messungen da, die in schwieriger Umgebung, also an schwer erreichbaren Orten oder in starken, elektromagnetischen Feldern vorgenommen werden.

¹³„*Wave-cluster*“ ist eine bestimmte Art von Clusteranalyse. Es ist ein Gitter basiertes Clusteringverfahren. Für nähere Erläuterungen siehe: [(Grunau und Riemer 2005)]

¹⁴Es wurden 10 Wiederholungen bei den Testreizen T3 und 20 Wiederholungen bei T6 vorgenommen.

Als Ergebnisse der statistischen Auswertung (mit SPSS) kann man folgende drei Punkte festhalten:

1. Die stärksten Effekte, also die stärksten Änderungen der Aktivität der Neurone wurden durch die Intensität, die Messreihenfolge und die Temperatur verursacht.
2. Der SAR-Wert spielte eine geringere Rolle als Haupteffekt, als vorher vermutet wurde.
3. Es wurde ein signifikanter Effekt bei der Interaktion von Intensität und SAR-Wert festgestellt. Dieses kann dadurch erklärt werden, dass der Effekt durch eine mit der Befeldung zusammenhängenden Temperaturänderung hervorgerufen wurde und folglich die Aussage mit Vorsicht betrachtet werden sollte.

Zum dritten Punkt ist noch Folgendes hinzuzufügen:

Es ist schwer bei dieser Untersuchung direkte Effekte auf die Temperatur von anderen, indirekten Temperatureffekten zu unterscheiden. Kleinere Temperaturschwankungen des Körpers sind normal. Falls Temperaturänderungen nun mit der Aktivität von Neuronen interagieren, können signifikante Interaktionen auftreten, die im Folgenden zu signifikanten Aktivitätsänderungen führen. Neuronen reagieren bei der synaptischen Übertragung und bei der Erregungsbildung immer auf Temperaturschwankungen. Somit reagieren die retinalen Ganglienzellen auch bei einer Temperaturerhöhung mit einer Erhöhung der Antwortrate auf einen bestimmten Lichtreiz. Die Latenz nimmt in diesem Fall allerdings ab. Dieses Phänomen ist auf physikalisch-chemische Mechanismen zurückzuführen. Ebenfalls dafür verantwortlich sind die temperatursensitiven Moleküle, man nennt sie auch *TRPV-Kanäle* (diese wurden in der Retina aufgewiesen), da diese Moleküle Membrankanäle darstellen. Diese speziellen Kanäle besitzen eine intrazelluläre Bindungsstelle für *Capsaicin*¹⁵, sie werden allerdings (hier von Bedeutung) auch durch Temperaturveränderungen aktiviert.

Die Studie zeigt klar, dass Änderungen von $\pm 0,3 - 0,5$ Grad Celsius schon zu einer akuten und statistisch messbaren Änderung der Aktivität der Neurone

¹⁵Capsaicin (CPS) ist ein aus Pflanzen der Gattung Capsicum gewonnenes Alkaloid und wirkt auf spezifische Wärme-Rezeptoren, wodurch chemisch ein Hitze- oder Schmerzreiz ausgelöst wird.

führen können. Hier ist es allerdings schwierig zu sagen, ob die Effekte durch die Befeldung verursacht wurden. Wenn die Effekte tatsächlich von der Befeldung rühren folgt, dass eine geringe Erwärmung wünschenswert ist, daher sollten möglichst niedrige SAR-Werte angestrebt werden. Jedoch kann man sich hier nicht sicher sein, ob die Temperaturschwankungen wirklich etwas mit der Befeldung zu tun haben.

3 Statistische Grundlagen

Nachdem im letzten Abschnitt zunächst auf die biologischen Grundlagen und auf die Forschungsstudie, auf welche ein Teil dieser Arbeit basiert, eingegangen wurde, werden in diesem Kapitel die in der Arbeit vorkommenden statistischen Grundlagen vorgestellt.

3.1 Verteilungsfunktionen

Zunächst werden unter Anderem für das bessere Verständnis der Modelle, die im folgenden Kapitel vorgestellt werden, die einzelnen Verteilungen, die zur Klasse der exponentiellen Familie gehören, dargestellt.

Im Folgenden habe die Zufallsvariable Y immer die Realisierung y

Definition 3.1 (Eindimensionale Normalverteilung)

Die (eindimensionale) Normalverteilung $N(\mu, \sigma^2)$, wobei μ der Erwartungswert und σ^2 die Varianz ist, ist die Verteilung der Zufallsvariablen Y mit der zugehörigen Realisierung y mit folgender Dichtefunktion:

$$f_{(\mu, \sigma^2)}(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(y - \mu)^2}{2\sigma^2}\right) \quad (2)$$

Da Daten häufig multivariat sind, wird auch die multivariate Normalverteilung definiert:

Definition 3.2 (Multivariate Normalverteilung)

Ein n -dimensionaler Zufallsvektor Y besitzt eine multivariate Normalverteilung $N_n(\mu, \Sigma)$, wobei μ der Erwartungswertvektor und Σ die Kovarianzmatrix mit $\Sigma \in \mathbb{R}^{n \times n}$ ist, falls die Dichte $f(y)$ von Y die Form

$$f(y) = \frac{1}{(\sqrt{2\pi})^n \det(\Sigma)^{\frac{1}{2}}} \exp\left(\frac{-(y - \mu)^t \Sigma^{-1} (y - \mu)}{2}\right) \quad (3)$$

für $y \in \mathbb{R}^n$ hat.

Definition 3.3 (Binominalverteilung)

Eine Binominalverteilung mit den Parametern n und p ($B(n, p)$) ist gegeben,

falls die Dichtefunktion der Verteilung durch

$$f(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 1, \dots, n \quad (4)$$

definiert ist. Hier ist der Erwartungswert np und die Varianz $np(1-p)$.

Definition 3.4 (Poissonverteilung)

Die Verteilung $P(\lambda)$ heisst Poissonverteilung, sofern die Dichtefunktion durch

$$f(k) = P(X = k) = \exp(-\lambda) \frac{\lambda^k}{k!}, k = 0, 1, 2, \dots \quad (5)$$

gegeben ist. Der Erwartungswert und die Varianz sind bei der Poissonverteilung gleich dem Parameter λ .

Definition 3.5 (Exponentialverteilung)

Eine stetige Zufallsvariable Y mit der zugehörigen Realisierung y , die die Verteilung $E(\lambda)$ mit Erwartungswert $\frac{1}{\lambda}$ und Varianz $\frac{1}{\lambda^2}$ und der Dichtefunktion

$$f(y) = \lambda \exp(-\lambda y), y \geq 0 \quad (6)$$

besitzt wird als Exponentialverteilung zur Zufallsvariablen Y bezeichnet.

Definition 3.6 (Gammaverteilung)

$\Gamma(\alpha, \lambda)$ mit den Parametern α und λ und der Dichtefunktion

$$f(y) = \begin{cases} \lambda^\alpha \frac{y^{\alpha-1}}{\Gamma(\alpha)} \exp(-\lambda y), & y \geq 0 \\ 0, & y < 0 \end{cases} \quad (7)$$

wird als Gammaverteilung zur Zufallsvariablen Y bezeichnet. Der Erwartungswert der Gammaverteilung beträgt $\frac{\alpha}{\lambda}$ und die Varianz $\frac{\alpha}{\lambda^2}$.

Definition 3.7 (Inverse Normalverteilung)

Die stetige Zufallsvariable Y mit Realisierung y besitzt eine inverse Normalverteilung mit den Parametern $\lambda > 0$ und $\mu > 0$, sofern ihre Dichte folgendermaßen gegeben ist:

$$f(y) = \left(\frac{\lambda}{2\pi y^3} \right)^{\frac{1}{2}} \exp \left(- \left(\frac{\lambda(y - \mu)^2}{2\mu^2 y} \right) \right), y > 0 \quad (8)$$

und

$$f(y) = 0, y \leq 0$$

Bei dieser Verteilung ist der Erwartungswert $E(y) = \mu$ und die Varianz $Var(y) = \frac{\mu^2}{\lambda}$.

3.2 Wichtige Definitionen

Folgende Definitionen sind für den weiteren Verlauf dieser Arbeit von Bedeutung:

Definition 3.8 (Erwartungswert)

Der Erwartungswert ist das Integral bezüglich eines Wahrscheinlichkeitsmaßes: Ist X eine Zufallsvariable eines Wahrscheinlichkeitsraumes (Ω, Σ, P) so ist der Erwartungswert der Zufallsvariable X definiert als:

$$E(X) = \int_{\Omega} X dP = \int_{\Omega} X(\omega) P(d\omega)$$

Der Erwartungswert von einer diskreten bzw. einer stetigen Zufallsvariable lässt sich individuell berechnen:

Definition 3.9 (Erwartungswert für den diskreten Fall)

Sei X eine diskrete Zufallsvariable mit den jeweiligen Werten x_1, x_2, \dots und den zugehörigen Wahrscheinlichkeiten p_1, p_2, \dots , $i = 1, \dots$, dann kann der Erwartungswert folgendermaßen berechnet werden:

$$E(X) = \sum_i x_i p_i = \sum_i x_i P(X = x_i)$$

Definition 3.10 (Erwartungswert für den stetigen Fall)

Besitzt X eine Wahrscheinlichkeitsdichtefunktion f , so berechnet sich der Erwartungswert folgendermaßen:

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

Da in der folgenden Arbeit des Öfteren von dem bedingten Erwartungswert die Rede ist, ist es wichtig diesen zu definieren:

Definition 3.11 (Bedingter Erwartungswert)

Sei Y eine Zufallsvariable auf dem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) und $\mathcal{C} \subseteq \mathcal{A}$ eine Teil- σ -Algebra. Eine Zufallsvariable Z auf (Ω, \mathcal{A}, P) heißt bedingter Erwartungswert von Y unter \mathcal{C} , falls gilt:

1. Z ist \mathcal{C} -messbar und $E(|Z|) < \infty$
2. $\int_C Z dP = \int_C Y dP$ für alle $C \in \mathcal{C}$

Definition 3.12 (Bedingte Wahrscheinlichkeit)

Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum. Für zwei Ereignisse $A, B \in \mathcal{A}$ mit $P(B) > 0$ ¹⁶ wird die bedingte Wahrscheinlichkeit von A unter B folgendermaßen definiert:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Definition 3.13 (Diskrete und stetige Randdichtefunktion)

Es werden die beiden folgenden Fälle unterschieden:

1. Wenn $f_{X,Y}$ die gemeinsame diskrete Dichtefunktion von den Zufallsvariablen (X, Y) mit den Wertetupeln (x_k, y_i) , $i, k = 1, 2, \dots$ ist, dann werden die diskreten Dichtefunktionen:

$$f_X(x_k) = \sum_i f_{X,Y}(x_k, y_i)$$

und

$$f_Y(y_i) = \sum_k f_{X,Y}(x_k, y_i)$$

als diskrete Randdichtefunktionen von X , bzw. Y bezeichnet.

2. Wenn (X, Y) ein 2-dimensionaler Zufallsvektor ist, dann werden die Dichtefunktionen

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy$$

und

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx$$

¹⁶ $P(B)$ stellt hier die Wahrscheinlichkeit des Ereignisses B dar

als stetige Randdichtefunktionen von X , bzw. Y bezeichnet.

Definition 3.14 (Bedingte Dichtefunktion)

Auch hier werden zwei Fälle unterschieden:

1. Seien X, Y diskrete Zufallsvariablen mit der gemeinsamen, diskreten Dichtefunktion $f_{X,Y}$. Die bedingte diskrete Dichtefunktion von Y gegeben ($X = x$) ist definiert durch:

$$f_{Y|X}(y|x) := \frac{f_{X,Y}(x, y)}{f_x(x)}$$

für $f_X(x) > 0$, wobei f_X die Randdichte von X and der Stelle x ist. Anstatt von $f_{Y|X}(y|x)$ kann auch $f(y|x)$, bzw. $f_x(y)$ geschrieben werden.

2. Die bedingte, stetige Verteilungsfunktion von Y gegeben ($X = x$) ist definiert durch:

$$f_{Y|X}(y|x) := \frac{f_{X,Y}(x, y)}{f_x(x)}$$

für $f_X(x) > 0$, wobei f_X die Randdichte von X an der Stelle x ist.

Wichtig sind auch die Kovarianzen von Zufallsvariablen:

Definition 3.15 (Kovarianz)

Sind X und Y zwei Zufallsvariablen, bei denen die quadrierten Erwartungswerte $E(X^2)$ und $E(Y^2)$ existieren, dann ist die Kovarianz von X und Y so definiert:

$$\text{cov}(X, Y) := E((X - E(X))(Y - E(Y)))$$

Definition 3.16 (Kovarianzmatrix)

Die Kovarianz ist die Matrix aller paarweisen Kovarianzen eines Zufallsvektors.

Sei $X = (X_1, \dots, X_n)$ ein Zufallsvektor. Dann ist die Kovarianzmatrix $\text{Cov}(X)$ definiert als:

$$\text{Cov}(X) := \text{cov}(X_i, X_j)_{i,j=1,\dots,n} \in \mathfrak{R}^{n,n}$$

Die Kovarianzmatrix hat also folgende Gestalt:

$$\text{Cov}(X) = \begin{pmatrix} \text{cov}(X_1, X_1) & \cdots & \text{cov}(X_1, X_n) \\ \vdots & \ddots & \vdots \\ \text{cov}(X_n, X_1) & \cdots & \text{cov}(X_n, X_n) \end{pmatrix}$$

Definition 3.17 (Erwartungstreue)

Eine Schätzfunktion \hat{a} für $a(\theta)$ heisst erwartungstreu genau dann, wenn:

$$\forall \theta \in \Theta : E_{\theta}(\hat{a}(Y_1, \dots, Y_n)) = a(\theta)$$

Ist $\hat{a}(Y_1, \dots, Y_n) = \hat{a}(Y)$ ein Zufallsvektor, also $\hat{a}(Y) = (\hat{a}_1(Y), \dots, \hat{a}_k(Y))^t$, so kann der Erwartungswert komponentenweise definiert werden:

$$E_{\theta}(\hat{a}(Y)) = (E_{\theta}(\hat{a}_1(Y)), \dots, E_{\theta}(\hat{a}_k(Y)))^t$$

Definition 3.18 (Konsistenz)

Eine Folge von Schätzfunktionen $(\hat{a}_n)_{n \in \mathbb{N}}$ heisst stark, bzw. schwach konsistent für den Aspekt $a(\theta)$ (Ein Aspekt ist eine Abbildung $a : \Theta \rightarrow A$.

$a(\theta)$ mit $\theta \in \Theta$ ist der Aspekt von θ) genau dann,

1. wenn:

$$P_{\theta}(\lim_{n \rightarrow \infty} \hat{a}_n(Y_1, \dots, Y_n) = a(\theta)) = 1 \forall \theta \in \Theta$$

2. oder wenn:

$$\lim_{n \rightarrow \infty} P_{\theta}(\|\hat{a}_n(Y_1, \dots, Y_n) - a(\theta)\| > \epsilon) = 0 \forall \epsilon > 0, \theta \in \Theta$$

Der erste Fall wird als fast sichere Konvergenz und der zweite Fall als stochastische Konvergenz bezeichnet.

Soweit zu den statistischen Grundlagen, die für diese Arbeit von Bedeutung sind. Einige individuelle Definitionen werden in den kommenden Kapiteln noch folgen.

4 Lineares Modell und Modelle für longitudinale Daten

Im Folgenden wird eine kurze Einführung in die verschiedenen Formen der linearen Modelle gegeben, um Grundlagen für das im nächsten Kapitel dargestellte „*Generalized linear Mixed Model*“ (GLMM) zu schaffen und ebenfalls auf andere Modelle einzugehen, die bei der Analyse von longitudinalen Daten verwendet werden. Hier sind in erster Linie die *subjektspezifischen Modelle*, unter die auch das GLMM fällt, und das *Marginalmodell* zu nennen.

Im Allgemeinen ist das GLMM eine verallgemeinerte Regressionsanalyse, welche mit der ANOVA, also der Varianzanalyse vergleichbar ist.¹⁷

Wichtig zu erwähnen ist, dass mit dem GLMM eine Unterscheidung zwischen multiplen, festen Effekten und Zufallseffekten ermöglicht wird. Wie sich im Folgenden herausstellen wird, ist das verallgemeinerte lineare gemischte Modell besonders gut für wiederholte Messungen geeignet, sowie es bei longitudinalen Daten der Fall ist, und ist zudem unempfindlich gegenüber „missing Values“.¹⁸ Bevor das verallgemeinerte lineare gemischte Modell weiter beschrieben wird, werden, wie schon erwähnt, zunächst weitere Modelle aufgeführt, welche bei der Analyse von longitudinalen Daten¹⁹ von Bedeutung sind. Bei der Analyse steht natürlich im Interesse das am besten geeignete Modell für die Datenauswertung zu finden und zu verwenden. Zu diesen Modellen gehören die *subjektspezifischen Modelle* und das *Marginalmodell*. Wie nachher auch noch herausgestellt wird, gehört das GLMM zu der Gruppe der subjektspezifischen Modelle. Nun aber zunächst zum allgemeinen linearen Modell.

4.1 Das allgemeine lineare Modell für den multivariaten Fall

Definition 4.1 (Lineares multivariates Modell)

Das einfache lineare Modell oder auch das allgemeine lineare Modell für den

¹⁷siehe: [(Abschlussbericht 2007)] S.12

¹⁸siehe Kapitel (7.5)

¹⁹Diese werden in Kapitel (5) noch ausführlich dargestellt

multivariaten Fall sieht wie folgt aus²⁰:

$$Y = XB + E \tag{9}$$

dabei sind:

$$Y = \begin{pmatrix} Y_1^t \\ \vdots \\ Y_n^t \end{pmatrix} = \begin{pmatrix} Y_{11} & \cdots & Y_{1p} \\ \vdots & & \vdots \\ Y_{n1} & \cdots & Y_{np} \end{pmatrix}$$

eine $n \times p$ -Zufallsmatrix,

$$E = \begin{pmatrix} E_1^t \\ \vdots \\ E_n^t \end{pmatrix} = \begin{pmatrix} E_{1*}, & \cdots, & E_{p*} \end{pmatrix} = \begin{pmatrix} E_{11} & \cdots & E_{1p} \\ \vdots & & \vdots \\ E_{n1} & \cdots & E_{np} \end{pmatrix}$$

eine $n \times p$ Zufallsmatrix von Fehlern ($i = 1, \dots, n$ und $j = 1, \dots, p$), $B = (\beta_1, \dots, \beta_p) \in \mathfrak{R}^{r \times p}$ ein Matrix von unbekanntem Parametern, bzw. von Regressionsparametern und X eine $n \times r$ Matrix von bekannten Variablen, welche auch als Designmatrix bezeichnet wird. Die Matrix B ist eine Matrix von sogenannten festen Effekten.

Es wird angenommen, dass E_i für $i = 1, \dots, n$ stochastisch unabhängig und identisch normalverteilt ist. Das bedeutet, dass $E(E_i) = 0_p$ und $Cov(E_i) = \Sigma \in \mathfrak{R}^{p \times p}$ gilt.

Damit sind auch die Y_1, \dots, Y_n stochastisch unabhängig mit $E(Y_i) = (x(t_n)^t B)^t$ und $Cov(Y_i) = \Sigma$

Aus diesen multivariaten linearen Modell erhält man ganz einfach auch das univariate lineare Modell, was hier allerdings nicht weiter erklärt werden soll, da die Daten, die in der folgenden Arbeit ausgewertet werden, multivariat sind.

²⁰ $(\cdot)^t$ steht ab hier immer für einen transponierten Vektor oder eine transponierte Matrix

Das allgemeine lineare Modell in Matrixschreibweise ist gegeben durch:

$$\begin{pmatrix} Y_{11} & \cdots & Y_{1p} \\ \vdots & & \vdots \\ Y_{n1} & \cdots & Y_{np} \end{pmatrix} = \begin{pmatrix} X_{11} & \cdots & X_{1r} \\ \vdots & & \vdots \\ X_{n1} & \cdots & X_{nr} \end{pmatrix} \begin{pmatrix} \beta_{11} & \cdots & \beta_{1p} \\ \vdots & & \vdots \\ \beta_{r1} & \cdots & \beta_{rp} \end{pmatrix} + \begin{pmatrix} E_{11} & \cdots & E_{1p} \\ \vdots & & \vdots \\ E_{n1} & \cdots & E_{np} \end{pmatrix}$$

Nachdem das allgemeine lineare Modell für den multivariaten Fall dargestellt wurde, stellt sich noch die Frage, wie die Schätzungen in diesem Modell aussehen. Wie bei jedem Modell spielen auch hier die Schätzungen eine bedeutende Rolle.

Da Maximum-Likelihood-Schätzungen für den weiteren Verlauf der Arbeit von Bedeutung sind, soll diese nun zunächst für das lineare Modell aufgeschrieben werden.

Zunächst wird allerdings kurz wiederholt, wie die Likelihood-Funktion und ihre zugehörige Maximum-Likelihood-Schätzung definiert sind:

Definition 4.2 (Likelihood-Funktion)

Es sei $P^Y \in (P_\theta^Y; \theta \in \Theta)$

1. Besitzt für jedes $\theta \in \Theta$ P_θ^Y die stetige Dichte f_θ , dann heißt die Funktion $L(\cdot, y) : \Theta \ni \theta \rightarrow L(\theta, y) := f_\theta(y) \in \mathfrak{R}$ für jedes $y \in Y$ die Likelihood-Funktion des statistischen Experiments bei der Beobachtung y .
2. Besitzt für jedes $\theta \in \Theta$ P_θ^Y die diskrete Dichte p_θ , dann heißt die Funktion $L(\cdot, y) : \Theta \ni \theta \rightarrow L(\theta, y) := p_\theta(y) \in [0, 1]$ für jedes $y \in Y$ die Likelihood-Funktion des statistischen Experiments bei der Beobachtung y .

Die Maximum-Likelihood-Schätzung ist nun folgendermaßen definiert:

Definition 4.3

Eine Schätzfunktion $\hat{\theta}$ für θ wird Maximum-Likelihood-Schätzung genannt, sofern:

$$L(\hat{\theta}(y), y) = \sup_{\theta \in \Theta} L(\theta, y)$$

für alle $y \in Y_n$ gilt. Das bedeutet, dass $\hat{\theta}(y) \in \arg \sup_{\theta \in \Theta} L(\theta, y)$

Satz 4.1 (Maximum-Likelihood-Schätzung für das multivariate lineare Modell)

Gilt $Y_i = x(t_i)^t B + E_i^t$ mit $E_i \sim N_p(0_p, \Sigma)$, wobei $B \in \mathbb{R}^{r \times p}$ und $\Sigma \in \mathbb{R}^{p \times p}$ unbekannt sind, sowie E_1, \dots, E_n stochastisch unabhängig. Dann ist $(X\hat{B}, \hat{\Sigma})$ die einzige Maximum-Likelihood-Schätzung für (XB, Σ) . Dabei sind²¹

$$\begin{aligned}\hat{B} &= (X^t X)^- X^t y \\ \hat{\Sigma} &= \frac{1}{n} \sum_{i=1}^n (y_i - \hat{B}^t x(t_i))(y_i - \hat{B}^t x(t_i))^t\end{aligned}$$

Beweis. Da der Beweis dieses Satzes ein Standardbeweis aus der Statistik ist, wird dieser hier nicht gemacht werden. \square

4.2 Subjektspezifische Modelle

In einer Studie mit longitudinalen Daten steht im Interesse zu untersuchen, wie die Subjekte sich über die Zeit verändern und welchen Eigenschaften diese Veränderungen zugrunde liegen.

Bei einem sogenannten subjektspezifischen Modell, welches auch oft als Zufallseffektmodell („random effect modell“)²² bezeichnet wird, werden die Antworten als unabhängig angenommen, was dazu führt, dass sich eine Anhäufung von den sogenannten subjektspezifischen Parametern ergibt. Neben den erklärenden Variablen, gibt es bei den subjektspezifischen Modellen also noch bestimmte spezifisch gebundene Parameter. Diese Regressionsparameter geben eine Aussage über die einzelnen Subjekte, nicht aber über den Populationsdurchschnitt. Will man Aussagen über den Populationsdurchschnitt treffen, muss im Anschluss an die Analyse anhand des subjektspezifischen Modells eine „Marginalisierung“²³ des Modells vorgenommen werden.

Bei einer longitudinalen Studie steht aber häufig die Beschreibung der einzelnen Subjekte, sowie die Vorhersage von subjektspezifischen Entwicklungen, im Vordergrund und in solch einem Fall stellen die subjektspezifischen Modelle eine gute Möglichkeit dar, solche Daten auszuwerten.

Nach [(Molenberghs und Verbeke 2005)] gibt es drei verschiedene Arten von subjektspezifischen Modellen, die durch unterschiedliche Handhabungen von

²¹ $(X^t X)^-$ ist die sogenannte G-Inverse von $(X^t X)$

²²siehe: z.B. [(Diggle 2002)] S.128

²³siehe: Unterkapitel: Das Marginalmodell

spezifisch-gebundenen Regressionsparametern zustande kommen. Wie schon erwähnt werden die Modelle auch oft als Zufallseffektmodelle bezeichnet, z.B. in [(Diggle 2002)], jedoch ist das nur eine der Teilfamilien der subjektspezifischen Modelle.

Bemerkung 4.1 (Teilfamilien der subjektspezifischen Modelle)

Es gibt folgende drei Handhabungen bei subjektspezifischen Modellen:

- *Die spezifischen Parameter werden als feste unbekannte Effekte aufgefasst*
- *Oder sie gelten als störende Parameter*
- *In der dritten Handhabung werden die Parameter als Zufallseffekte gesehen. Da diese Methode die häufigste der drei Arten ist, wird in der Literatur auch meist von Zufallseffektmodellen gesprochen.*

Bzgl. der dritten Handhabung der subjektspezifischen Modelle bedarf es in dieser Arbeit einer näheren Betrachtung. Zu den Modellen aus dieser dritten Gruppe gehören zum Einen das Beta-Binomial Modell, welches für binäre Daten verwendet werden kann, zum Anderen das am Anfang dieses Kapitels schon erwähnte GLMM²⁴.

4.3 Das lineare gemischte Modell

Ein spezieller Fall der eben erläuterten subjektspezifischen Modelle wird nun kurz thematisiert: Das lineare gemischte Modell.

Das lineare gemischte Modell ist wie das allgemeine lineare Modell aufgebaut, nur dass es noch einen zusätzlichen Zufallseffekt besitzt. Dieses Modell wird meist bei geclusterten, also gebündelten Daten verwendet, folglich also bei abhängigen Daten. Ein Beispiel dafür ist z. B., wenn Daten über eine gewisse Zeit von den gleichen Individuen beobachtet, bzw. gemessen werden.²⁵

Definition 4.4 (Lineares gemischtes Modell)

Das lineare gemischte Modell ist gegeben durch:

$$Y = XB + WA + E \tag{10}$$

²⁴siehe Kapitel (5)

²⁵siehe: [(Fox 2002)] S.1

wobei Y, X, B und E wie in (9) definiert sind und W eine weitere Designmatrix, sowie A eine Matrix von Zufallseffekten ist.

Der Unterschied zum allgemeinen linearen Modell ist hier also ganz einfach zu sehen. Beim linearen Modell liegen nur solche Variablen vor, von denen ein Zusammenhang zu den Antwortvariablen vermutet wird, welche als „feste Faktoren“ bezeichnet werden, beim linearen gemischten Modell kommt noch eine weitere Gruppe von Faktoren hinzu, die sogenannten „zufälligen Faktoren“. Diese Faktoren spielen bei einer Untersuchung meist keine primäre Rolle. Dies kann z.B. der Name einer Versuchsperson, bzw Versuchstieres sein oder aber auch die Nummer eines experimentellen Blocks.

Wie im Folgenden noch zu sehen ist, ist das lineare gemischte Modell ein spezieller Fall des GLMMs.

4.4 Das verallgemeinerte lineare Modell (GLM)

Das verallgemeinerte oder auch generalisierte lineare Modell (GLM) stellt (wie der Name schon sagt) eine Verallgemeinerung der klassischen linearen Modelle dar.

Der Unterschied besteht darin, dass bei den klassischen linearen Modellen immer vorausgesetzt wird, dass die Zielvariable normalverteilt ist. Beim GLM ist das nicht zwingend der Fall, die Verteilung muss hier jediglich aus der Klasse der exponentiellen Familien stammen, welche die Normal-, Binomial-, Poisson-, Gamma- und inverse Normalverteilung enthält. Welche Verteilung die Zielvariable besitzt, muss individuell nachgeprüft werden.

Definition 4.5 (Verallgemeinertes lineares Modell (GLM))

Das verallgemeinerte lineare Modell besteht nun aus folgenden drei Komponenten:

1. dem beobachteten Zufallsvektor $Y = (Y_1, \dots, Y_n)^t \in \mathfrak{R}^n$
2. dem linearen Prädiktor $x_1, \dots, x_n \in \mathfrak{R}^p$, welcher auch als systematische Komponente bezeichnet wird.
3. der Linkfunktion g

Der lineare Prädiktor für das verallgemeinerte lineare Modell ist folgendermaßen definiert:

Definition 4.6 (Linearer Prädiktor für das GLM)

Wenn die Variablenvektoren x_1, \dots, x_n für $i = 1, \dots, n$ gegeben sind, welche die Verteilung von $Y = (Y_1, \dots, Y_n)^t$ nur durch eine lineare Funktion beeinflussen, wird diese Funktion als linearer Prädiktor bezeichnet und ist definiert durch:

$$\eta = x_i^t \beta$$

wobei β einen festen Effekt darstellt.

Definition 4.7 (1-parametrische Exponentialfamilie)

Wenn die Zufallsvariable Y_i für $i = 1, \dots, n$ mit den Realisierungen y_i die Dichte der Form:

$$f_{\theta, \phi}(y_i) = \exp \left([y_i \theta - b(\theta)] \frac{w}{\phi} + c(y_i, \phi) \right) \quad (11)$$

besitzt, mit $\theta, \phi \in \mathfrak{R}$ unbekannt, $w \in \mathfrak{R}$, $b : \mathfrak{R} \rightarrow \mathfrak{R}$ und $c : \mathfrak{R}^2 \rightarrow \mathfrak{R}$ bekannt, dann hat Y_i die Dichte einer 1-parametrischen natürlichen Exponentialfamilie mit dem Dispersionsparameter ϕ

Beispiele der 1-parametrischen Exponentialfamilie sind unter Anderem die Normal-²⁶, Exponential-, Gamma-, und die Poissonverteilung.

Definition 4.8 (Linkfunktion)

Eine Linkfunktion g mit $g : \mathfrak{R} \rightarrow \mathfrak{R}$ ist dann gegeben, wenn die Beobachtungen $Y_i, i = 1, \dots, n$ die Dichtefunktionen einer 1-parametrischen natürlichen Exponentialfamilie mit Dispersionsparameter ϕ haben (siehe: Definition (11)) und es Regressoren $x_1, \dots, x_n \in \mathfrak{R}^p$ und ein $\beta \in \mathfrak{R}^p$ (der zu schätzender Parameter) gibt, so dass $\eta := x_i^t \beta = g(\mu_i)$, wobei hier $\mu_i = E(Y_i)$ ist. Die Linkfunktion stellt somit den Zusammenhang von den Zufallsgrößen Y_i und dem linearen Prädiktor her.

Die Linkfunktion wird für den weiteren Verlauf dieser Arbeit von Bedeutung sein.²⁷

Beispiel 4.1 (Normalverteilung)

In diesem Beispiel wird gezeigt, dass die Normalverteilung zur Gruppe der 1-parametrischen Exponentialfamilie gehört.

Man nimmt an, dass die Zufallsvariable Y_i normalverteilt ist mit Erwartungs-

²⁶siehe: Beispiel (4.1)

²⁷siehe Kapitel (5)

wert μ und Varianz σ^2 ($N(\mu, \sigma^2)$):

$$\begin{aligned} f_{(\mu, \sigma^2)}(y_i) &= \frac{1}{\sqrt{2\pi\sigma}} \exp\left(\frac{-(y_i - \mu)^2}{2\sigma^2}\right) \\ &= \exp\left(-\ln(\sqrt{2\pi\sigma}) - \frac{1}{\sigma^2}((y_i)^2 - 2y_i\mu + \mu^2)\right) \\ &= \exp\left([y_i\mu - \frac{1}{2}\mu^2] \frac{1}{\sigma^2} - \ln(\sqrt{2\pi\sigma}) - \frac{(y_i)^2}{2\sigma^2}\right) \\ &= \exp\left([y_i\theta - b(\theta)] \frac{w}{\phi} + c(y_i, \phi)\right) \end{aligned}$$

wobei hier $\theta = \mu$, $b(\theta) = \frac{1}{2}\mu^2$, $w = 1$, $\phi = \sigma^2$, $c(y_i, \phi) = -\ln(\sqrt{2\pi\sigma}) - \frac{(y_i)^2}{2\sigma^2}$

Folgender Satz ist in diesem Zusammenhang noch von Bedeutung und wird in der Arbeit noch vermehrt zur Anwendung kommen:

Satz 4.2

Besitzt y_n eine Verteilung der 1-parametrischen natürlichen Exponentialfamilie (siehe Definition) mit Dispersionsparameter ϕ und Parameter θ , dann gilt für den Fall, dass Differentiation nach θ und Integration vertauscht werden dürfen:

$$E(y_i) = b'(\theta) \tag{12}$$

$$Var(y_i) = b''(\theta) \frac{\phi}{w} \tag{13}$$

Der Beweis dieses Satzes wird hier nicht gemacht.

4.5 Das Marginalmodell

In diesem Zusammenhang muss nun auch noch das sogenannte Marginalmodell näher erläutert werden. Dieses Modell, welches auch als Populationsdurchschnittsmodell („population-averaged model“²⁸) bezeichnet wird, ist eine Erweiterung des verallgemeinerten linearen Modells und wird häufig bei der Analyse von longitudinalen Studien angewandt.

In den sogenannten marginalen Modellen werden die Beobachtungen marginal über allen anderen Beobachtungen modelliert.²⁹

Bei einem Marginalmodell geht es nach [(Diggle 2002)] darum, den Einfluss

²⁸siehe: [(Molenberghs und Verbeke 2005)]

²⁹siehe: [(Molenberghs und Verbeke 2005)] S.47

der erklärenden Variablen auf die Beobachtung getrennt von der Korrelation der Beobachtungen bei einem Subjekt zu modellieren.

Wenn also $Y_i = (Y_{i1}, \dots, Y_{in_i})$ mit $i = 1, \dots, n$, $j = 1, \dots, n_i$ und die Designmatrix $X_i = (x_{i1}, \dots, x_{in_i})$ gegeben sind, dann besteht das Prinzip eines solchen marginalen Modells darin, für jedes $i \in \{1, \dots, n\}$ die Regression von Y_i in Abhängigkeit von X_i unabhängig der Struktur der Y_i zu modellieren.

Es wird die Regression des marginalen Erwartungswertes $E(Y_{ij})$ als eine Funktion der erklärenden Variablen aufgefasst. Mit dem marginalen Erwartungswert ist die durchschnittliche Antwort einer Subpopulation, also einer Untergruppe der Grundgesamtheit, gemeint, die die gleichen erklärenden Variablen besitzt. Es ist also vergleichbar mit der Auswertung von unkorrelierten Daten. Somit kann das Marginalmodell als eine Erweiterung des verallgemeinerten linearen Modells (wie oben schon erwähnt) für korrelierte Daten angesehen werden.

Definition 4.9 (Das Marginalmodell)

Das allgemeine Marginalmodell lässt sich aus folgenden drei Komponenten zusammensetzen³⁰:

1. Der marginale Erwartungswert der Beobachtung Y_{ij} hängt von dem Vektor der erklärenden Variablenvektor $x_{ij} \in \mathbb{R}^p$ ab, gemäß der beiden Gleichungen:

$$h(\mu_{ij}) = x_{ij}^t \beta \forall i \in \{1, \dots, m\}, j \in \{1, \dots, n_i\} \quad (14)$$

$$E(Y_{ij}) = \mu_{ij} \forall i \in \{1, \dots, m\}, j \in \{1, \dots, n_i\} \quad (15)$$

wobei $h(\cdot)$ eine bekannte Linkfunktion ist und $\beta \in \mathbb{R}^p$ der Regressionsparameter.

2. Die marginale Varianz hängt von dem marginalen Erwartungswert ab, entsprechend der Gleichung:

$$\text{Var}(Y_{ij}) = \nu(\mu_{ij})\phi \quad (16)$$

Hier ist $\nu(\cdot)$ eine bekannte Varianzfunktion und $\phi \in \mathbb{R}$ welcher eventuell noch geschätzt werden muss.

3. Die Korrelation zwischen Y_{ij} und Y_{ik} ist eine Funktion der marginalen

³⁰siehe: [(Diggle 2002)] S.126

Erwartungswerte μ_{ij} und μ_{ik} , sowie eines zusätzlichen Parameters α , also:

$$\text{Corr}(Y_{ij}, Y_{ik}) = \rho(\mu_{ij}, \mu_{ik}; \alpha) \quad (17)$$

ρ ist hier eine bekannte Funktion.

Wichtig an dieser Stelle zu erwähnen ist, dass für die Auswertung, bzw. für die Beschreibung von Daten oft eine Poissonverteilung verwendet wird, wobei natürlich vorausgesetzt ist, dass die Varianz und der Erwartungswert übereinstimmen.

Ein wesentlicher Unterschied zwischen den marginalen Modellen und den Subjektspezifischen liegt darin, dass bei einem Marginalmodell nicht, wie bei den subjektspezifischen Modellen, neben den erklärenden Variablen noch zusätzliche subjektspezifische Parameter vorhanden sind.

Das Marginalmodell wird an dieser Stelle nicht weiter erläutert, da sich diese Arbeit in erster Linie mit dem verallgemeinerten linearen gemischten Modell auseinandersetzt, welches im folgenden Kapitel (5) ausführlich bearbeitet wird.

Bevor das getan wird, wird jedoch noch gezeigt, dass zwischen dem Marginalmodell und dem subjektspezifischen Modell im Falle des normalverteilten linearen gemischten Modells ein Zusammenhang vorliegt.

4.6 Der normalverteilte Fall

In den vorhergehenden Kapiteln wurden unter Anderem das lineare gemischte Modell, als ein subjektspezifisches Modell, sowie das Marginalmodell, vorgestellt, welche beide häufig bei der Analyse longitudinaler Daten zur Anwendung kommen. Diese beiden Modellarten sind zwei der drei wichtigen Untergruppen³¹ für „repeated measurements“.³²

Wie eben schon erwähnt, ist es nun von Interesse zu zeigen, dass im normalverteilten Fall, das lineare gemischte Modell und das Marginalmodell zusammenhängen. Dieses wird in folgendem Satz deutlich:

Satz 4.3

Das gaußsche gemischte Modell, also das lineare gemischte Modell für den normalverteilten Fall, enthält ein einfaches Marginalmodell.

³¹auf die dritte Untergruppen, die Transitionsmodelle wird hier nicht weiter eingegangen.

³²siehe: [(Molenberghs und Verbeke 2005)] S.46

Beweis. Ein marginales Modell wird durch seinen marginalen Erwartungswert

$$E(Y_{ij}|X_i) = X_i\beta \quad (18)$$

beschrieben. In einem subjektspezifischen Modell, wie dem linearen gemischten Modell, sieht der Erwartungswert, der zusätzlich von einem Vektor von Zufallseffekten α_i abhängt, folgendermaßen aus:

$$E(Y_{ij}|\alpha_i) = X_i\beta + Z_i\alpha_i \quad (19)$$

Dass das lineare gemischte Modell nun ein marginales Modell enthält, geht auf die Eigenschaften der multivariaten Normalverteilung zurück. Gleichung (18) folgt dann von Gleichung (19) entweder durch das marginalisieren über den Zufallseffekten oder unter der Bedingung, dass $\alpha_i = 0$ ist. Damit haben dann die festen Effekte β eine marginale und eine subjektspezifische Interpretation zur selben Zeit. Es ist daher zu zeigen, dass die Erwartungswerte der beiden Modelle gleich sind. Im Folgenden wird die Variante des Marginalisieren über den Zufallseffekten gezeigt. Um die Regressionsvektor β im Folgenden besser zu unterscheiden, wird nun der Parametervektor für das marginale Modell mit β^M und für das subjektspezifische Modell mit β^{RE} bezeichnet.³³ Es gibt nun einen Zusammenhang zwischen den beiden Parametervektoren für den Fall, dass das subjektspezifische Modell bestimmt ist durch das lineare gemischte Modell mit bedingtem Erwartungswert:

$$E(Y_{ij}|\alpha_i) = X_i\beta^{RE} + Z_i\alpha_i$$

und für das i -te Subjekt durch:

$$Y_i|\alpha_i \sim N(X_i\beta^{RE} + Z_i\alpha_i, \Sigma_i)$$

wobei X_i und Z_i bekannte invertierbare Designmatrizen darstellen und $\alpha_i \sim N(0, D)$ ist. Weiter sei $Y_i = X_i\beta^{RE} + Z_i\alpha_i + \epsilon_i$, wobei $\epsilon_i \sim N(0, \Sigma_i)$ und ϵ_i sowie α_i unabhängig. Den marginalisierten Erwartungswert erhält man nun durch nochmaliges Integrieren über dem Erwartungswert ($E(Y_{ij}|\alpha_i)$).

Somit ist

$$E(Y_i) = E(Y_{ij}|\alpha_i) = X_i\beta^{RE}$$

³³siehe: [(Molenberghs und Verbeke 2005)] S.299

der marginalisierte Erwartungswert. Der marginale Erwartungswert sieht folgendermaßen aus³⁴:

$$E(Y_i) = X_i\beta^M$$

Damit ist: $X_i\beta^M = X_i\beta^{RE}$ und es folgt aufgrund der Invertierbarkeit der Designmatrix X_i , dass

$$\beta^M = \beta^{RE}$$

gilt und damit sind die beiden Erwartungswerte gleich.

Wenn also $f(y_i|\alpha_i)$ die Dichtefunktion von $Y_i|\alpha_i$ und $f(\alpha_i)$ die Dichte der Zufallseffekte des normalverteilten linearen gemischten Modells ist, dann ist die marginale Dichtefunktion von Y_i gegeben durch³⁵:

$$f(y_i) = \int f(y_i|\alpha_i)f(\alpha_i)d\alpha_i \quad (20)$$

Die Funktion (20) ist aufgrund der beiden Dichtefunktionen $f(y_i|\alpha_i)$ und $f(\alpha_i)$, welche ja aufgrund der eben beschriebenen Eigenschaften eine Dichte der Normalverteilung sind, die Dichtefunktion einer n_i -dimensionalen Normalverteilung. Das marginale Modell ist nun gegeben durch:

$$Y_i \sim N(X_i\beta, Z_iDZ_i^t + \Sigma_i)$$

Damit ist klar, dass das normalverteilte lineare gemischte Modell ein einfaches marginales Modell enthält. □

³⁴siehe: (18)

³⁵siehe: [(Verbeke und Molenberghs 2000)] S.24

5 Das verallgemeinerte lineare gemischte Modell (GLMM)

Im Folgenden wird das verallgemeinerte lineare gemischte Modell vorgestellt, welches für die Auswertung der Daten aus der im zweiten Kapitel vorgestellten Forschungsstudie verwendet wird. Das GLMM fällt unter die Kategorie der Zufallseffektmodelle, die eine Teilfamilie der subjektspezifischen Modelle darstellen und ist das am häufigsten verwendete Zufallseffektmodell für diskrete Daten.³⁶

Da nicht alle Daten mit normalen oder auch verallgemeinerten linearen Modellen ausgewertet werden können, ist das GLMM ein wichtiger Fortschritt, um eine große Varietät von Datenstrukturen, die in vielen Untersuchungen, unter Anderem natürlich auch in biologischen Studien, vorliegt, zu untersuchen und auszuwerten.³⁷ Und gerade die longitudinalen Daten, die im Kapitel (7) noch näher behandelt werden, ebenso wie „repeated measures“ Daten, stellen bei der statistischen Analyse oft Probleme dar und können meist am besten mit einem GLMM ausgewertet werden.³⁸

Das verallgemeinerte lineare gemischte Modell ist somit ein geeignetes Modell für die Auswertung von Daten, die neben den festen Effekten noch Zufallseffekte enthalten. Des Weiteren kann das GLMM, welches ein gemischtes Modell ist, gut verwendet werden, um nichtlineare Beziehungen über die Zeit darzustellen. Diese liegen häufig bei longitudinalen Datensätzen mit fehlenden Werten vor.³⁹

Das GLMM ist zum Einen eine Erweiterung des verallgemeinerten linearen Modells für univariate Daten im Zusammenhang mit geclusterten Größen⁴⁰, zum Anderen gibt es für dieses Modell eine große Auswahl von geeigneten Softwareprogrammen; hier ist zum Beispiel das Statistikprogramm SAS, in dem für die Datenanalyse anhand des GLMMs die beiden Prozeduren „GLIMMIX“ und

³⁶siehe: [(Molenberghs und Verbeke 2005)] S.265

³⁷siehe: [(Avital,Laird,Slasor 1997)] S.1

³⁸Mehr zu der Analyse von longitudinalen Daten in Kapitel (7)

³⁹siehe:(Krueger und Tian 2004) S.151

⁴⁰Ein Cluster ist eine Gruppe von Daten bzw. Objekten, welche Ähnlichkeiten (oder auch Unterschiede) in ihren Eigenschaften oder Eigenschaftsprägungen aufweisen.

„NLMIXED“ verwendet werden können, oder die freie Statistiksoftware R zu nennen.

Das GLMM ähnelt in vielerlei Hinsicht der multiplen Regression. Es erfordert jedoch noch zusätzliche Arbeit, um das Modell weiter zu bestimmen, sowie die Anpassungsgüte (goodness of fit) zu berechnen. Die Anpassungsgüte beschreibt, wie gut ein statistisches Modell die Menge der Beobachtungen trifft. Das verallgemeinerte lineare gemischte Modell stellt gute Möglichkeiten dar, um longitudinale Daten zu analysieren und wird vor allem in der Forschung von biologischen Fragestellungen verwendet, weil es ein geeignetes System für die Analyse von gerade den Daten darstellt, welche aus biologischen Studien gewonnen werden.⁴¹

Bemerkung 5.1

Die folgenden fünf Punkte legen dar, dass es oft von Vorteil ist, bei der Analyse von longitudinalen Daten das GLMM zu verwenden.

1. *Es können die fehlenden Datenpunkte (Missing Values) besser berechnet werden.*
2. *Es besteht bei dem Modell noch die Möglichkeit nichtlineare Daten, mit ihrem individuellen Charakter zu analysieren.*
3. *Anhand des Modells können die verschiedenen Gruppen gut charakterisiert, sowie das Verhaltensmuster auf eine formale Weise dargestellt werden.*
4. *Die Unterschiede von Gruppen und Individuum können herausgestellt werden.*
5. *Bei der Analyse anhand des GLMMS können zusätzliche Kovarianzen berücksichtigt werden.*

Um einen ersten kleinen Einblick in das verallgemeinerte lineare gemischte Modell zu erhalten, wird nun eine Darstellung erfolgen, welche das GLMM symbolisch wiedergibt.

$$\left\{ \begin{array}{l} \sigma \xrightarrow{G} \alpha \rightarrow \\ \beta \rightarrow \end{array} \right. \quad Xb + Z\alpha \rightarrow \eta \xrightarrow{h(\eta)} \mu \xrightarrow{V} y$$

σ sind die Varianzkomponenten, α die zufälligen und β die festen Effekte,

⁴¹[(Krueger und Tian 2004)] S.2

mit $\alpha \sim N(0, \Sigma)$, η stellt den linearen Prädiktor⁴² dar, der gleich $x^t\beta + z^t\alpha$ ⁴³ ist. $h(\eta)$ ist die inverse Linkfunktion, auf die im Folgenden noch genauer eingegangen wird, μ ist der Erwartungswert von der Beobachtung y und V stellt die Kovarianzmatrix dar, welche von μ durch eine Varianzfunktion $\nu(\cdot)$ abhängt.

Für den weiteren Verlauf wird ab jetzt angenommen, dass longitudinale Daten vorliegen.

Seien nun $i = 1, \dots, n$ Subjekte und $j = 1, \dots, n_i$ wiederholte Beobachtungen gegeben, welche innerhalb jeder Beobachtungseinheit gemessen werden.

Ein Zufallseffektmodell, welches das einfache gemischte Modell darstellt, mit einem ausgeweiteten linearen Prädiktor und einem Zufallseffekt für das Subjekt i sieht folgendermaßen aus:

$$\eta_{ij} = x_{ij}^t\beta + \alpha_i, \quad (21)$$

wobei α_i die Zufallseffekte darstellt. Diese Zufallseffekte sind $N(0, \sigma_i^2)$ verteilt und repräsentieren den Einfluss vom Subjekt i auf die wiederholte Beobachtung.

Der erwartete Wert, welcher in Beziehung zum linearen Prädiktor steht, ist gegeben durch:

$$\mu_{ij} = E(y_{ij}|\alpha_i, x_{ij})$$

Das gerade beschriebene Modell kann noch ausgeweitet werden, so dass es einen Vektor von Zufallseffekten enthält.

Bei den Analysen von longitudinalen Daten liegt neben einem Zufallseffekt zudem meist noch ein linearer Zeittrend vor. Bei diesem ist zusätzlich ein z_{ij} vorhanden, ein Vektor der Länge $r + 1$ von Variablen mit weiteren Zufallseffekten, $\alpha_i \sim N_i(0, \sigma^2 B)$ sei der Vektor der Zufallseffekte. Der lineare Prädiktor sieht für das verallgemeinerte lineare gemischte Modell folgendermaßen aus:

$$\eta_{ij} = x_{ij}^t\beta + z_{ij}^t\alpha_i \quad (22)$$

Man vergleiche hier mit Gleichung (21).

⁴²siehe vorheriges Kapitel „Verallgemeinertes lineares Modell“

⁴³siehe unten

Definition 5.1 (Verallgemeinertes lineares gemischtes Modell)

Das GLMM durch folgende Gleichung dargestellt:⁴⁴

$$Y_i = X_i\beta + Z_i\alpha_i + \epsilon_i \quad (23)$$

Hier ist Y_i der $n_i \times 1$ Spaltenvektor von der Antwortvariable für das Subjekt i , X_i die $n_i \times g$ Designmatrix, β ein $b \times 1$ Vektor der Regressionskoeffizienten der festen Effekte, Z_i die Designmatrix der Zufallseffekte, α_i sind wieder die Zufallseffekte, welche $N(0, \sigma^2 B)$ verteilt (B ist eine beliebige Kovarianzmatrix) und die Fehler ϵ_i besitzen die Verteilung $N(0, \sigma^2 W_i)$, wobei W_i wieder eine Kovarianzmatrix darstellt. Diese wird häufig als Identitätsmatrix angenommen.⁴⁵ Bedingt auf den α_i sind die Y_i paarweise unabhängig und die α_i sind ebenfalls paarweise unabhängig und identisch verteilt mit der oben genannten Verteilung. Die ϵ_i sind unabhängig von den verschiedenen Beobachtungseinheiten, sowie unabhängig von den Zufallseffekten α_i und identisch verteilt.

Definition 5.2 (Dichtefunktion beim GLMM)

Bei einem GLMM wird angenommen, dass die Dichtefunktion des n_i -dimensionalen Vektors Y_i oder besser gesagt die Dichtefunktion der Beobachtungen Y_{ij} (welche unter der Annahme, dass die q -dimensionalen Zufallseffekte α_i der Verteilung $N(0, \sigma^2 B)$ folgen, unabhängig sind) folgende Form hat:

$$(f_i)_{\alpha_i, \beta, \phi}(y_{ij}) = \exp(\phi^{-1}[y_{ij}\theta_{ij} - b(\theta_{ij})] + c(y_{ij}, \phi)) \quad (24)$$

Um das GLMM $Y = X\beta + Z\alpha + \epsilon$ in seiner Matrixschreibweise deutlich zu machen, seien nun $Y \in \mathfrak{R}^{n \times n_i}$, $i = 1, \dots, n$, $j = 1, \dots, n_i$ die Designmatrix, $X \in \mathfrak{R}^{n \times b}$ die Matrix der festen Effekte $\beta \in \mathfrak{R}^{b \times n_i}$, $Z \in \mathfrak{R}^{n \times g}$ die Zufallsmatrix $\alpha \in \mathfrak{R}^{g \times n_i}$ und die Fehlermatrix $\epsilon \in \mathfrak{R}^{n \times n_i}$ gegeben. Dann sieht das GLMM in

⁴⁴siehe: [(Liard und Ware 1982)] S.965

⁴⁵siehe: [(Jones 1993)]S.26

Matrixschreibweise folgendermaßen aus:

$$\begin{aligned}
 \begin{pmatrix} Y_{11} & \cdots & Y_{1n_i} \\ \vdots & & \vdots \\ Y_{n1} & \cdots & Y_{nn_i} \end{pmatrix} &= \begin{pmatrix} X_{11} & \cdots & X_{1b} \\ \vdots & & \vdots \\ X_{n1} & \cdots & X_{nb} \end{pmatrix} \begin{pmatrix} \beta_{11} & \cdots & \beta_{1n_i} \\ \vdots & & \vdots \\ \beta_{b1} & \cdots & \beta_{bn_i} \end{pmatrix} \\
 &+ \begin{pmatrix} Z_{11} & \cdots & Z_{1g} \\ \vdots & & \vdots \\ Z_{n1} & \cdots & Z_{ng} \end{pmatrix} \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n_i} \\ \vdots & & \vdots \\ \alpha_{g1} & \cdots & \alpha_{gn_i} \end{pmatrix} \\
 &+ \begin{pmatrix} \epsilon_{11} & \cdots & \epsilon_{1n_i} \\ \vdots & & \vdots \\ \epsilon_{n1} & \cdots & \epsilon_{nn_i} \end{pmatrix} \tag{25}
 \end{aligned}$$

Donald Hedeker⁴⁶ macht in seinem Artikel deutlich, dass im GLMM häufig der Logit- Link verwendet wird. Die normale Linkfunktion, welche beim GLMM sonst verwendet wird, wird später erläutert.

Definition 5.3 (Logit- Link- Funktion)

Ist eine Linkfunktion definiert durch:

$$g(\mu_{ij}) = \text{logit}(\mu_{ij}) = \log\left(\frac{\mu_{ij}}{1 - \mu_{ij}}\right) = \eta_{ij}$$

wobei hier wieder $\mu_{ij} = E(y_{ij}|\alpha, x_{ij})$ ist, dann wird die Linkfunktion als „Logit Link Funktion“ bezeichnet.

Man sieht, dass es zwei Schlüsselemente gibt, welche ein GLMM definieren. Zum Einen ist es die bedingte Unabhängigkeit (der Zufallseffekte selbst), die gegeben sein muss und zum Anderen die Verteilung der Zufallseffekte.

Dieses sind beides auch die Schlüsselemente eines gaußschen linearen gemischten Modells, worauf schon im Kapitel (4) (letzter Abschnitt) eingegangen wurde.

Wenn man nun einen Vektor von Zufallsvariablen hat, sind die Realisierungen y_1, \dots, y_n stochastisch unabhängig, so dass die Verteilung von y folgendermaßen aussieht:

$$(f_i)_{\alpha, \phi}(y_i) = \exp\left(\frac{y_i \theta_i - b(\theta_i)}{a_i(\phi)} + c_i(y_i, \phi)\right) \tag{26}$$

⁴⁶siehe: [(Hedeker)]

Dies ist wieder die Dichtefunktion einer Exponentialfamilie nach Definition (4.7), $b(\cdot)$, $a_i(\cdot)$, $c_i(\cdot, \cdot)$ sind bekannte Funktionen und ϕ ist hier wieder der Dispersionsparameter (siehe Definition (4.7))

Wie bereits erwähnt kann in den meisten Fällen die Logit-Link-Funktion verwendet werden. Die allgemeine Linkfunktion für das GLMM ist definiert durch:

$$g(\mu_i) = \eta_i$$

und es ist $b'(\theta_i) = \mu_i$.

Unter dem kanonischen Link ist nun $\theta_i = \eta_i = g(\mu_i)$ und es gilt folgendes Lemma:

Lemma 5.1

Für die kanonische Linkfunktion gilt, dass $g = h^{-1}$, wobei $h(\cdot) = b'(\cdot)$, bzw. $g(\cdot) = (b')^{-1}(\cdot)$. h^{-1} ist hier die sogenannte inverse Linkfunktion.

Beweis. Es ist zu zeigen, dass $g = (b')^{-1}$

Nach Satz 4.2 gilt, dass $\mu_i = E(y_i) = b'(\theta_i)$, daraus folgt, da der kanonische Link vorliegt, (d.h. es gilt: $\theta_i = g(\mu_i)$):

$$(b')^{-1}(\mu_i) = \theta_i = (b')^{-1}(g^{-1}(x_i^t \beta))$$

damit folgt: $g(\mu_i) = (b')^{-1}(\mu_i) \Rightarrow g = (b')^{-1}$ □

Zur Veranschaulichung des GLMMs, wird nun ein kleines Beispiel folgen, welches zeigt, dass unter bestimmten Bedingungen das lineare gemischte Modell ein GLMM ist.

Beispiel 5.1 (Normales lineares gemischtes Modell)

Das allgemeine lineare gemischte Modell

$$y = X\beta + Z\alpha + \epsilon$$

mit dem Beobachtungsvektor y , den Vektoren β und α der festen und zufälligen Effekte und den Designmatrizen X und Z ist ein Spezialfall des verallgemeinerten linearen gemischten Modells, sofern die Varianz des Fehlervektors $\text{var}(\epsilon) = \sigma^2$.

Die Exponentialfamilie ist hier normalverteilt mit Erwartungswert μ und Varianz σ^2 . In Beispiel (4.1) wurde bereits gezeigt, dass in diesem Fall der Dispersionsparameter $\phi = \sigma^2$ und der Parametervektor $\theta = \mu$. Die Linkfunktion

ist hier der kanonische Link mit $g(\mu) = \theta = \mu$.

Der kanonische Link wird neben dem Logit Link häufig beim GLMM verwendet. Eine weitere Linkfunktion, welche beim GLMM oft zur Anwendung kommt, ist der Probit, bei dem $g(\mu) = \Phi^{-1}(\mu)$, wobei Φ hier die Standardnormalverteilung also die Normalverteilung mit Erwartungswert 0 und Varianz 1 darstellt.

Auf weitere Beispiele wird hier nicht eingegangen, sie können aber in [(Jiang 2007)] S.122 nachgelesen werden.

5.1 Maximum-Likelihood-Funktion und deren Annäherung

Wie bei jedem statistischen Modell spielen auch beim GLMM Schätzungen eine bedeutende Rolle. Im weiteren Verlauf dieser Arbeit wird nun die Maximum-Likelihood-Schätzung beim GLMM vertieft. Wie gleich zu sehen ist, ist es nur sehr schwierig möglich die Likelihood-Funktion auf einem normalen Weg zu berechnen, sondern es wird hier mit Annäherungen gearbeitet.

Anders als beim linearen gemischten Modell hat die Likelihood-Funktion ausgenommen vom Normalfall, der im vorhergehenden Beispiel (5.1) aufgeführt wurde, hier keine geschlossene Form.

Bemerkung 5.2

Da bei der Likelihood-Funktion des GLMMs oft Integrale mit höheren Dimensionen entstehen, kann es bei der Berechnung zu gravierenden Problemen kommen. Genau in diesem Fall muss man eine andere Möglichkeit finden, sich die Funktion herzuleiten, d.h. man muss eine Alternative finden, die in diesem Fall die Annäherung ist. Eine gängige Methode eine Annäherung zu berechnen ist die Laplace-Methode, die nun erklärt wird.

Satz 5.1 (Laplace-Approximation)

Sei

$$\int_{-\infty}^{+\infty} \exp(-q(x)) dx \quad (27)$$

das Integral, welches angenähert werden soll. Dabei ist $q(\cdot)$ eine Funktion, welche bei dem Wert $x = x_0$ ein Minimum besitzt, es ist also $q'(x_0) = 0$ und $q''(x_0) > 0$, wobei die Striche hier den Grad der Ableitung angeben. Die

Annäherung ist gegeben durch:

$$\int_{-\infty}^{+\infty} \exp(-q(x)) dx \approx \exp(-q(x_0)) \sqrt{\frac{2\pi}{q''(x_0)}} \quad (28)$$

Beweis. Es muss gezeigt werden, dass $\exp(-q(x_0)) \sqrt{\frac{2\pi}{q''(x_0)}}$ tatsächlich, die Annäherung für das Integral: $\int_{-\infty}^{+\infty} \exp(-q(x)) dx$ ist:

Dafür benötigt man das *Gaußsche Fehlerintegral*. Das ist das uneigentliche Riemann-Integral $\int_{-\infty}^{+\infty} \exp(-x^2) dx$. Dieses gaußsche Fehlerintegral spielt in der Wahrscheinlichkeitstheorie und für das eben aufgestellte Integral (27), wie gleich zu sehen ist, eine bedeutende Rolle. Es gilt nun folgende Gleichung:

$$\int_{-\infty}^{+\infty} \exp(-x^2) dx = \sqrt{\pi} \quad (29)$$

Das Gleichung (29) gilt, muss zunächst bewiesen werden:

Dafür sei:

$$\exp(-(x^2 + y^2)) = \exp(-r^2)$$

wobei das Ganze auf dem \mathfrak{R}^2 gilt.

Das Integral wird nun durch folgende zwei Schritte berechnet:

1. Zunächst wird eine doppelte Integration berechnet. Es ist:

$$\begin{aligned} \int_{-\infty}^{+\infty} \exp(-x^2) dx &= \sqrt{\left(\int_{-\infty}^{+\infty} \exp(-x^2) dx \right) \left(\int_{-\infty}^{+\infty} \exp(-x^2) dx \right)} \\ &= \sqrt{\left(\int_{-\infty}^{+\infty} \exp(-y^2) dy \right) \left(\int_{-\infty}^{+\infty} \exp(-x^2) dx \right)} \\ &= \sqrt{\left(\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp(-(x^2 + y^2)) dy dx \right)} \end{aligned}$$

Hier wird von Integraleigenschaften Gebrauch gemacht und davon, dass die Variable innerhalb des Integrals eine sogenannte „dummy variable“ (Platzhaltervariable) ist, somit kann x durch y ersetzt werden.

2. Als zweites muss das Integral berechnet werden, indem man sich zunächst die Polarkoordinaten anschaut. Es sei in den folgenden Gleichungen: $r = \sqrt{x^2 + y^2}$ und $\theta = \tan^{-1}(\frac{y}{x})$. Man substituiert nun mit $x := r \cos \theta$ und

$y := r \sin \theta$, $dx dy := r(dr d\theta)$:

$$\begin{aligned}
 \int_{-\infty}^{+\infty} \exp(-x^2) dx &= \sqrt{\left(\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp(-(x^2 + y^2)) dy dx \right)} \\
 &= \sqrt{\int_0^{2\pi} \int_0^{\infty} r \exp(-r^2) dr d\theta} \\
 &= \sqrt{2\pi \int_0^{\infty} r \exp(-r^2) dr} \\
 &= \sqrt{2\pi \int_0^{\infty} -\frac{1}{2} \exp(s) ds} \\
 &= \sqrt{2\pi \int_{-\infty}^0 \frac{1}{2} \exp(s) ds} \\
 &= \sqrt{\pi \int_{-\infty}^0 \exp(s) ds} \\
 &= \sqrt{\pi(\exp(0) - \exp(-\infty))} \\
 &= \sqrt{\pi(1 - 0)} \\
 &= \sqrt{\pi}
 \end{aligned}$$

dabei wurde substituiert mit $s := -r^2$, $ds := -2r dr$

Mit dem gaußschen Fehlerintegral kann nun auch das Integral

$\int_{-\infty}^{+\infty} \exp(-q(x)) dx$ (27) berechnet werden:

Dafür wird die Taylor-Entwicklung⁴⁷ verwendet. Diese ist gegeben durch:

$$q(x) = q(x_0) + \frac{q''(x_0)}{2!} (x - x_0)^2 + \dots \quad (30)$$

Der zweite Term der Taylor-Entwicklung ($\frac{q''(x_0)}{2!} (x - x_0)^2$) kann weggelassen werden, da er Null ergibt ($q'(x_0) = 0$). Für die Taylorreihe muss die Funktion natürlich stetig auf einem Intervall und dort n-mal differenzierbar sein.

Mit der Taylorreihen-Entwicklung (wo man nun also anstatt $q(x)$ die angenä-

⁴⁷siehe: [(Heuser 2001)] S.253-257

herte Funktion $q(x_0) + \frac{q''(x_0)}{2!}(x - x_0)$ einsetzt) gilt dann:

$$\begin{aligned} Q(x) &:= \int_{-\infty}^{+\infty} \exp(-q(x)) dx \\ &\approx \int_{-\infty}^{+\infty} \exp\left(-q(x_0) + \frac{q''(x_0)}{2!}(x - x_0)^2\right) dx \\ &= \exp(-q(x_0)) \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}q''(x_0)(x - x_0)^2\right) dx \end{aligned}$$

Nun wird substituiert mit:

$$y := (x - x_0)$$

und

$$\begin{aligned} c &:= \sqrt{\frac{2}{q''(x_0)}} \\ \Rightarrow Q(x) &\approx \exp(-q(x_0)) \int_{-\infty}^{+\infty} \exp\left(-\frac{y^2}{c^2}\right) dy \end{aligned}$$

Durch nochmalige Substitution mit

$$z := \frac{y}{c}, \frac{dy}{dz} \Rightarrow dy = dz \cdot c$$

ergibt sich:

$$\begin{aligned} \Rightarrow Q(x) &= \exp(-q(x_0)) c \int_{-\infty}^{+\infty} \exp(-z^2) dz \\ &= \exp(-q(x_0)) \sqrt{\frac{2}{q''(x_0)}} \sqrt{\pi} \\ &= \exp(-q(x_0)) \sqrt{\frac{2\pi}{q''(x_0)}} \end{aligned}$$

Hier wurde die Konstante c wieder eingesetzt und berücksichtigt, dass $\int_{-\infty}^{+\infty} \exp(-z^2) dz$ das gaußsche Fehlerintegral ist.

Damit gilt:

$$\begin{aligned}
 Q(x) &= \int_{-\infty}^{+\infty} \exp(-q(x)) dx \\
 &\approx \int_{-\infty}^{+\infty} \exp\left(-\left(q(x_0) + \frac{q''(x_0)}{2!}(x-x_0)^2\right)\right) dx \\
 &= \exp(-q(x_0)) \sqrt{\frac{2\pi}{q''(x_0)}}
 \end{aligned}$$

□

Der nächste Satz folgt ohne Beweis, da es die multivariate Variante von Satz (5.1) ist.⁴⁸

Satz 5.2 (Laplace-Approximation für den multivariaten Fall)

Sei

$$\int_{-\infty}^{+\infty} \exp(-q(\alpha)) d\alpha \tag{31}$$

das Integral, welches angenähert werden soll, mit der Funktion $q(\alpha)$, die ein Minimum an der Stelle $\alpha = \alpha_0$ hat.

(Es ist also $q'(\alpha_0) = 0$ und $A := q''(\alpha_0) > 0$, wobei hier q' den Vektor der ersten Ableitung und $A := q''(\alpha_0)$ die Hessematrix darstellt, welche positiv definit ist.)

Die Annäherung für (31) ist im multivariaten Fall gegeben durch:

$$\int_{-\infty}^{+\infty} \exp(-q(\alpha)) d\alpha \approx \int_{-\infty}^{+\infty} \exp\left(-\left(q(\alpha_0) + \frac{A}{2!}(\alpha - \alpha_0)\right)\right) \tag{32}$$

$$= \exp(-q(\alpha_0)) c |A|^{-\frac{1}{2}} \tag{33}$$

c ist hier eine Konstante, die von der Dimension des Integrals abhängt und die Begrenzungssymbole $|A|$, geben in diesem Fall die Determinante der Hessematrix A an. Es ist also: $|A| = |q''(\alpha_0)|$

In [(Molenberghs und Verbeke 2005)] wird neben der Annäherung des Integrals (anhand der „Gauß-Quadratur“⁴⁹) noch eine weitere numerische Approximationsmethode bezüglich der Maximum-Likelihood-Schätzung erwähnt. Es handelt sich hier um die Approximation der Daten. Diese wird im Folgenden

⁴⁸siehe [(Jiang 2007)] S.127

⁴⁹Siehe: (Molenberghs und Verbeke 2005) S.274-275

näher untersucht und es wird zudem auf ihre beiden Methoden (siehe Unterkapitel (5.2) und (5.3)) eingegangen.

Diese weitere Klasse von Annäherungen erhält man durch eine spezielle Zerlegung der vorhandenen Daten in den Erwartungswert μ_{ij} und seinen entsprechenden Fehlerterm ϵ_{ij} . Wenn man dieses mit der Taylor-Entwicklung des Erwartungswertes macht, dann ist das eine nicht-lineare Funktion des linearen Prädiktors. Alle Vorgehensweisen unterscheiden sich, je nachdem welche Ordnung die Taylorreihen-Entwicklung besitzt. Eine Zerlegung sieht folgendermaßen aus:

$$Y_{ij} = \mu_{ij} + \epsilon_{ij} = h((x_{ij})^t \beta + (z_{ij})^t \alpha_i) + \epsilon_{ij}$$

Hierbei ist $h(\cdot)$ wieder die inverse Linkfunktion und die Fehlerterme ϵ_{ij} mit $i = 1, \dots, n$ und $j = 1, \dots, n_i$ besitzen eine geeignete Verteilung.

Lemma 5.2

Die Varianz ist gegeben durch:

$$Var(Y_{ij}|\alpha_i) = \phi\nu(\mu_{ij}) \tag{34}$$

wobei hier $\nu(\cdot)$ die Varianzfunktion der Exponentialfamilie ist.⁵⁰

Beweis. Die Varianz (34) kommt folgendermaßen zu Stande: Nach Satz (4.2) gilt, dass $Var(Y) = \phi b''(\theta)$ (unter Auslassung der Indizes i und j), sofern die Gewichte $w = 1$ sind. (normalerweise gilt: $Var(Y) = \frac{\phi}{w} b''(\theta)$) b'', ϕ, θ sind hier wie in der Definition der 1-parametrischen Exponentialfamilie definiert. Nun hängen Erwartungswert und Varianz in der folgenden Beziehung zusammen:

$$Var(Y) = \phi b''(\theta) = \phi b''((b')^{-1}(\mu)) = \phi\nu(\mu)$$

mit $E(Y) = \mu = b'(\theta) \Leftrightarrow (b')^{-1}(\mu) = \theta$ und $\nu(\cdot) := b''((b')^{-1}(\cdot))$ □

Bemerkung 5.3

Wenn die natürliche Linkfunktion betrachtet wird, so sieht die Varianzfunktion

⁵⁰Durch die Varianzfunktion werden lineare Exponentialfamilien charakterisiert. Sie tragen eine wichtige Funktion dazu bei, den Parameter β zu schätzen. Die Varianzfunktion für die Normalverteilung ist z.B. $V(\mu) = 1$, für die Poissonverteilung $V(\mu) = \mu$, für die Gammaverteilung $V(\mu) = \mu^2$ und für die Binominalverteilung $V(\mu) = \mu(1 - \mu)$. Siehe: [(Hatzinger 1991)]

folgendermaßen aus:

$$\nu(\widehat{\mu}_{ij}) = h'(x_{ij}^t \widehat{\beta} + z_{ij}^t \widehat{\alpha}_i) \quad (35)$$

Die Schätzung für den allgemeinen Parameter θ erhält man, wie schon erwähnt, von der Maximum-Likelihood-Schätzung. Bei den Zufallseffektmodellen, unter das auch das GLMM fällt, erhält man diese Schätzung durch die Maximierung der marginalen Likelihood-Funktion, welche durch das Integrieren über den Zufallseffekten erhalten wurde.⁵¹ Nun sei die marginale Dichtefunktion von Y_i gegeben durch:

$$(f_i)_{\theta, Q}(y_i) = \int (f_i)_{\alpha_i, \theta}(y_i) dQ(\alpha_i) \quad (36)$$

mit $Q(\alpha_i)$ der Verteilungsfunktion von den Zufallseffekten α_i . Diese Funktion wird in [(Molenberghs und Verbeke 2005)] auch als gemischte Verteilung (*mixing distribution*) bezeichnet.

Die Modelle, die Zufallseffekte beinhalten, können somit anhand der Maximum-Likelihood-Schätzung, durch die Integration der Zufallseffekte, angepasst werden. Die marginale Dichtefunktion (36) für jedes einzelne Subjekt beschreibt nun die folgende Formel:

$$(f_i)_{\beta, D, \phi}(y_i) = \int \prod_{j=1}^{n_i} (f_{ij})_{\alpha_i, \beta, \phi}(y_{ij}) f_D(\alpha_i) d\alpha_i \quad (37)$$

Hier setzt man also die Parameter β , D und ϕ gemäß (36) ein. Man erhält nun die Likelihood-Funktion, indem man die Dichtefunktionen der einzelnen Subjekte betrachtet und multipliziert. Folglich ergibt sich für die Likelihood-Funktion:

$$L_{(\beta, D, \phi)}(y) = \prod_{i=1}^N (f_i)_{\beta, D, \phi}(y_i) \quad (38)$$

$$= \prod_{i=1}^N \int \prod_{j=1}^{n_i} (f_{ij})_{\alpha_i, \beta, \phi}(y_{ij}) f_D(\alpha_i) d\alpha_i \quad (39)$$

Hier ist $f_D(\alpha_i)$ die Dichtefunktion der Normalverteilung

$$N(0, D) := N(0, \sigma^2 B)$$

⁵¹siehe: [(Molenberghs und Verbeke 2005)] S.266

der Zufallseffekte α_i . (siehe oben)

In den folgenden beiden Abschnitten wird es nun um zwei Methoden zur Approximation der Daten gehen. Dabei spielt die „*Quasi-Likelihood-Schätzung*“ eine wichtige Rolle.

5.2 Penalized-Quasi-Likelihood-Schätzung (PQL)

Da für die Maximum-Likelihood-Schätzung die Likelihood-Funktion und nicht nur eine angenäherte Form davon benötigt wird, wird nun folgendermaßen fortgefahren:

Mit der angenäherten Laplace-Approximation erhält man eine Likelihood-Funktion, welche in diesem Fall als „*Quasi-Likelihood-Funktion mit Strafterm*“ oder auch als „*Penalized- oder Pseudo-Quasi-Likelihood-Funktion (PQL)*“ bezeichnet wird und arbeitet mit dieser nun so weiter als wäre es die wahre Likelihood-Funktion. Diese Pseudo-Quasi-Likelihood-Schätzung maximiert die (Taylor-) Approximation an die Likelihood-Funktion (bzw. an die Quasi-Likelihood-Funktion) und ergibt, wie bereits erwähnt, nur approximative Schätzer.

Bevor die PQL präzisiert wird, muss zunächst erklärt werden, was eine Quasi-Likelihood-Funktion ist und wie sie gebildet wird. In den meisten Fällen, in denen diskrete unabhängige Daten vorliegen, ist es numerisch sehr aufwendig, oder gar unmöglich, die Maximum-Likelihood-Schätzung zu bestimmen, egal was für ein Algorithmus angewendet wird. Es besteht allerdings die Möglichkeit die nicht-lineare Beziehung zwischen den abhängigen und unabhängigen Variablen durch eine Taylorreihenentwicklung (siehe auch Unterkapitel (5.1)) zu linearisieren, um doch noch zu Schätzungen der Parameter zu gelangen. Nach der Linearisierung können die Parameter dann leichter bestimmt werden und diese Methode wird als „*Quasi-Likelihood-Schätzung*“ bezeichnet.

Nach [(McCullagh und Nelder 1989)] S.324 wird die Quasi-Likelihood-Funktion folgendermaßen konstruiert: Seien die Komponenten eines Beobachtungsvektors Y als unabhängig vorausgesetzt mit Erwartungswertvektor μ und Kovarianzmatrix $\sigma^2 V(\mu)$, wobei σ^2 unbekannt und $V(\mu)$ die Matrix von bekannten Varianzfunktionen ist. Wenn man nun eine einzelne Komponente y von dem Vektor Y betrachtet, gilt:

$$\frac{\partial l}{\partial \mu} := U = \frac{y - \mu}{\sigma^2 V(\mu)}$$

Lemma 5.3 (Eigenschaften von U)

Die Funktion $U := \frac{y-\mu}{\sigma^2 V(\mu)}$ hat die folgenden Eigenschaften mit der Ableitung einer Loglikelihood gemeinsam:

$$\begin{aligned} E(U) &= 0 \\ -E\left(\frac{\partial U}{\partial \mu}\right) &= \frac{1}{\sigma^2 V(\mu)} \\ \text{Var}(U) &= \frac{1}{\sigma^2 V(\mu)} \end{aligned}$$

Beweis.

$$\begin{aligned} E(U) &= E\left(\frac{y-\mu}{\sigma^2 V(\mu)}\right) = \frac{1}{\sigma^2 V(\mu)} E(y) - \frac{1}{\sigma^2 V(\mu)} E(\mu) \\ &= \frac{\mu}{\sigma^2 V(\mu)} - \frac{\mu}{\sigma^2 V(\mu)} = 0 \\ -E\left(\frac{\partial U}{\partial \mu}\right) &= -E\left(\frac{\partial^2 l}{\partial \mu^2}\right) = E\left(-\frac{\partial^2 l}{\partial \mu^2}\right) = E\left(\frac{V(\mu) + (y-\mu)V(\mu)}{\sigma^2 \{V(\mu)\}^2}\right) \\ &= \frac{E(V(\mu)) + E(yV(\mu)) - E(\mu V(\mu))}{E(\sigma^2 \{V(\mu)\}^2)} \\ &= \frac{V(\mu) + \mu V(\mu) - \mu V(\mu)}{\sigma^2 \{V(\mu)\}^2} = \frac{V(\mu)}{\sigma^2 \{V(\mu)\}^2} = \frac{1}{\sigma^2 V(\mu)} \\ \text{Var}(U) &= E((U)^2) - (E(U))^2 \\ &= E\left(\left(\frac{y-\mu}{\sigma^2 V(\mu)}\right)^2\right) - 0^2 \\ &= -E\left(\frac{\partial^2 l}{\partial \mu^2}\right) = \frac{1}{\sigma^2 V(\mu)} \end{aligned}$$

□

Bemerkung 5.4

Das folgende Integral verhält sich nun aufgrund der Eigenschaften von U , sofern es existiert, wie eine Loglikelihood-Funktion für μ :

$$\int_y^\mu \frac{y-u}{\sigma^2 V(\mu)} du$$

Definition 5.4 (Quasi-Likelihood-Funktion)

$Q(\mu, y)$ wird Quasi-Likelihood-Funktion für μ bzgl. y genannt, wenn $Q(\mu, y)$

durch folgende beiden Eigenschaften definiert ist:

$$\frac{\partial Q(\mu, y)}{\partial \mu} = \frac{y - \mu}{\sigma^2 V(\mu)} \quad (40)$$

oder äquivalent dazu:⁵²

$$Q(\mu, y) = \int_y^\mu \frac{y - u}{\sigma^2 V(u)} du + \text{Dichtefunktion von } y \quad (41)$$

Bemerkung 5.5

Sind die Komponenten von Y unabhängig, sieht die Quasi-Likelihood-Funktion folgendermaßen aus:

$$Q(\mu, y) = \sum_i^n Q_i(\mu_i, y_i)$$

Sei nun der bedingte Erwartungswert von der Beobachtung y_i mit $(1 \leq i \leq n)$ gegeben durch:

$$E(y_i | \alpha) = h(x_i \beta + z_i \alpha)$$

wobei $\alpha = (\alpha_1, \dots, \alpha_n)^t$ den Vektor der Zufallseffekte darstellt. h ist die inverse Linkfunktion von der bekannten Linkfunktion g ($h^{-1} = g$) und β ist wieder der Vektor der unbekannt Parameter der festen Effekte, x_i und z_i sind bekannte Vektoren. Weiter sei

$$\mu_i = E(y_i | \alpha)$$

und damit dann auch:

$$\begin{aligned} \mu_i &= h(x_i \beta + z_i \alpha) \\ \Rightarrow h^{-1}(\mu_i) &= x_i \beta + z_i \alpha \\ \Rightarrow \eta_i = g(\mu_i) &= x_i \beta + z_i \alpha \end{aligned}$$

da $h^{-1} = g$. Es wird weiter angenommen, dass die bedingte Varianz von y_i unter α gegeben ist durch:

$$var(y_i | \alpha) = a_i(\phi) \nu(\mu_i)$$

⁵²siehe: (Wedderburn 1974) S.440

wobei hier ϕ ein zusätzlicher Dispersionsparameter, $a_i(\cdot)$ eine bekannte Funktion der Form $a_i(\phi) = \frac{\phi}{w_i}$ und $\nu(\cdot)$ eine bekannte Varianzfunktion ist. Die Varianzfunktion wird in Übereinstimmung mit einem bestimmten Wert aus der Exponentialfamilie gewählt.⁵³

Es ist zu erkennen, dass die Annahmen schwächer sind, als bei der eben gegebenen Definition eines GLMMs, wo zusätzlich vorausgesetzt wird, dass y_i die Verteilung einer 1-parametrischen Exponentialfamilie besitzt.

Bei der Quasi-Likelihood-Schätzung ist als erstes die Ableitung der Quasi-Likelihood-Funktion zu bestimmen:

Dazu wird angenommen, dass die Realisierungen y_1, \dots, y_n der Zufallsvariablen Y_1, \dots, Y_n unabhängig sind und der Vektor der Zufallseffekte α eine multivariate Normalverteilung mit Erwartungswert 0 und Kovarianzmatrix D besitzt.

D hat eine spezielle Form, die auf den Vektor des Dispersionsparameters θ zurückgeht.

Die Quasi-Likelihood-Funktion L_Q von $y = (y_1, \dots, y_n)^t$ (berechnet nach Formel (41)) ist gegeben durch:⁵⁴

$$L_Q \propto |D|^{-\frac{1}{2}} \int \exp \left(-\frac{1}{2} \sum_{i=1}^n d_i - \frac{1}{2} \alpha^t D^{-1} \alpha \right) d\alpha \quad (42)$$

mit

$$d_i = -2 \int_{y_i}^{\mu_i} \frac{y_i - u}{a_i(\phi)v(u)} du \quad (43)$$

der sogenannten *Deviance*.

Bemerkung 5.6

Die *Deviance* ist ein Maß für die Diskrepanz⁵⁵ und gibt die Abweichung vom Idealwert an. Man kann sie mathematisch auch als „minus zweifache der logarithmierten Likelihood-Funktion“ ansehen.

Wenn nun die Laplace-Approximation (32) verwendet wird, sieht die loga-

⁵³siehe: [(Molenberghs und Verbeke 2005)]

⁵⁴siehe: [(Jiang 2007)] S.128

⁵⁵Die Diskrepanz ist bei Anpassungstests (Goodness-of-Fit Tests) von Bedeutung, wobei überprüft wird, ob ein bestimmtes Modell vorliegt.

rithmierte Quasi-Likelihood-Funktion von L so aus:

$$l_Q \approx \log \left(c |D|^{-\frac{1}{2}} \left| q''(\alpha_0) \right|^{-\frac{1}{2}} \exp(-q(\alpha_0)) \right) \quad (44)$$

$$= c - \frac{1}{2} \log |D| - \frac{1}{2} \log \left| q''(\alpha_0) \right| - q(\alpha_0) \quad (45)$$

wobei α_0 derjenige Punkt ist, an dem die Funktion

$$\begin{aligned} q(\alpha) &= \frac{1}{2} \left(\sum_{i=1}^n d_i + \alpha^t D^{-1} \alpha \right) \\ &= \frac{1}{2} \left(\sum_{i=1}^n -2 \int_{y_i}^{u_i} \frac{y_i - u}{a_i(\phi)v(u)} du + \alpha^t D^{-1} \alpha \right) \\ &= -2 \frac{1}{2} \sum_{i=1}^n \int_{y_i}^{u_i} \frac{y_i - u}{a_i(\phi)v(u)} du + \frac{1}{2} \alpha^t D^{-1} \alpha \\ &= \frac{1}{2} \alpha^t D^{-1} \alpha - \sum_{i=1}^n \int_{y_i}^{u_i} \frac{y_i - u}{a_i(\phi)v(u)} du \end{aligned}$$

(mit der Deviance d_i von eben eingesetzt), ein Minimum besitzt. Es ist also $q'(\alpha_0) = 0$. Es ergibt sich die folgende partielle Ableitung:

$$\frac{\partial q(\alpha)}{\partial \alpha} = q'(\alpha) = D^{-1} \alpha - \sum_{i=1}^n \left(\int_{y_i}^{u_i} \frac{y_i - u}{a_i(\phi)v(u)} du \right)' \quad (46)$$

$$= D^{-1} \alpha - \sum_{i=1}^n \frac{y_i - \mu_i}{a_i(\phi)v(\mu_i)g'(\mu_i)} z_i = 0 \quad (47)$$

mit

$$\mu_i = x_i^t \beta + z_i^t \alpha \quad (48)$$

Wird die Funktion $q'(\alpha)$ nochmals abgeleitet, erhält man die folgende zweite Ableitung (Hessematrix):

$$q''(\alpha) = D^{-1} + \sum_{i=1}^n \frac{z_i z_i^t}{a_i(\phi)v(\mu_i)[g'(\mu_i)]^2} + r \quad (49)$$

mit $r = - \sum_{i=1}^n (y_i - \mu_i) z_i \frac{\partial}{\partial \alpha} \left[\frac{1}{a_i(\phi)v(\mu_i)g'(\mu_i)} \right]$. Bedenkt man, dass $a_i(\phi) = \frac{\phi}{w_i}$ gilt und den Nenner des Bruches, also $(a_i(\phi)v(\mu_i)[g'(\mu_i)])$ gleich w_i^{-1} setzt, erhält man, unter Auslassen des Terms r , folgende Approximation der Hessematrix

von Gleichung (49):

$$q''(\alpha) \approx Z^t W Z + D^{-1}$$

Z ist die Matrix mit dem Vektor z_i in der i -ten Zeile, wobei $1 \leq i \leq n$ und W ist die Diagonalmatrix mit w_1, \dots, w_n in der Diagonale.

Es gilt somit: $W = \begin{pmatrix} w_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_n \end{pmatrix}$. Kombiniert man nun die beiden Annä-

herungen (45) und (49) miteinander, erhält man für die logarithmierte Likelihood-Funktion folgende Approximation:

$$\begin{aligned} l_Q &\approx c - \frac{1}{2} \log |D| - \frac{1}{2} \log |Z^t W Z + D^{-1}| - \frac{1}{2} \left(\sum_{i=1}^n \tilde{d}_i + \alpha_0^t D^{-1} \alpha_0 \right) \\ &= c - \frac{1}{2} \left(\log |D| + \log |Z^t W Z + D^{-1}| + \sum_{i=1}^n \tilde{d}_i + \alpha_0^t D^{-1} \alpha_0 \right) \\ &= c - \frac{1}{2} \left(\log |Z^t W Z D + D^{-1} D| + \sum_{i=1}^n \tilde{d}_i + \alpha_0^t D^{-1} \alpha_0 \right) \\ &= c - \frac{1}{2} \left(\log |I + Z^t W Z D| + \sum_{i=1}^n \tilde{d}_i + \alpha_0^t D^{-1} \alpha_0 \right) \end{aligned}$$

wobei \tilde{d}_i das d_i mit α_0 anstelle α aus (43) ist.

Wird angenommen, dass die Gewichte als eine Funktion von dem Erwartungswert gegeben sind, dann kann der erste Term der letzten Gleichung, also $\frac{1}{2} (\log |I + Z^t W Z D|)$ ignoriert werden⁵⁶, da dieser nur von β von W abhängt. Eine Approximation ist nun gegeben durch:

$$l_{PQ} = c - \frac{1}{2} \left(\sum_{i=1}^n \tilde{d}_i + \alpha_0^t D^{-1} \alpha_0 \right) \quad (50)$$

Gleichung (50) maximiert nun die PQL von [(Green 1987)]. Das ist die folgende Funktion:

$$L(\beta, \alpha) - \frac{1}{2} \lambda J(\alpha) \quad (51)$$

J stellt den Strafterm dar und λ ist eine nichtnegative Konstante. Die Schätzungen, die man erhalten hat, werden Quasi-Likelihood-Schätzungen mit Strafterm genannt, da sie durch eine Optimierung der Quasi-Likelihood-Funktion

⁵⁶siehe: [(Breslow und Clayton 1993)] S.11

mit einem Strafterm J der Zufallseffekte zustande kommen.

Man hat nun eine angenäherte logarithmierte Likelihood-Funktion erhalten und kann damit weiterarbeiten. Es stellt sich hier noch die Frage, ob die PQL-Schätzer konsistent sind oder ob die Approximation Verzerrungen der Schätzer mit sich zieht. Nach [(Jiang 2007)] gibt es natürlich auf Grund der verschiedenen Annäherungen bei der PQL Verzerrungen, allerdings dürfen diese nicht einfach ignoriert werden. Der Vorteil der PQL ist, dass diese im Gegensatz zur „normalen“ Log-Likelihood-Funktion leicht zu berechnen ist.

Zusammenfassend geht es bei der Penalized- oder Pseudo-Quasi-Likelihood-Schätzung darum, eine Annäherung für den Erwartungswert μ_{ij} mit $i = 1, \dots, n$ und $j = 1, \dots, n_i$ zu finden. Das wird anhand einer Taylor-Entwicklung getan und zwar mit den Schätzungen $\hat{\beta}$ und $\hat{\alpha}_i$ der festen und zufälligen Effekte. Die Taylor-Entwicklung wird auf die Zerlegung der Zufallsvariable in den Erwartungswert und den Fehlerterm angewendet:

$$Y_{ij} = \mu_{ij} + \epsilon_{ij} = h(x_{ij}^t \beta + z_{ij}^t \alpha_i) + \epsilon_{ij} \quad (52)$$

wobei h wieder die inverse Linkfunktion darstellt und die Fehler ϵ_{ij} haben eine Verteilung mit einer Varianz die gleich der bedingten Varianz von Y_{ij} und α_i ist, folglich ist $Var(Y_{ij}|\alpha_i) = \phi\nu(\mu_{ij})$. Des Weiteren ist zu beachten, dass mit der normalen Linkfunktion gilt, dass $\nu(\mu_{ij}) = h'(x_{ij}^t \beta + z_{ij}^t \alpha_i)$

Um eine Approximation für Y_{ij} zu finden, verwendet man die Taylorentwicklung für den Erwartungswert $\mu_{ij} = h(x_{ij}^t \beta + z_{ij}^t \alpha_i)$. Man ersetzt folglich den Erwartungswert durch seine Taylorentwicklung:

$$h(x) = h(x_0) + h'(x_0)(x - x_0)$$

mit $x := x_{ij}^t \beta + z_{ij}^t \alpha_i$ und $x_0 = x_{ij}^t \hat{\beta} + z_{ij}^t \hat{\alpha}_i$.

Die Entwicklung für den Erwartungswert, wenn man x und x_0 einsetzt, ist

gegeben durch:

$$\begin{aligned}
 h(x_{ij}^t \beta + z_{ij}^t \alpha_i) &\approx h(x_{ij}^t \hat{\beta} + z_{ij}^t \hat{\alpha}_i) + h'(x_{ij}^t \hat{\beta} + z_{ij}^t \hat{\alpha}_i)((x_{ij}^t \beta + z_{ij}^t \alpha_i) \\
 &\quad - (x_{ij}^t \hat{\beta} + z_{ij}^t \hat{\alpha}_i)) \\
 &= h(x_{ij}^t \hat{\beta} + z_{ij}^t \hat{\alpha}_i) + h'(x_{ij}^t \hat{\beta} + z_{ij}^t \hat{\alpha}_i)(x_{ij}^t \beta - x_{ij}^t \hat{\beta} \\
 &\quad + z_{ij}^t \alpha_i - z_{ij}^t \hat{\alpha}_i) \\
 &= h(x_{ij}^t \hat{\beta} + z_{ij}^t \hat{\alpha}_i) + h'(x_{ij}^t \hat{\beta} + z_{ij}^t \hat{\alpha}_i)(x_{ij}^t (\beta - \hat{\beta}) \\
 &\quad + z_{ij}^t (\alpha_i - \hat{\alpha}_i)) \\
 &= h(x_{ij}^t \hat{\beta} + z_{ij}^t \hat{\alpha}_i) + h'(x_{ij}^t \hat{\beta} + z_{ij}^t \hat{\alpha}_i) x_{ij}^t (\beta - \hat{\beta}) \quad (53) \\
 &\quad + h'(x_{ij}^t \hat{\beta} + z_{ij}^t \hat{\alpha}_i) z_{ij}^t (\alpha_i - \hat{\alpha}_i)
 \end{aligned}$$

Mit der Annäherung (53) ergibt sich die folgende Approximation für Y_{ij} :

$$Y_{ij} \approx h(x_{ij}^t \hat{\beta} + z_{ij}^t \hat{\alpha}_i) + h'(x_{ij}^t \hat{\beta} + z_{ij}^t \hat{\alpha}_i) x_{ij}^t (\beta - \hat{\beta}) + h'(x_{ij}^t \hat{\beta} + z_{ij}^t \hat{\alpha}_i) z_{ij}^t (\alpha_i - \hat{\alpha}_i) + \epsilon_{ij}$$

weiter folgt:

$$Y_{ij} \approx \hat{\mu}_{ij} + \nu(\hat{\mu}_{ij}) x_{ij}^t (\beta - \hat{\beta}) + \nu(\hat{\mu}_{ij}) z_{ij}^t (\alpha_i - \hat{\alpha}_i) + \epsilon_{ij}$$

wobei

$$\hat{\mu}_{ij} = h(x_{ij}^t \hat{\beta} + z_{ij}^t \hat{\alpha}_i) \quad (54)$$

und $\nu(\cdot)$ die Varianzfunktion ist und es gilt: (siehe Gleichung (35))

$$\nu(\hat{\mu}_{ij}) = h'(x_{ij}^t \hat{\beta} + z_{ij}^t \hat{\alpha}_i)$$

In Matrixschreibweise gilt:

$$Y_i \approx \hat{\mu}_i + \hat{V}_i X_i (\beta - \hat{\beta}) + \hat{V}_i Z_i (\alpha_i - \hat{\alpha}_i) + \epsilon_i \quad (55)$$

\hat{V}_i ist eine Diagonalmatrix mit den Einträgen $\nu(\hat{\mu}_{ij})$ in der Diagonale. Es ist

$$\text{also: } \hat{V}_i = \begin{pmatrix} \nu(\hat{\mu}_i) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \nu(\hat{\mu}_i) \end{pmatrix}$$

Wenn man Gleichung (55) umstellt, ergibt sich für die Taylor-Entwicklung von

$$Y_{ij} = h(x_{ij}^t \beta + z_{ij}^t \alpha_i) + \epsilon_{ij}$$

die Annäherung:

$$Y_i^* \equiv \widehat{V}_i^{-1}(Y_i - \widehat{\mu}_i) + X_i\widehat{\beta} + Z_i\widehat{\alpha}_i \quad (56)$$

$$\begin{aligned} &\approx \widehat{V}_i^{-1}((\widehat{\mu}_i + \widehat{V}_i X_i(\beta - \widehat{\beta}) + \widehat{V}_i Z_i(\alpha_i - \widehat{\alpha}_i) + \epsilon_i) - \widehat{\mu}_i) + X_i\widehat{\beta} + Z_i\widehat{\alpha}_i \\ &= \widehat{V}_i^{-1}\widehat{V}_i X_i(\beta - \widehat{\beta}) + \widehat{V}_i^{-1}\widehat{V}_i Z_i(\alpha_i - \widehat{\alpha}_i) + \widehat{V}_i^{-1}\epsilon_i + X_i\widehat{\beta} + Z_i\widehat{\alpha}_i \\ &= X_i\beta - X_i\widehat{\beta} + Z_i\alpha_i - Z_i\widehat{\alpha}_i + \widehat{V}_i^{-1}\epsilon_i + X_i\widehat{\beta} + Z_i\widehat{\alpha}_i \\ &= X_i\beta + Z_i\alpha_i + \epsilon_i^* \end{aligned} \quad (57)$$

Hier ist $\epsilon_i^* := \widehat{V}_i^{-1}\epsilon_i$, wobei deren Erwartungswert 0 ist. Dieses Modell stellt ein vereinfachtes und verbessertes Modell dar, wenn eben diese Pseudodaten betrachtet werden.

Gleichung (57), also die Pseudodaten, befinden sich in der Form eines GLMM, wie immer mit β den festen Effekten, α_i den Zufallseffekten und ϵ_i^* den Fehlertermen. Gleichung (56) ist die Quasi-Likelihood-Schätzung oder besser gesagt die Penalized-Quasi-Likelihood-Schätzung, da sie von einer Optimierung der Likelihood-Funktion erhalten wurde. Da hier Pseudodaten betrachtet werden erklärt sich auch, warum die PQL auch als Pseudo-Likelihood-Schätzung bezeichnet wird. Bei der Approximation der Daten werden die Daten durch normalverteilte Pseudodaten ersetzt mit denen dann weitergerechnet werden kann.

5.3 Marginale-Quasi-Likelihood-Schätzung (MQL)

Als eine Art Alternative zur angenäherten Likelihood-Schätzung gibt es noch eine andere Variante, die *marginale Quasi-Likelihood-Schätzung*. Diese wird kurz vorgestellt, damit im Anschluss ein Vergleich zwischen den beiden Methoden gezogen werden kann, welche Schätzung sich bei der Analyse von Daten besser eignet.

Die Methode der marginalen Quasi-Likelihood-Schätzung (MQL), ist als eine alternative Methode zur Annäherung zu betrachten und ist mit der PQL vergleichbar. Die MQL basiert jedoch auf einer linearen Taylor-Entwicklung des Erwartungswertes $\mu_{ij} = h(x_{ij}^t\beta + z_{ij}^t\alpha_i)$, wobei hier wieder $i = 1, \dots, n$ und $j = 1, \dots, n_i$ ist, die um den Schätzungen von β , also $\widehat{\beta}$ der festen Effekte und um $\alpha_i = 0$ der Zufallseffekte liegt. Bei der MQL wird somit eine ungenauere Taylor-Approximation als bei der PQL verwendet.

Es ergibt sich eine ähnliche Form von Y_{ij} wie bei der PQL, jedoch sieht der

Prädiktor bei der MQL folgendermaßen aus:

$$\widehat{\mu}_{ij} = h(x_{ij}^t \widehat{\beta})$$

Man vergleiche mit (54). Die Annäherung von Y_{ij} ist bei der MQL gegeben durch:

$$\begin{aligned} Y_{ij} &\approx h(x_{ij}^t \widehat{\beta}) + h'(x_{ij}^t \widehat{\beta}) x_{ij}^t (\beta - \widehat{\beta}) + \epsilon_{ij} \\ &= \widehat{\mu}_{ij} + \nu(\widehat{\mu}_{ij}) x_{ij}^t (\beta - \widehat{\beta}) + \epsilon_{ij} \end{aligned}$$

und die Annäherung für Y_i sieht folglich so aus:

$$Y_i \approx \widehat{\mu}_i + \widehat{V}_i X_i (\beta - \widehat{\beta}) + \epsilon_i$$

Das \widehat{V}_i ist wie im letzten Abschnitt gewählt.

Genau wie eben werden nun auch hier die Pseudodaten betrachtet. Diese sehen bei der MQL anders aus als bei der PQL, da es hier, wie eben schon erwähnt, keine Zufallseffekte gibt:

$$\begin{aligned} Y_i^* &\equiv \widehat{V}_i^{-1} (Y_i - \widehat{\mu}_i) + X_i \widehat{\beta} \\ &= \widehat{V}_i^{-1} ((\widehat{\mu}_i + \widehat{V}_i X_i (\beta - \widehat{\beta}) + \epsilon_i) - \widehat{\mu}_i) + X_i \widehat{\beta} \\ &= \widehat{V}_i^{-1} \widehat{V}_i X_i (\beta - \widehat{\beta}) + \widehat{V}_i^{-1} \epsilon_i + X_i \widehat{\beta} \\ &= X_i \beta - X_i \widehat{\beta} + \widehat{V}_i^{-1} \epsilon_i + X_i \widehat{\beta} \\ &= X_i \beta + \epsilon_i^* \end{aligned}$$

wobei, wie im letzten Abschnitt, wieder $\epsilon_i^* := \widehat{V}_i^{-1} \epsilon_i$ ist. Da der Prädiktor der MQL keine Zufallseffekte enthält, haben die Pseudodaten dementsprechend auch keine.

5.3.1 Vergleich von PQL und MQL

Da es zwei Arten von Quasi-Likelihood-Schätzungen gibt, stellt sich die Frage, welche der beiden Funktionen besser geeignet ist, zumal sich die beiden Methoden um Schätzungen zu erhalten nur durch ihren Prädiktor $\widehat{\mu}_{ij}$ ($h(x_{ij}^t \widehat{\beta} + z_{ij}^t \widehat{\alpha}_i)$ für die PQL und $h(x_{ij}^t \widehat{\beta})$ für die MQL) unterscheiden. Der Unterschied besteht somit darin, dass der lineare Prädiktor der MQL keine Zufallseffekte enthält. Es ist deutlich zu erkennen, dass die beiden Methoden auf denselben Schlüs-

selelementen beruhen.⁵⁷ Bei beiden Methoden wird eine Annäherung für das GLMM der Pseudodaten Y_i^* berechnet. Da die Pseudodaten normalverteilt sind, muss die Funktion $\prod_{j=1}^{n_i} (f_{ij})_{\alpha_i, \beta, \phi}(y_{ij})$ aus Gleichung (39) durch die multivariate Normalverteilungsfunktion von Y_i^* ersetzt werden. Es gilt:

$$\begin{aligned} \prod_{j=1}^{n_i} (f_{ij})_{\alpha_i, \beta, \phi}(y_{ij}) &= \exp \left(\sum_{j=1}^{n_i} [y_{ij} \theta_{ij} - b(\theta_{ij})] \frac{w}{\phi} + \sum_{j=1}^{n_i} c(y_{ij}, \phi) \right) \\ &\stackrel{\text{vgl. mit BSP. (4.1)}}{=} \exp \left(\sum_{j=1}^{n_i} \phi^{-1} [y_{ij} \theta_{ij} - b(\theta_{ij})] + \sum_{j=1}^{n_i} c(y_{ij}, \phi) \right) \\ &= \exp \left(\phi^{-1} \left(\beta^t \sum_{j=1}^{n_i} x_{ij} y_{ij} + \alpha_i^t \sum_{j=1}^{n_i} z_{ij} y_{ij} - b(\theta_{ij}) \right) + \sum_{j=1}^{n_i} c(y_{ij}, \phi) \right) \quad (58) \end{aligned}$$

Hier muss also wieder die Dichtefunktion der 1-parametrischen Exponentialfamilie betrachtet werden. Siehe Formel (11) (w ist hier gleich 1).

In (58) sind x_{ij} und z_{ij} die Regressoren für die festen Effekte β und die Zufallseffekte α_i . Das heisst, dass die Annäherung sich genau dann dem korrekten Ergebnis nähert, wenn diese Statistiken, also die Variablen x_{ij} und z_{ij} , approximativ normalverteilt sind. In solch einem Fall muss deshalb ein Test auf Normalverteilung vorgenommen werden. Das kann zum Beispiel der χ^2 Anpassungstest auf Normalverteilung sein. Für den Fall das keine Normalverteilung vorliegt, gibt es noch eine andere Form von Maximum Likelihood Schätzung, die „hierarchische“ Maximum Likelihood (MHL). Diese Schätzung ist eine Erweiterung der PQL und sie erlaubt, dass die Zufallseffekte α_i eine andere Verteilung als die Normalverteilung besitzen.⁵⁸

Nach [(Molenberghs und Verbeke 2005)] sind die Variablen dann normalverteilt, sofern die Beobachtungen y_{ij} stetig sind und die Anzahl n_i der Beobachtungen pro Beobachtungseinheit sehr groß ist.

Ein bedeutender Unterschied von PQL und MQL liegt, wie eben schon dargestellt, in ihrem individuellen linearen Prädiktor. Die Zufallseffekte werden bei der MQL vollkommen ignoriert, so dass sich bei der MQL nur brauchbare Schätzungen ergeben, sofern die Varianz der Zufallseffekte sehr klein ist. Ist die Varianz groß, liegt es nahe sich für die PQL zu entscheiden, da sie die Zufallseffekte berücksichtigt. Die MQL ist somit zwar einfacher zu berechnen

⁵⁷ siehe [(Molenberghs und Verbeke 2005)]

⁵⁸ Siehe: [(Jiang 2007)] S.134

aufgrund der nicht vorhandenen Zufallseffekte, jedoch ist sie in den meisten Fällen keinesfalls die bessere Wahl. Selbst wenn die Anzahl der Beobachtungen pro Subjekt größer wird lassen sich Verzerrungen der MQL nicht vermeiden. In diesem Fall ist die PQL eindeutig zu bevorzugen. Das bedeutet, dass je nach Datenmenge und Datentypen individuell entschieden werden muss, welche Quasi-Likelihood-Schätzung die besten Annäherungen berechnet. Allerdings ist in den meisten Fällen die PQL eine gute Wahl.⁵⁹ In der Datenauswertung in Kapitel (8) wird ebenfalls die PQL-Methode gewählt.

5.4 Tests beim GLMM (Testen der Varianzkomponente)

Die Laplace-Approximation ist lediglich eine Annäherung an das richtige Integral. Nach [(Jiang 2007)] (S.132) entspricht diese Approximation genau dann dem exakten Wert des Integrals, sofern die Varianzkomponente gleich Null ist. Aus diesem Grund wird im Folgenden beschrieben, wie das Testen der Varianz aussieht. Dazu soll der „*globale-Score-Test*“ für das Testen der Varianzkomponente $\theta = 0$ erläutert werden.⁶⁰

Tests spielen, wie bei jedem anderen statistischen Modell, so auch beim GLMM eine bedeutende Rolle.

Oft ist es bei der Auswertung von Daten wichtig auf Überstreuung der Daten zu testen und ob es zwischen den Beobachtungen einen Zusammenhang gibt, d.h. ob sie korreliert sind. Dabei ist zu beachten, dass es zwei Arten von GLMMs gibt, wobei für diese Arbeit lediglich die erste Klasse relevant ist: Es wird unterschieden zwischen dem longitudinalen und dem ANOVA GLMM.

Definition 5.5 (Die beiden Arten des GLMMs)

Es gibt folgende zwei GLMM Typen:

1. *Das longitudinale GLMM ist gegeben durch:*

$$g(\mu_i) = X_i\beta + Z_i\alpha_i \quad (59)$$

mit $i = 1, \dots, n$. Der Vektor μ_i , welcher der Vektor des bedingten Erwartungswertes der Antwortvariable im i -ten Bereich ist, legt die Zufallseffekte fest.

⁵⁹siehe [(Molenberghs und Verbeke 2005)] S.272

⁶⁰Siehe auch: [(Lin 1997)] S.311

$g(\cdot)$ ist wieder eine Linkfunktion, X_i und Z_i sind die Kovarianzmatrizen, β ist wieder der Vektor der festen Effekte und α_i der q -dimensionale Vektor der Zufallseffekte, welcher eine Verteilung, die auf dem Dispersionsparameter θ beruht, besitzt.

2. Das ANOVA⁶¹ GLMM ist gegeben durch:

$$g(\mu) = X\beta + Z_1\alpha_1 + \dots + Z_{n_i}\alpha_{n_i} \quad (60)$$

mit $j = 1, \dots, n_i$ Ähnlich wie beim longitudinalen GLMM, bestimmt der bedingte Erwartungswertvektor $\mu = (\mu_i)_{1 \leq i \leq n}$ der Antwortvariablen die Zufallseffekte.

X ist eine bekannte Kovarianzmatrix, Z_1, \dots, Z_{n_i} sind weitere Designmatrizen, β der Vektor der festen Effekte, $\alpha_1, \dots, \alpha_{n_i}$ stellt die Zufallseffekte dar und jedes $\alpha_j, j = 1, \dots, n_i$ ist jeweils unabhängig und identisch verteilt mit einer Verteilungsfunktion F_j mit Erwartungswert 0 und Varianz θ_j .

Um einen statistischen Test durchführen zu können, muss zunächst eine Hypothese aufgestellt werden.

In diesem Fall sei

$$H_0 : \theta = 0 \quad (61)$$

mit $\theta = (\theta_1, \dots, \theta_{n_i})^t$ die Nullhypothese. Unter dieser Nullhypothese sind keine Zufallseffekte vorhanden und das GLMM wird zum verallgemeinerten linearen Modell, da die Zufallseffekte a_j ($j = 1, \dots, n_i$) die Verteilung F_j besitzen mit Erwartungswert 0 und Varianz θ_j . Da bei dieser Nullhypothese diese allerdings als Null angenommen werden, gibt es natürlich keine Zufallseffekte.

Als erstes ist es nun erforderlich eine angenäherte Erweiterung für die Quasi-Likelihood-Funktion zu berechnen.

Definition 5.6 (Globale-Score-Statistik)

Um die allgemeine Form der globalen-Score-Statistik zu verstehen sei $S_\theta(\beta)$ der Vektor der ersten partiellen Ableitung der Loglikelihood-Funktion hinsichtlich dem Parameter θ ($(\frac{\partial l}{\partial \theta})$), $S_\theta(\beta)$ ist die sogenannte Score-Funktion ($S_\theta(\beta) = (\frac{\partial l(\beta, \theta)}{\partial \theta})^t$) und $H_\theta(\beta)$ sei die Hessematrix der zweiten partiellen Ableitung der Loglikelihood Funktion auch bezüglich θ . Weiter sei $I(\beta) = -H_\theta(\beta)$, H_0 sei die

⁶¹ANOVA= Analysis of Variance

Nullhypothese.

Die Maximum-Likelihood-Schätzung unter H_0 sei gegeben als $\widehat{\beta}$.

Sei jetzt $l(\beta, \theta)$ die approximierte, logarithmierte Quasi-Likelihood-Funktion. Die Globale-Score-Statistik, oder auch χ^2 -Statistik genannt, für das Testen der Hypothese $H(0)$ ist folgendermaßen definiert:

$$\chi_G^2 = (S_\theta(\widehat{\beta}))^t (\widetilde{I}(\widehat{\beta}))^{-1} S_\theta(\widehat{\beta}) \quad (62)$$

wobei G für Global steht und im individuellen Fall gleich der Anzahl der Freiheitsgrade ist.

\widetilde{I} stellt hier die sogenannte *Informationsmatrix*, oder *Fisher- Informationsmatrix* von θ unter H_0 dar.

Definition 5.7 (Informationsmatrix)

Ist $\theta \in \mathfrak{R}^p$, so heißt

$$I_{Y_n}(\theta) := \left(\text{cov}_\theta \left(\frac{\partial}{\partial \theta_i} \log f_\theta(Y_n), \frac{\partial}{\partial \theta_j} \log f_\theta(Y_n) \right) \right) \in \mathfrak{R}^p \quad (63)$$

mit $i, j = 1, \dots$ Informationsmatrix über θ basierend auf Y_n

\widetilde{I} ist in der Globalen-Score-Statistik folgendermaßen definiert:

$$\widetilde{I} := I_\theta - I_{\beta\theta}^t I_\beta^{-1} I_{\beta\theta}$$

mit

$$\begin{aligned} I_\theta &:= E \left(\frac{\partial^2 l(\beta, \theta)}{\partial \theta \partial \theta^t} \right) = E \left(\left(\frac{\partial l}{\partial \theta} \right) \left(\frac{\partial l}{\partial \theta} \right)^t \right) \\ I_{\beta\theta} &:= E \left(\frac{\partial^2 l(\beta, \theta)}{\partial \beta \partial \theta^t} \right) = E \left(\left(\frac{\partial l}{\partial \beta} \right) \left(\frac{\partial l}{\partial \theta} \right)^t \right) \\ I_\beta &:= E \left(\frac{\partial^2 l(\beta, \theta)}{\partial \beta \partial \beta^t} \right) = E \left(\left(\frac{\partial l}{\partial \beta} \right) \left(\frac{\partial l}{\partial \beta} \right)^t \right) \end{aligned}$$

wobei $l := l(\beta, \theta)$.

Es ist also:

$$\tilde{I} = E \left(\left(\frac{\partial l}{\partial \theta} \right) \left(\frac{\partial l}{\partial \theta} \right)^t \right) - \frac{\left(E \left(\left(\frac{\partial l}{\partial \beta} \right) \left(\frac{\partial l}{\partial \theta} \right)^t \right) \right)^t E \left(\left(\frac{\partial l}{\partial \beta} \right) \left(\frac{\partial l}{\partial \theta} \right)^t \right)}{\left(E \left(\left(\frac{\partial l}{\partial \beta} \right) \left(\frac{\partial l}{\partial \beta} \right)^t \right) \right)}$$

Man nennt diese Fischer-Informationsmatrix auch *erwartete* Fischer- Informationsmatrix, da jeweils der Erwartungswert betrachtet wird.

Um verständlicher zu machen, warum die Globale-Score-Statistik so aussieht, sei nun:

$$\begin{aligned} S_{\theta}(\hat{\beta}) &\sim N_G(0, \tilde{I}(\hat{\beta})) \\ \Rightarrow \tilde{I}(\hat{\beta})^{-\frac{1}{2}} S_{\theta}(\hat{\beta}) &\sim N_G(0, \text{diag}_G(1, \dots, 1)) \end{aligned}$$

Dabei ist $\tilde{I}(\hat{\beta})^{\frac{1}{2}}$ die Wurzel der Inversen und mit der Cholesky-Zerlegung gilt dann:

$$(\tilde{I}(\hat{\beta})^{-\frac{1}{2}})^t (\tilde{I}(\hat{\beta})^{-\frac{1}{2}}) = \tilde{I}(\hat{\beta})^{-1}$$

Somit ist:

$$\begin{aligned} (\tilde{I}(\hat{\beta})^{-\frac{1}{2}} S_{\theta}(\hat{\beta}))^t (\tilde{I}(\hat{\beta})^{-\frac{1}{2}} S_{\theta}(\hat{\beta})) &= S_{\theta}(\hat{\beta})^t (\tilde{I}(\hat{\beta})^{-\frac{1}{2}})^t (\tilde{I}(\hat{\beta})^{-\frac{1}{2}} S_{\theta}(\hat{\beta})) \\ &= (S_{\theta}(\hat{\beta}))^t (\tilde{I}(\hat{\beta}))^{-1} S_{\theta}(\hat{\beta}) \\ &\sim \chi_G^2 \end{aligned}$$

Wenn der Schätzer $\hat{\beta}$ gegeben ist, kann unter der Nullhypothese H_0 auch die Informationsmatrix geschätzt werden. Nach [(McCullagh und Nelder 1989)] ist dieses möglich, wenn man sich die Eigenschaften der Exponentialfamilie genauer anschaut und diese durch bestimmte, allerdings schwächere Voraussetzungen der Beobachtungen ersetzt.

Das Testen der Varianz unter dem ANOVA GLMM, wird im Rahmen dieser Arbeit nicht behandelt. Wichtig zu erwähnen ist, dass es noch mehrere Tests dieser Art gibt, in erster Linie sind hier der *Wald-Test* und der *Likelihood-Ratio-Test* zu nennen. Diese beiden Tests sind, ebenso wie der Globale-Score-Test, statistische Hypothesentests.

5.5 Modellwahl

Für die Daten der Forschungsstudie, die im folgenden Kapitel (8) mit dem Statistikprogramm SAS ausgewertet wird, ist das GLMM ein geeignetes Modell mit welchem man eine statistische Analyse durchführen kann. Jedoch ist es oft schwer unter vielen Modellen das Beste auszusuchen. Die Modellwahl hängt von vielen unterschiedlichen Faktoren, z.B. von der Stichprobengröße ab.

Es gibt verschiedene sogenannte Informationskriterien, um ein Modell auszusuchen, also die Anpassungsgüte des Modells zu testen.

Lemma 5.4 (Informationskriterien)

Folgende beiden Kriterien gehören zu den Bekanntesten, um die Anpassungsgüte eines Modells zu testen:

- *Akaike's Informationskriterium (AIC):*

$$AIC = -2l + 2p$$

wobei l der maximierte Wert der Loglikelihood-Funktion (oder der Log-Quasi,- oder Log-Pseudo-Likelihood) und p die Anzahl der Parameter ist.⁶²

- *Bayesian's Informationskriterium (BIC):*

$$BIC = -2l + p \ln n$$

l und p sind hier wie beim AIC und n ist die Anzahl aller Beobachtungen aller Subjekte, also der Umfang der Daten.⁶³

In nicht konventionellen Situationen, wie es z.B. beim GLMM der Fall ist, treten oftmals Probleme auf. Meist ist die Frage nach dem Stichprobenumfang (z.B. im Falle von korrelierten Beobachtungen) nicht richtig geklärt. Bei beiden Kriterien sollte dasjenige Modell gewählt werden, welches den kleinsten AIC-, bzw. BIC-Wert besitzt. Voraussetzung ist, dass mehrere Modelle zur Auswahl stehen. Hat man nur ein Modell vorliegen, nutzt der AIC oder BIC hingegen wenig.

⁶²Siehe: [(Akaike 1974)]

⁶³Siehe: [(Schwarz 1978)]

Für die Modellwahl haben [(Jiang et al. 2008)] als neue Methode die sogenannte „fence-Methode“ gefunden. Diese Methode beinhaltet eine breite Klasse von Problemen, welche bei der Modellwahl auftreten können, insbesondere auch bei der GLMM-Wahl. Bei dieser „fence-Methode“ werden vorsichtig nicht korrekte Modelle eliminiert.⁶⁴ Die Grundidee dieser Methode besteht darin, eine Prozedur zu verwenden, um Untergruppen eines bereits bekannten, korrekten Modells zu isolieren.

Definition 5.8 (Die Fence-Methode)

Die fence-Methode basiert auf der Prozedur, in der es um die Größe

$$Q_M := Q(y, \theta_M)$$

geht, wobei M das jeweilige Modell repräsentiert, y der $n \times 1$ Vektor der Beobachtungen und θ_M der Parameter unter dem Modell M ist. Das angenommene Modell M ist das wahre Modell, wenn der Erwartungswert $E(Q_m)$ minimiert ist. Mit dem wahren Modell ist hier zwar ein korrektes Modell, jedoch nicht zwingend das effizienteste Modell gemeint.

Im Folgenden wird ein Beispiel für die fence-Methode von [(Jiang et al. 2008)] gegeben, also ein Beispiel für die Größe Q_M .

Beispiel 5.2 (Maximum-Likelihood Modellwahl)

Ein Beispiel für Q_M ist die negative Loglikelihood-Funktion unter dem Modell M . Es ist

$$Q_M = -\log(f_M(y|\theta_M))$$

Hier ist $f_M(\cdot|\theta_M)$ die Dichtefunktion von y hinsichtlich eines Maßes ν , welches den wahren Parametervektor θ festlegt. Um zu zeigen, dass M das „wahre“ Modell ist, dass also $E(Q_M)$ minimal ist und θ_M den zugehörigen wahren Parameter darstellt, sei $f(y)$ die wahre Dichtefunktion von y . Dann gilt:

$$\begin{aligned} Q_M &= -\log(f_M(y|\theta_M)) \\ \Leftrightarrow -Q_M &= \log(f_M(y|\theta_M)) \end{aligned}$$

Es muss nun gezeigt werden, dass $-E(Q_M)$ minimiert ist.

⁶⁴siehe:[(Jiang et al. 2008)]

Es ist:

$$\begin{aligned}
 -E(Q_M) &= E(\log \{f_M(y|\theta_M)\}) \\
 &= \int \log \{f_M(y|\theta_M)\} f(y)\nu(dy) \\
 &= \int \log \left\{ \frac{f(y)f_M(y|\theta_M)}{f(y)} \right\} f(y)\nu(dy) \\
 &= \int \log \left\{ f(y) \frac{f_M(y|\theta_M)}{f(y)} \right\} f(y)\nu(dy) \\
 &= \int \left(\log \{f(y)\} + \log \left\{ \frac{f_M(y|\theta_M)}{f(y)} \right\} \right) f(y)\nu(dy) \\
 &= \int \log \{f(y)\} f(y)\nu(dy) + \int \log \left\{ \frac{f_M(y|\theta_M)}{f(y)} \right\} f(y)\nu(dy) \\
 &\leq \int \log \{f(y)\} f(y)\nu(dy) + \log \left\{ \int \frac{f_M(y|\theta_M)}{f(y)} f(y)\nu(dy) \right\} \\
 &= \int \log \{f(y)\} f(y)\nu(dy) + \log \left\{ \int f_M(y|\theta_M)\nu(dy) \right\} \\
 &= \int \log \{f(y)\} f(y)\nu(dy) + \log \{1\} \\
 &\stackrel{\log\{1\}=0}{=} \int \log \{f(y)\} f(y)\nu(dy)
 \end{aligned}$$

Die Ungleichung gilt, da hier eine konkave Funktion⁶⁵ gegeben ist und somit keine Gleichheit mehr vorliegt, da die Jensische Ungleichung⁶⁶ angewendet werden kann und somit der Term in der Gleichung größer oder gleich der vorherigen Gleichung ist. Gleichheit besteht genau dann, wenn M das wahre Modell und θ_M der zugehörige wahre Parametervektor ist.

Des Weiteren ist $\log \left\{ \int f_M(y|\theta_M)\nu(dy) \right\} = \log \{1\}$ aufgrund der Normierungseigenschaft⁶⁷ der Dichtefunktion. Da $\log \{1\}$ gleich 0 ist, fällt somit der rechte Term in der vorletzten Gleichung weg.

Andere Beispiele der fence-Methode und den entsprechenden Algorithmus sind in [(Jiang 2007)] oder in [(Jiang et al. 2008)] zu finden.

⁶⁵Im Allgemeinen ist eine konkave Funktion folgendermaßen definiert: Die Funktion f ist konkav auf dem Intervall I , wenn für je zwei verschiedene Punkte $x_1, x_2 \in I$ und für alle $\lambda \in (0, 1)$ folgende Ungleichung gilt: $(1 - \lambda)f(x_1) + \lambda f(x_2) \leq f((1 - \lambda)x_1 + \lambda x_2)$

⁶⁶Da für eine konkave Funktion f die Funktion $-f$ konvex ist, gilt die Jensische Ungleichung hier in umgekehrter Richtung

⁶⁷siehe: [(Fahrmeier 2003)] S.271

6 Schätzmethoden für das GLMM: EBP und GEE

Im zweiten Kapitel des GLMMs werden die Schätzmethoden für das GLMM präzisiert. Dabei werden Schätzmethoden, wie der EBP „Empirical Best Predictor“ für das Schätzen der Zufallseffekte und die GEE „Generalized Estimating Equation“, als eine Schätzmethode für die festen Effekte dargestellt. Zunächst geht es um das Schätzen der Zufallseffekte.

6.1 Schätzungen der Zufallseffekte, (EBP)

Bemerkung 6.1

Oft ist es von Interesse die Zufallseffekte $\alpha_i, i = 1, \dots, n$ zu schätzen. Dieses kann jedoch, ebenso wie die Schätzung, bzw. die Vorhersage von möglichen gemischten Effekten, in einem GLMM einige Probleme bereiten.

Nach [(Jones 1993)] ist der EM-Algorithmus von [(Liard und Ware 1982)] für das Schätzen der Parameter in einem GLMM eine geeignete Methode. Vor allem für den Fall, dass die Daten nicht vollständig sind, sie also „missing values“ enthalten,⁶⁸ ist der EM-Algorithmus laut Jones eine geeignete Methode Vorhersagen für die Parameter zu treffen.

Das Problem wird von [(Jiang 2007)] anders angegangen. Hier werden zwei unterschiedliche Wege dargestellt, wie man Vorhersagen für die Zufallseffekte treffen kann.

Zum Einen gibt es eine Methode, in der man die festen Effekte und die im Mittelpunkt stehenden Zufallseffekte gemeinsam schätzt. Diese Methode ist eine Erweiterung der BLUE.⁶⁹

Die zweite Methode, die in Jangs Buch dargestellt wird, wurde erst vor einigen Jahren von Jiang selbst entwickelt und ist eine Methode der „empirisch besten Vorhersage“(EBP) im Hintergrund der Schätzung von „small areas“, bzw. kleinen Stichproben. Da die empirisch beste Vorhersage natürlich von großer

⁶⁸siehe Kapitel (7.5)

⁶⁹BLUE = „Best linear unbiased estimation“. Im einfachsten Fall der linearen Regression ist die Kleinste-Quadrat-Summen-Schätzung $\hat{\beta} = (X^t X)^{-1} X^t Y$ die BLUE

Bedeutung ist, wird diese im Folgenden für das longitudinale GLMM dargelegt.

Bei dieser Art von GLMM werden die Beobachtungen in verschiedene Gruppen, bzw. Cluster eingeteilt.

Bemerkung 6.2 (Methoden des EBPs)

Für das longitudinale GLMM gibt es zwei verschiedene Wege, wie bei der EMP vorgegangen werden kann:

1. *Die kleine Bereichsschätzung („small area estimation“)*
2. *Die verallgemeinerten Schätzgleichungen; Sie werden häufig bei der Analyse von longitudinalen Daten⁷⁰ verwendet, um die festen Effekte, bzw. den festen Effekt β zu schätzen.*

Es wird zunächst der erste Punkt vorgestellt. Dabei ist zu beachten, dass es bei der „small area estimation“ meistens zu Problemen kommt, die mit der Vorhersage von gemischten Effekten in Beziehung steht. Bei der Analyse von longitudinalen Daten hingegen besteht großes Interesse den Mittelwert der Beobachtungen zu schätzen, was mit den festen Effekten im GLMM zu tun hat.

6.2 „Small area estimation“ unter dem GLMM

Nach [(Jiang 2007)] gibt es zwei verschiedene Ansätze der „small area estimation“ um den EBP zu erhalten. Zum Einen gibt es die Schätzung des EBP unter dem GLMM und zum Anderen mit „design consistency“.⁷¹ Die zweite Modell-basierte Methode wird hier nicht weiter dargelegt.

Als ein Ansatz für die empirisch beste Vorhersage (EBP) wird nun die *small area estimation* unter dem GLMM vorgestellt.

Es wird dazu angenommen, dass die Beobachtungen $y_{i1}, \dots, y_{in_i}, i = 1, \dots, n, j = 1, \dots, n_i$ als stochastisch unabhängig gegeben sind mit einer bedingten Dichtefunktion bzgl. α_i , wobei $\alpha_i = (\alpha_{ij})_{1 \leq j \leq n_i}$ wieder der Vektor der Zufallseffekte ist. Die bedingte Dichtefunktion,⁷² gemäß der parametrischen Exponentialfamilie, ist gegeben durch:

$$f_{\alpha_i}(y_{ij}) = \exp \left(\left(\frac{w_{ij}}{\phi} \right) [y_{ij}\theta_{ij} - b(\theta_{ij})] + c \left(y_{ij}, \frac{w_{ij}}{\phi} \right) \right)$$

⁷⁰siehe auch Kapitel 7

⁷¹siehe: [(Jiang 2007)]S.142

⁷²siehe Grundlagen

Hier sind $b(\cdot)$ und $c(\cdot, \cdot)$ wie in Definition (4.7), ϕ ist wieder der Dispersionsparameter und w_{ij} sind die Gewichte. Für ungruppierte Daten gilt $w_{ij} = 1$, d.h. die Antworten y_{i1}, \dots, y_{in_i} haben in diesem Fall die Dichte:

$$f_{\alpha_i}(y_{ij}) = \exp(\phi^{-1}[y_{ij}\theta_{ij} - b(\theta_{ij})] + c(y_{ij}, \phi)) \quad (64)$$

Des Weiteren soll θ_{ij} in Verbindung mit dem linearen Prädiktor

$$\eta_{ij} = x_{ij}^t \beta + z_{ij}^t \alpha_i$$

stehen gegeben durch die Linkfunktion $g(\cdot)$, derart dass: $g(\theta_{ij}) = \eta_{ij}$, bzw. mit der inversen Linkfunktion $h(\cdot)$: $\theta_{ij} = h(\eta_{ij})$.

Außerdem gibt es Parameter $v = v_1, \dots, v_m$, welche unabhängig sind und eine Dichtefunktion $f_\xi(\cdot)$ besitzen, wobei ξ ein Vektor von Varianzkomponenten ist. Sei $\psi = (\beta^t, \xi^t)$ und $\vartheta = (\psi^t, \phi)$. Sofern die v_1, \dots, v_m eine Binominal-, oder Poissonverteilung besitzen, bei der der Dispersionsparameter ϕ bekannt ist (bei beiden Verteilungen ist $\phi = 1$), stellt ψ den Vektor aller unbekannt Parameter dar.

Im Folgenden wird für einen gemischten Effekt die beste Vorhersage, bzw. beste Schätzung („best prediktor (BP)“) dargelegt. Der gemischte Effekt ist gegeben durch:

$$\zeta = \zeta(\beta, \alpha_S) \quad (65)$$

wobei S eine Teilmenge von $\{1, \dots, n\}$ ist und $\alpha_S = (\alpha_i)_{i \in S}$. Sei $y_S = (y_i)_{i \in S}$ und $y_i = (y_{ij})_{1 \leq j \leq n_i}$.

Bevor der beste Schätzer (BP) für diesen gemischten Effekt unter dem oben dargestellten Modell gefunden wird, wird zuvor noch die mittlere quadratische Abweichung definiert.

Definition 6.1 (Mittlere quadratische Abweichung)

Im Fall von ausgeglichenen Daten ist die „mittlere quadratische Abweichung“ (MSE), oder auch empirische Varianz genannt, definiert durch:

$$\begin{aligned} MSE &= \frac{SSE}{m(k-1)} \\ &= \frac{\sum_{i=1}^m \sum_{j=1}^k (y_{ij} - \bar{y}_i.)^2}{m(k-1)} \end{aligned}$$

dabei ist: $SSE = \text{Sum of Squares Error}$

Man kann die mittlere quadratische Abweichung auch folgendermaßen darstellen:

$$MSE = \text{Bias(Verzerrung)} + \text{Varianz}$$

Für den erwartungstreuen⁷³ Fall entspricht die mittlere quadratische Abweichung der Varianz, da die Verzerrung dann Null ist.

Satz 6.1 (BP für gemischten Effekt ζ)

Der BP unter dem Modell für den gemischten Effekt ζ im Sinne des Minimums der mittleren quadratischen Abweichung ist gegeben durch:

$$\tilde{\zeta} = \frac{\int \zeta(\beta, \alpha_S) \exp(\phi^{-1} \sum_{i \in S} s_i(\beta, \alpha_i)) \prod_{i \in S} f_\xi(\alpha_i) \prod_{i \in S} d\alpha_i}{\int \prod_{i \in S} \exp(\phi^{-1} s_i(\beta, v_i)) f_\xi(v_i) dv_i} \quad (66)$$

Beweis. Es ist:

$$\tilde{\zeta} = E(\zeta|y) \quad (67)$$

$$= E(\zeta(\beta, \alpha_S)|y_S) \quad (68)$$

$$= \frac{\int \zeta(\beta, \alpha_S) f_{\alpha_S}(y_S) f_\xi(\alpha_S) d\alpha_S}{\int f_{\alpha_S}(y_S) f_\xi(\alpha_S) d\alpha_S} \quad (69)$$

$$\stackrel{(64)}{=} \frac{\int \zeta(\beta, \alpha_S) \exp(\phi^{-1} \sum_{i \in S} s_i(\beta, \alpha_i)) \prod_{i \in S} f_\xi(\alpha_i) \prod_{i \in S} d\alpha_i}{\int \prod_{i \in S} \exp(\phi^{-1} s_i(\beta, v_i)) f_\xi(v_i) dv_i} \quad (70)$$

Im letzten Schritt wird für $f_{\alpha_S}(y_S)$ die Dichtefunktion nach Formel (64) eingesetzt. Zur Veranschaulichung wurde eine neue Funktion $s_i(\beta, v)$ derart definiert, dass:

$$s_i(\beta, v_i) := \sum_{j=1}^{n_i} w_{ij} (y_{ij} h(x_{ij}^t \beta + z_{ij}^t v_i) + b \{h(x_{ij}^t \beta + z_{ij}^t v_i)\}) \quad (71)$$

ist, wobei $h(x_{ij}^t \beta + z_{ij}^t v_i) = \xi_{ij}$ der lineare Prädiktor und w_{ij} wieder die Gewichte sind.

Wichtig ist sich über die Dimension der Integrale in Gleichung (70) klar zu werden.

⁷³siehe Definition(3.17)

Für den Zähler gilt:

$$\dim\left(\int \zeta(\beta, \alpha_S) \exp\left(\phi^{-1} \sum_{i \in S} s_i(\beta, \alpha_i)\right) \prod_{i \in S} f_\xi(\alpha_i) \prod_{i \in S} d\alpha_i\right) = rs = \dim(\alpha_i) |S|$$

wobei $s := |S|$ die Anzahl von der Menge $S \subseteq \{1, \dots, n\}$ angibt. Folglich hat das Integral des Zählers in (70) die Dimension sr . Für den Nenner gilt:

$$\dim\left(\int \prod_{i \in S} \exp(\phi^{-1} s_i(\beta, v_i)) f_\xi(v_i) dv\right) = r = \dim(\alpha_i)$$

Somit hat das Integral des Nenners die Dimension $r = \dim(\alpha_i)$. □

Der BP $\tilde{\zeta}$ hängt nun von y_S und ψ ab. Es sei die Funktion $u(\cdot, \cdot)$ gegeben, derart dass:

$$\tilde{\zeta} := u(y_S, \psi) \tag{72}$$

Im Normalfall ist $\psi = (\beta^t, \xi^t)$ unbekannt. Dann wird ψ durch den konsistenten⁷⁴ Schätzer $\hat{\psi}$ ersetzt, so dass:

$$\hat{\zeta} = u(y_S, \hat{\psi})$$

den empirisch besten Schätzer (EBP) darstellt.

Das Interesse des Betrachters liegt nicht immer darin den EBP zu schätzen, sondern auch seine Streuung ist von Bedeutung. Für die Streuung wird das passende Maß benötigt. Ein solches Maß ist gegeben durch den MSE aus Definition (6.1). Dieses Maß ist definiert als:

$$MSE(\hat{\zeta}) = E(\hat{\zeta} - \zeta)^2 \tag{73}$$

Leider ist (73) oftmals schwierig zu schätzen. In solch einer Situation hilft die MSE von $\tilde{\zeta}$ weiter. Es ist:

$$MSE(\tilde{\zeta}) = b(\psi) \tag{74}$$

Mit Gleichung (74) ist der „naive“ Schätzer von MSE von $\hat{\zeta}$, der gesucht wird, gleich $b(\hat{\psi})$. Es gilt also: $MSE(\hat{\zeta}) = b(\hat{\psi})$

Wichtig zu beachten ist, dass dies natürlich nicht der wahre MSE Schätzer ist,

⁷⁴siehe Definition (3.18)

denn dieser hat die folgende Form:

$$MSE(\widehat{\zeta}) = MSE(\widetilde{\zeta}) + E(\widehat{\zeta} - \widetilde{\zeta})^2 \quad (75)$$

$$= b(\psi) + E(\widehat{\zeta} - \widetilde{\zeta})^2 \quad (76)$$

Es ist zu erkennen, dass der naive Schätzer von $MSE(\widehat{\zeta}) = b(\widehat{\psi})$ den zweiten Term der Gleichung vollkommen ignoriert und es durch den naiven Schätzer $b(\widehat{\psi})$ zu einer Verzerrung kommt.⁷⁵

Um diese Verzerrung zu korrigieren, muss man sich eine weitere Gleichung für den MSE anschauen. Zur vereinfachten Darstellung, sei ab jetzt ϕ bekannt (es wird angenommen, dass nun eine Binominal-, oder eine Poissonverteilung vorliegt), damit gilt: $b(\psi) = b(\xi)$ aus der vorherigen Gleichung und $b(\xi)$ ist von der Form:

$$\begin{aligned} b(\xi) &\stackrel{(74)}{=} MSE(\widetilde{\zeta}) \\ &= E(\zeta^2) - (E(\widetilde{\zeta}))^2 \\ &\stackrel{(65),(72)}{=} E(\zeta(\beta, \alpha_S)^2) - (E(u(Y_S, \xi)))^2 \end{aligned}$$

wobei S eine Teilmenge von $\{1, \dots, m\}$ ist. Nun kann der MSE-Schätzer durch folgende Gleichung dargestellt werden:

$$\widehat{MSE}(\widehat{\zeta}) = b(\widehat{\xi}) + m^{-1} \left\{ e(\widehat{\xi}) - B(\widehat{\xi}) \right\} \quad (77)$$

Die beiden Funktionen $e(\cdot)$ und $B(\cdot)$ werden später noch genau definiert. Zuvor wird nun allerdings Gleichung (76) noch einmal betrachtet. Wie $b(\psi)$ bzw. hier ja $b(\xi)$ aussieht wurde gerade dargestellt. Es stellt sich nun noch die Frage, wie der zweite Term der Gleichung, also $E(\widehat{\zeta} - \widetilde{\zeta})^2$ aussieht:

Man nimmt an, dass die Taylorreihen-Entwicklung verwendet wird, um $\widehat{\zeta} - \widetilde{\zeta}$ anzunähern.

Es wird angenommen, dass der Betrag $\left| \widehat{\xi} - \xi \right| = O(m^{-\frac{1}{2}})$ ist.

Definition 6.2 (Definition von O und o)

Es gilt:

1. Für den Fall gleicher Ordnung: Falls es ein $c > 0$ gibt mit $|f(x)| \leq |g(x)|$

⁷⁵siehe: [(Jiang 2007)]S.144

für $x \rightarrow x_0$, dann ist $f = O(g)$, d.h. f geht schneller gegen 0 als g für $x \rightarrow x_0$

2. Für den Fall höherer Ordnung: Falls gilt: $\frac{|f(x)|}{|g(x)|} \rightarrow 0$ für $x \rightarrow x_0$, dann ist $f = o(g)$

Mit der Taylor-Entwicklung ergibt sich:

$$\begin{aligned}\widehat{\zeta} - \widetilde{\zeta} &= u(Y_S, \widehat{\xi}) - u(Y_S, \xi) \\ &= \left(\frac{\partial u}{\partial \xi} \right)^t (\widehat{\xi} - \xi) + o(m^{-\frac{1}{2}})\end{aligned}$$

und damit folgt nach [(Jiang 2007)] S.156:

$$E(\widehat{\zeta} - \widetilde{\zeta})^2 = m^{-1} E \left\{ \left(\frac{\partial u}{\partial \xi} \right)^t \sqrt{m}(\widehat{\xi} - \xi) \right\}^2 + o(m^{-1})$$

In Gleichung (77) ist noch nicht bekannt, was $e(\cdot)$ und $B(\cdot)$ für Funktionen sind. Zur Erläuterung:

Im Fall, dass $\widehat{\xi}$ nun nicht mehr auf y_S basiert, sondern auf $y_{-S} = (y_i)_{i \notin S}$, ist $\widehat{\xi}$ unabhängig von der Zufallsvariable Y_S ist. Dann hat die Funktion $e(\cdot) = e_{-S}(\cdot)$ folgende Darstellung:

$$e(\xi) := e_{-S}(\xi) := E \left\{ \left(\frac{\partial u}{\partial \xi} \right)^t \sqrt{m}(\widehat{\xi} - \xi) \right\}^2$$

wobei $\widehat{\xi} := \widehat{\xi}_{-S}$ ist und $B(\cdot)$ ist gegeben durch:

$$B(\xi) := \left(\frac{\partial b}{\partial \xi} \right)^t m E(\widehat{\xi} - \xi) + \frac{1}{2} E \left\{ (\sqrt{m}(\widehat{\xi} - \xi))^t \left(\frac{\partial b}{\partial \xi} \right) \left(\frac{\partial b}{\partial \xi} \right)^t (\sqrt{m}(\widehat{\xi} - \xi)) \right\}$$

Mit diesen beiden Funktionen ist nun Gleichung (77) definiert. Weiter gilt für die zweite Ordnung:

$$E(\widehat{MSE}(\widehat{\zeta}) - MSE(\widehat{\zeta})) = o(m^{-1}) \quad (78)$$

Diese Gleichung bedeutet (nach (6.2)), dass

$$\frac{|E(\widehat{MSE}(\hat{\zeta}) - MSE(\hat{\zeta}))|}{|m^{-1}|} \rightarrow 0$$

d.h. $E(\widehat{MSE}(\hat{\zeta}) - MSE(\hat{\zeta}))$ geht schneller gegen 0 als m^{-1} .

Soweit zur Schätzung der mittleren quadratischen Abweichung (MSE) des EBP unter dem GLMM für die „small area estimation“.

Bemerkung 6.3

Die Methode den EBP zu schätzen ist eine Modell-basierte Methode. Dabei ist zu beachten, dass der Schätzer nicht der EBP ist, sofern das angenommene Modell nicht das Richtige ist.

6.3 GEE als Schätzmethode für die festen Effekte

Bemerkung 6.4

Der zweite Ansatz für die empirisch beste Vorhersage (BP) sind im Falle des longitudinalen GLMMs die verallgemeinerten Schätzgleichungen oder auch die GEEs (GEE = Generalized Estimating Equations).

Diese wurden von [(Liang und Zeger 1986)] eingeführt.

Es kann also neben der EBP auch von Interesse sein, sogenannte Schätzgleichungen-, bzw. Funktionen zu verwenden um Parameter zu schätzen. In der allgemeinen Theorie der GEE dienen die verallgemeinerten Schätzgleichungen dazu bei, den festen Parameter β zu schätzen. Die folgende Definition bezieht sich zunächst auf eine einfache Schätzgleichung.

Definition 6.3 (Schätzfunktion für Schätzgleichungen)

Eine Schätzfunktion für eine Schätzgleichung ist eine Funktion, in den meisten Fällen eine vektorwertige Funktion, welche von dem Vektor der Beobachtungen $y = (y_i)_{1 \leq i \leq n}$ und dem Parametervektor θ abhängt (Wenn beim Erwartungswert sowie bei der Kovarianzmatrix der Indize θ fehlt, entspricht es dem wahren Parameter θ .) und außerdem die Bedingung:

$$E_{\theta} \{g_{\theta}(y)\} = 0 \tag{79}$$

für jedes θ erfüllt, wobei $g_{\theta}(y)$ die Schätzfunktion darstellt.

Um im Folgenden auf die sogenannte *Schätzgleichung* einzugehen, sei nun der einfache Fall gegeben, dass y_1, \dots, y_n alle unabhängig sind mit $E(y_i) = \theta$. Es gibt eine Klasse \mathcal{G} , die alle Schätzfunktionen der Form:

$$g_\theta(y) = \sum_{i=1}^n a_i(\theta)(y_i - \theta)$$

enthält, wobei die Funktionen $a_i(\theta)$ differenzierbare Funktionen sind mit $\sum_{i=1}^n a_i(\theta) \neq 0$. Dann gilt folgende Definition:

Definition 6.4 (Schätzgleichung)

Wenn für $1 \leq i \leq n$ die Varianz $\text{Var}(y_i) = \sigma^2$ und $g^* = \sum_{i=1}^n (y_i - \theta)$ eine „optimale“ Schätzfunktion innerhalb der Klasse \mathcal{G} ist, sowie die Gleichung $g^* = 0$ festlegt, dass das arithmetische Mittel \bar{y} ein Schätzer für θ ist, dann wird die Gleichung

$$g_\theta(y) = 0 \tag{80}$$

Schätzgleichung genannt.

In der Definition einer Schätzgleichung ist es Voraussetzung, dass eine optimale Schätzfunktion gegeben ist. Die Definition (6.5) gibt Aufschluss über diese optimale Schätzfunktion.

Definition 6.5 (Optimale Schätzfunktion)

Eine Schätzfunktion ist eine optimale Schätzfunktion, sofern die beiden folgenden Kriterien erfüllt sind:⁷⁶

1. Für Gleichung (80) ist die zugehörige Schätzfunktion näherungsweise Null, sofern θ der wahre Parameter ist. Für Gleichung (79) bedeutet das, dass man die Varianz von g (also $\text{var}(g)$ mit $g = g_\theta(y)$) minimieren muss.
2. Der wahre Parameter θ wird von einem falschen unterschieden, indem die partielle Ableitung der Schätzfunktion g nach dem Vektor θ , also $\frac{\partial g}{\partial \theta}$ maximiert wird.

Das folgende Lemma wird ohne Beweis hingenommen.⁷⁷

⁷⁶siehe: [(Jiang 2007)] S.183

⁷⁷siehe: [(Jiang 2007)] S.183

Lemma 6.1

Wenn die beiden Kriterien aus Definition (6.5) für die optimale Schätzfunktion miteinander kombiniert werden, ergibt sich Folgendes:

$$\frac{\text{var}(g)}{\left\{E\left(\frac{\partial g}{\partial \theta}\right)\right\}^2} = \text{var}(g_s) \tag{81}$$

wobei $g_s = \frac{g}{E\left(\frac{\partial g}{\partial \theta}\right)}$ die standardisierte Funktion von g ist.

Im Folgenden wird ein Satz definiert, in dem es um die „optimale Schätzgleichung“ geht. Für diesen Satz werden einige Voraussetzungen benötigt: Es liegt eine multivariate Version der Schätzfunktion vor, d.h. $y = (y_1, \dots, y_n)$ ist ein Vektor von Beobachtungen und $x = (x_1, \dots, x_n)$ ist der Vektor der erklärenden Variablen. $\mu = E_\theta(y|x) = \mu(x, \theta)$ ist der bedingte Erwartungswert von y und x bzgl. dem unbekanntem Parametervektor θ und $V := \text{Var}(y|x)$ die bedingte Varianz von y bedingt x .

$\left(\frac{\partial \mu}{\partial \theta}\right)^t$ ist die Matrix der partiellen Ableitungen, $\mathcal{H} = \{G = A(y - \mu)\}$ mit $A = A(x, \theta)$ einer Klasse von vektorwertigen Schätzfunktionen, derart dass $E\left(\left(\frac{\partial G}{\partial \theta}\right)^t\right)$ nicht singulär⁷⁸ ist. Dann gilt der folgende Satz:

Satz 6.2

Sei $V := \text{Var}(y|x)$ bekannt, $\mu := E_\theta(y|x) = \mu(x, \theta)$ und $E\left(\left(\frac{\partial \mu}{\partial \theta}\right) V \left(\frac{\partial \mu}{\partial \theta}\right)^t\right)$ nicht singulär. Dann ist die optimale Schätzfunktion innerhalb von \mathcal{H} gegeben durch:

$$G^* = \left(\frac{\partial \mu}{\partial \theta}\right)^t V^{-1}(y - \mu)$$

mit $A = A^* = \left(\frac{\partial \mu}{\partial \theta}\right)^t V^{-1}$

Für den Beweis des Satzes wird der folgende Satz benötigt:⁷⁹

Satz 6.3

Die Funktion $G^* \in \mathcal{H}$ ist eine optimale Schätzfunktion innerhalb von \mathcal{H} , sofern gilt:

1.

$$E(G^{*(s)}(G^{(s)t})) = E(G^{(s)}(G^{*(s)t})) = E(G^{(s)}(G^{(s)t})) \tag{82}$$

⁷⁸Nicht singulär bedeutet, dass die Determinante der Matrix ungleich Null ist.

⁷⁹siehe: [(Heyde 1997)] S.14

wobei $G^{(s)}$ die standardisierte Schätzfunktion
 $G^{(s)} := -(E\dot{G})^t(EGG^t)^{-1}G$ mit $E\dot{G} := \frac{E\partial G}{\partial \theta}$ darstellt.

2. oder äquivalent dazu: wenn

$$\left(E \left(\dot{G}\right)\right)^{-1} E(G(G^*)^t)$$

eine konstante Matrix für alle $G \in \mathcal{H}$ ist.

3. Umgekehrt gilt Gleichung (82), falls $G^* \in \mathcal{H}$ eine optimale Schätzfunktion und \mathcal{H} konvex ist.⁸⁰

Für den Beweis des Satzes (6.3) ist zunächst noch eine Definition notwendig:⁸¹

Definition 6.6

$G^* \in \mathcal{H}$ ist eine optimale Schätzfunktion innerhalb von \mathcal{H} , falls

$$\mathcal{E}(G^*) - \mathcal{E}(G)$$

nicht negativ definit ist für alle $G \in \mathcal{H}, \theta \in \Theta$. Dabei ist

$$\mathcal{E}(G) = E(G^{(s)}G^{(s)t}) = (E\dot{G})^t(EGG^t)^{-1}(E\dot{G})$$

Beweis. von Satz (6.3): Zunächst ist zu zeigen, dass G^* eine optimale Schätzfunktion ist, falls (82) gilt.

Um zu zeigen, dass G^* eine optimale Schätzfunktion ist, sei nun: (82) gegeben. Nach Definition (6.6) muss nun gezeigt werden, dass $E(G^{*(s)}G^{*(s)t}) - E(G^{(s)}G^{(s)t})$ nicht negativ definit ist. Das ist allerdings leicht zu sehen, da:

$$E(G^{*(s)}G^{*(s)t}) - E(G^{(s)}G^{(s)t}) = E(G^{*(s)} - G^{(s)})(G^{*(s)} - G^{(s)})^t$$

gilt und die rechte Seite dieser Gleichung eine Kovarianzfunktion ist, welche nicht negativ definit ist, da die Kovarianz durch den Erwartungswert definiert ist und dieser immer nicht negativ definit ist. Daraus folgt, dass auch die linke Seite der Gleichung, also: $E(G^{*(s)}G^{*(s)t}) - E(G^{(s)}G^{(s)t})$ nicht negativ definit ist und nach Definition (6.6) ist G^* somit unter der Bedingung, dass (82) gilt,

⁸⁰Eine Menge $M \in \mathbb{R}^n$ ist konvex, falls $\forall x, y \in M$ und $\forall \delta \in (0, 1) : \delta x + (1 - \delta)y \in M$

⁸¹siehe: [(Heyde 1997)]S.12

eine optimale Schätzfunktion.

Nun sei \mathcal{H} konvex und G^* eine optimale Schätzfunktion. Es muss gezeigt werden, dass (82) gilt:

Nach Definition (6.6) gilt, dass

$$\mathcal{E}(G^*) - \mathcal{E}(G) = E(G^{*(s)}G^{*(s)t}) - E(G^{(s)}G^{(s)t})$$

nicht negativ definit ist. Da nach Definition (6.6) die nicht negative Definitheit für alle $G \in \mathcal{H}$ gilt und \mathcal{H} hier konvex ist, gilt die Definition auch für $H = \alpha G + G^*$, da aufgrund der Konvexität von \mathcal{H} gilt: $G, G^* \in \mathcal{H} \Rightarrow \alpha G + G^* \in \mathcal{H}$ mit $\alpha \in (0, 1)$, es ist folglich $\alpha \neq 0$. Es ist somit:

$$E(G^{*(s)}G^{*(s)t}) - E(H^{(s)}H^{(s)t})$$

nicht negativ definit. Weiter gilt, dass

$$\begin{aligned} & E(G^{*(s)}G^{*(s)t}) - E(H^{(s)}H^{(s)t}) \\ = & E(G^{*(s)}G^{*(s)t}) - E((\alpha G^{(s)} + G^{*(s)})(\alpha G^{(s)} + G^{*(s)})^t) \\ = & E(G^{*(s)}G^{*(s)t}) - E(\alpha^2 G^{(s)}G^{(s)t} + \alpha G^{(s)}G^{*(s)t} + \alpha G^{*(s)}G^{(s)t} + G^{*(s)}G^{*(s)t}) \\ = & E(G^{*(s)}G^{*(s)t}) - E(\alpha^2 G^{(s)}G^{(s)t}) - E(\alpha G^{(s)}G^{*(s)t}) - E(\alpha G^{*(s)}G^{(s)t}) \\ & - E(G^{*(s)}G^{*(s)t}) \\ = & -E(\alpha^2 G^{(s)}G^{(s)t}) - E(\alpha G^{(s)}G^{*(s)t}) - E(\alpha G^{*(s)}G^{(s)t}) \end{aligned}$$

nicht negativ definit ist.

Wird die letzte Zeile der Gleichung mit -1 multipliziert, ergibt sich, dass:

$$\begin{aligned}
 & E(\alpha^2 G^{(s)} G^{(s)t}) + E(\alpha G^{(s)} G^{*(s)t}) + E(\alpha G^{*(s)} G^{(s)t}) \\
 = & \alpha^2 E(G^{(s)} G^{(s)t}) - \alpha(-E(G^{(s)} G^{*(s)t})) - \alpha(-E(G^{*(s)} G^{(s)t})) \\
 = & \alpha^2 \left\{ E \left[\left(-(E\dot{G})^t (EGG^t)^{-1} G \right) \left(-(E\dot{G})^t (EGG^t)^{-1} G \right)^t \right] \right\} \\
 - & \alpha \left\{ -E \left[\left(-(E\dot{G})^t (EGG^t)^{-1} G \right) \left(-(E\dot{G}^*)^t (EG^* G^{*t})^{-1} G^* \right)^t \right] \right\} \\
 - & \alpha \left\{ -E \left[\left(-(E\dot{G}^*)^t (EG^* G^{*t})^{-1} G^* \right) \left(-(E\dot{G})^t (EGG^t)^{-1} G \right)^t \right] \right\} \\
 = & \alpha^2 \left(EGG^t - (E\dot{G})(E\dot{G}^*)^{-1} EG^* G^{*t} ((E\dot{G}^*)^{-1})^t (E\dot{G})^t \right) \\
 - & \alpha \left(-EGG^{*t} + (E\dot{G})(E\dot{G}^*)^{-1} (EG^* G^{*t}) \right) \\
 - & \alpha \left(-EG^* G^t + (EG^* G^{*t})((E\dot{G}^*)^{-1})^t (E\dot{G})^t \right)
 \end{aligned}$$

nicht negativ definit ist. Wie schnell zu erkennen ist, ist dies von der Form: $\alpha^2 A - \alpha B$, mit $A = EGG^t - (E\dot{G})(E\dot{G}^*)^{-1} EG^* G^{*t} ((E\dot{G}^*)^{-1})^t (E\dot{G})^t$ und

$$\begin{aligned}
 B = & \left(-EGG^{*t} + (E\dot{G})(E\dot{G}^*)^{-1} (EG^* G^{*t}) \right) \\
 & + \left(-EG^* G^t + (EG^* G^{*t})((E\dot{G}^*)^{-1})^t (E\dot{G})^t \right)
 \end{aligned}$$

wobei A und B symmetrische Matrizen darstellen und A wieder nach Definition (6.6) nicht negativ definit ist. Da A nicht negativ definit ist, gilt mit einem beliebigen p -dimensionalen Vektor $u \neq 0$, dass $u^t A u \geq 0$ ist. Aufgrund der nicht negativen Definitheit von A folgt somit:

$$\begin{aligned}
 \alpha^2 u^t A u - \alpha u^t B u & \geq 0 \\
 \Rightarrow \alpha^2 u^t A u & \geq \alpha u^t B u \\
 \Rightarrow u^t A u & \geq \frac{\alpha}{\alpha^2} u^t B u \\
 \Rightarrow u^t A u & \geq \alpha^{-1} u^t B u = 0
 \end{aligned}$$

für alle α . Da $\alpha \neq 0$ folgt sofort, dass $u^t B u = 0$ sein muss und somit ist $B = 0$, da der Vektor nach Voraussetzung $u \neq 0$ ist.

$B = 0$ kann nun geschrieben werden, als:

$$(EGG^t)((E\dot{G})^t)^{-1} C + C(E\dot{G})^{-1}(EGG^t) = 0 \quad (83)$$

wobei

$$C = (EG^{(s)}EG^{(s)t} - EG^{(s)}EG^{*(s)t})(E\dot{G}^*)^{-1}EG^*G^{*t} \quad (84)$$

Das gilt aufgrund der standardisierten Schätzfunktion $G^{(s)}$. Wenn man diese in Gleichung (84) jeweils einsetzt und anschließend das „neue“ C in Gleichung (83) einsetzt, erhält man das B von eben. Da (83) für alle $G \in \mathcal{H}$ gilt, kann G mit einer beliebigen konstanten Diagonalmatrix

$$D := \text{diag}(\lambda_1, \dots, \lambda_p) = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_p \end{pmatrix}$$

durch DG ersetzt werden. Nun sieht die Gleichung $B = 0$, mit $G = DG$ für jedes i, j folgendermaßen aus:

$$\begin{aligned} (EDG(DG)^t)((EDG)^t)^{-1}C + C(EDG)^{-1}(EDG(DG)^t) &= 0 \\ \lambda_i \left((EGG^t)((E\dot{G})^t)^{-1}C \right)_j + C \left((E\dot{G})^{-1}(EGG^t) \right)_i \lambda_j &= 0 \end{aligned}$$

Daraus folgt, dass $C = 0$ ist. Es ist also:

$$\begin{aligned} C &= (EG^{(s)}EG^{(s)t} - EG^{(s)}EG^{*(s)t})(E\dot{G}^*)^{-1}EG^*G^{*t} = 0 \\ \Leftrightarrow EG^{(s)}EG^{(s)t}(E\dot{G}^*)^{-1}EG^*G^{*t} - EG^{(s)}EG^{*(s)t}(E\dot{G}^*)^{-1}EG^*G^{*t} &= 0 \\ \Leftrightarrow EG^{(s)}EG^{(s)t}(E\dot{G}^*)^{-1}EG^*G^{*t} &= EG^{(s)}EG^{*(s)t}(E\dot{G}^*)^{-1}EG^*G^{*t} \\ \Leftrightarrow EG^{(s)}EG^{(s)t} &= EG^{(s)}EG^{*(s)t} \end{aligned}$$

und es gilt (82).⁸² □

Mit Hilfe von Satz (6.3) folgt nun der Beweis von Satz (6.2).

Beweis. Mit Satz (6.3) gilt, dass G^* eine optimale Schätzfunktion ist, sofern

$$\left\{ E \left(\left(\frac{\partial G}{\partial \theta} \right)^t \right) \right\}^{-1} E(G(G^*)^t)$$

eine konstante Matrix für alle $G \in \mathcal{H}$ ist.

Sei nun $G := A(y - \mu) \in \mathcal{H}$. Wichtig für den Beweis sind die Formen von $E(G(G^*)^t)$ und $E\left(\left(\frac{\partial G}{\partial \theta}\right)^t\right)$.

⁸²siehe: [(Heyde 1997)]S.14

Mit $G^* = \left(\frac{\partial \mu}{\partial \theta}\right) V^{-1}(y - \mu)$ und $(G^*)^t = (y - \mu)^t V^{-1} \left(\frac{\partial \mu}{\partial \theta}\right)^t$ gilt:

$$\begin{aligned}
 E \{G(G^*)^t\} &= E \left\{ A(y - \mu)(y - \mu)^t V^{-1} \left(\frac{\partial \mu}{\partial \theta}\right)^t \right\} \\
 &= E \left\{ AE [(y - \mu)(y - \mu)^t | x] V^{-1} \left(\frac{\partial \mu}{\partial \theta}\right)^t \right\} \\
 &= E \left\{ AE [(y - \mu)(y - \mu)^t | x] \text{Var}(y|x)^{-1} \left(\frac{\partial \mu}{\partial \theta}\right)^t \right\} \\
 &= E \left\{ AE [(y - \mu)(y - \mu)^t | x] (E(y^2|x) - E(y|x)^2)^{-1} \left(\frac{\partial \mu}{\partial \theta}\right)^t \right\} \\
 &= E \left\{ AE [(y - \mu)(y - \mu)^t | x] E [(y - \mu)^2 | x]^{-1} \left(\frac{\partial \mu}{\partial \theta}\right)^t \right\} \\
 &= E \left\{ AE [(y - \mu)(y - \mu)^t | x] E [(y - \mu)(y - \mu)^t | x]^{-1} \left(\frac{\partial \mu}{\partial \theta}\right)^t \right\} \\
 &\Rightarrow E \{G(G^*)^t\} = E \left\{ A \left(\frac{\partial \mu}{\partial \theta}\right)^t \right\}
 \end{aligned}$$

und mit $\left(\frac{\partial G}{\partial \theta}\right)^t = \left(\frac{\partial A}{\partial \theta}\right)^t (y - \mu) - A \left(\frac{\partial \mu}{\partial \theta}\right)^t$ gilt:

$$\begin{aligned}
 E \left\{ \left(\frac{\partial G}{\partial \theta} \right)^t \right\} &= E \left\{ \left(\frac{\partial A}{\partial \theta} \right)^t (y - \mu) - A \left(\frac{\partial \mu}{\partial \theta} \right)^t \right\} \\
 &= E \left\{ \left(\frac{\partial A}{\partial \theta} \right)^t (y - \mu) \right\} - E \left\{ A \left(\frac{\partial \mu}{\partial \theta} \right)^t \right\} \\
 &= E \left\{ \left(\frac{\partial A}{\partial \theta} \right)^t E(y - \mu|x) \right\} - E \left\{ A \left(\frac{\partial \mu}{\partial \theta} \right)^t \right\} \\
 &= E \left\{ \left(\frac{\partial \mu}{\partial \theta} \right)^t V^{-1} E(y - \mu|x) \right\} - E \left\{ A \left(\frac{\partial \mu}{\partial \theta} \right)^t \right\} \\
 &= E \left\{ \left(\frac{\partial \mu}{\partial \theta} \right)^t \text{Var}(y|x)^{-1} E(y - \mu|x) \right\} - E \left\{ A \left(\frac{\partial \mu}{\partial \theta} \right)^t \right\} \\
 &= E \left\{ \left(\frac{\partial \mu}{\partial \theta} \right)^t E((y - \mu)(y - \mu)^t|x)^{-1} E(y - \mu|x) \right\} \\
 &\quad - E \left\{ A \left(\frac{\partial \mu}{\partial \theta} \right)^t \right\} \\
 &= E \left\{ \left(\frac{\partial \mu}{\partial \theta} \right)^t E((y - \mu)|x)^{-1} \right\} - E \left\{ A \left(\frac{\partial \mu}{\partial \theta} \right)^t \right\} \\
 &= -E \left\{ A \left(\frac{\partial \mu}{\partial \theta} \right)^t \right\}
 \end{aligned}$$

Daraus ergibt sich:

$$\begin{aligned}
 \left\{ E \left(\left(\frac{\partial G}{\partial \theta} \right)^t \right) \right\}^{-1} E(G(G^*)^t) &= \left(-E \left\{ A \left(\frac{\partial \mu}{\partial \theta} \right)^t \right\} \right)^{-1} E \left\{ A \left(\frac{\partial \mu}{\partial \theta} \right)^t \right\} \\
 &= -I
 \end{aligned}$$

wobei I die Identitätsmatrix darstellt. Da die Identitätsmatrix eine konstante Matrix ist, ist Satz (6.2) bewiesen. \square

Nachdem Schätzfunktion und Schätzgleichung dargestellt wurden, werden nun die verallgemeinerten Schätzgleichungen (GEE) erläutert.

Definition 6.7 (Verallgemeinerte Schätzgleichung für unabhängige Beobachtungen)

Die optimale Schätzfunktion aus Satz (6.2) ist definiert als:

$$G^* := \sum_{i=1}^n \left(\frac{\partial \mu_i}{\partial \beta} \right)^t V_i^{-1} (y_i - \mu_i)$$

dabei ist $y_i = (y_{ij})_{1 \leq j \leq n_i}$, $\mu_i = E(y_i)$ und $V_i = \text{Cov}(y_i)$. Die zugehörige Schätzgleichung wird verallgemeinerte Schätzgleichung (GEE) genannt und ist, sofern unabhängige Beobachtungen vorliegen, gegeben durch:

$$\sum_{i=1}^n \left(\frac{\partial \mu_i}{\partial \beta} \right)^t V_i^{-1} (y_i - \mu_i) = 0 \quad (85)$$

Satz 6.4

Seien Y_{ij} unabhängige Beobachtungen, die eine Verteilung einer Exponentialfamilie besitzen, wobei der Parameter ϕ unabhängig von β ist. $f_{y_{ij}}(\theta_{ij}, \phi)$ sei die zugehörige Dichtefunktion. Dann ist die verallgemeinerte Schätzgleichung

$$\sum_{i=1}^n \left(\frac{\partial \mu_i}{\partial \beta} \right)^t V_i^{-1} (y_i - \mu_i) = 0 \quad (86)$$

eine Gleichung für die Maximum-Likelihood-Schätzung

Beweis. Seien Y_{ij} mit $i = 1, \dots, n$ und $j = 1, \dots, n_i$ Zufallsvariablen mit einer Verteilung der Exponentialfamilie, also mit einer Dichtefunktion der 1-parametrischen Exponentialfamilie gegeben, sowie ϕ und β voneinander unabhängig. Nach Satz (4.2) gilt: $E(Y_{ij}) = b'(\theta_{ij})$ und damit auch:

$$g(E(Y_{ij})) = g(b'(\theta_{ij}))$$

und

$$g(E(Y_{ij})) = x_{ij}^t \beta$$

wobei g die Linkfunktion darstellt. Die letzte Gleichung gilt alleine aufgrund

der Voraussetzungen einer Linkfunktion. Es folgt also:

$$g(b'(\theta_{ij})) = x_{ij}^t \beta \quad (87)$$

$$\Leftrightarrow \theta_{ij} = (b')^{-1}(g^{-1}(x_{ij}^t \beta)) \quad (88)$$

Wegen der Unabhängigkeit der Daten ist die gemeinsame Dichte gleich dem Produkt der einzelnen Dichtefunktionen von Y_{ij} . Nun können die Maximum-Likelihood-Schätzungen bestimmt werden. Dafür sei zunächst:

$$\begin{aligned} \ln \prod_{i=1}^n \prod_{j=1}^{n_i} f_{y_{ij}}(\theta_{ij}, \phi) &= \sum_{i=1}^n \sum_{j=1}^{n_i} \ln f_{y_{ij}}(\theta_{ij}, \phi) \\ &\stackrel{88}{=} \sum_{i=1}^n \sum_{j=1}^{n_i} \ln f_{y_{ij}}(((b')^{-1}(g^{-1}(x_{ij}^t \beta)), \phi)) \end{aligned}$$

Die Maximum-Likelihood-Schätzung kann nun durch folgende Gleichung beschrieben werden:

$$\begin{aligned} &\frac{\partial \sum_{i=1}^n \sum_{j=1}^{n_i} \ln f_{y_{ij}}(((b')^{-1}(g^{-1}(x_{ij}^t \beta)), \phi))}{\partial \beta} \\ &= \sum_{i=1}^n \sum_{j=1}^{n_i} \frac{\partial \ln f_{y_{ij}}(((b')^{-1}(g^{-1}(x_{ij}^t \beta)), \phi))}{\partial \beta} = 0 \end{aligned}$$

Mit der Kettenregel⁸³ gilt weiter:

$$\sum_{i=1}^n \sum_{j=1}^{n_i} \frac{\partial \ln f_{y_{ij}}(((b')^{-1}(g^{-1}(x_{ij}^t \beta)), \phi))}{\partial((b')^{-1}(g^{-1}(x_{ij}^t \beta)), \phi)} \left(\frac{\partial(b')^{-1}(g^{-1}(x_{ij}^t \beta))}{\partial \beta}, \frac{\partial \phi}{\partial \beta} \right)^t$$

da ϕ abgeleitet nach β gleich Null ist. Durch Einsetzen von θ_{ij} ergibt sich:

$$= \sum_{i=1}^n \sum_{j=1}^{n_i} \frac{\partial \ln f_{y_{ij}}(\theta_{ij}, \phi)}{\partial \theta_{ij}} \left(\frac{\partial(b')^{-1}(g^{-1}(x_{ij}^t \beta))}{\partial \beta} \right)$$

Durch nochmalige Anwendung der Kettenregel ist das gleich:

$$= \sum_{i=1}^n \sum_{j=1}^{n_i} \frac{\partial \ln f_{y_{ij}}(\theta_{ij}, \phi)}{\partial \theta_{ij}} \left(\frac{\partial(b')^{-1}(g^{-1}(x_{ij}^t \beta))}{\partial g^{-1}(x_{ij}^t \beta)} \right) \left(\frac{\partial g^{-1}(x_{ij}^t \beta)}{\partial \beta} \right)$$

⁸³Die Kettenregel für zwei differenzierbare Funktionen g und h sieht folgendermaßen aus:
 $(g \circ h)'(x) = g'(h(x))h'(x)$

Nach diesen Umformungen wird die Dichtefunktion abgeleitet. Dabei ist zu beachten, dass die Dichte der einer 1-parametrischen Exponentialfamilie entspricht.

Die Ableitung ist gegeben durch:

$$\sum_{i=1}^n \sum_{j=1}^{n_i} (y_{ij} - b'(\theta_{ij})) \frac{w}{\phi} \left(\frac{\partial(b')(\theta_{ij})}{\partial\theta_{ij}} \right)^{-1} \left(\frac{\partial x_{ij}^t \beta}{\partial\beta} \right)$$

Da die Ableitung von $b'(\cdot)$ $b''(\cdot)$ ist, ist das

$$\begin{aligned} &= \sum_{i=1}^n \sum_{j=1}^{n_i} (y_{ij} - b'(\theta_{ij})) \frac{w}{\phi} (b''(\theta_{ij}))^{-1} \left(\frac{\partial\mu_{ij}}{\partial\beta} \right) \\ &= \sum_{i=1}^n (y_i - b'(\theta_i))^t \text{diag} \left(b''(\theta_i) \frac{\phi}{w} \right)^{-1} \left(\frac{\partial\mu_i}{\partial\beta} \right) \\ &= \sum_{i=1}^n (y_i - b'(\theta_i))^t V_i^{-1} \left(\frac{\partial\mu_i}{\partial\beta} \right) \end{aligned}$$

da nach Satz (4.2) gilt, dass $b''(\theta_i) \frac{\phi}{w} = \text{Var}(y_i)$ und $b'(\theta_{ij}) = \mu_{ij}$ bzw. $b'(\theta_i) = \mu_i$ und $V_i := \text{Var}(y_i)$ hier die Kovarianzmatrix von y_i ist. Damit ist, wenn die Transponierte der letzten Gleichung betrachtet und gleich Null gesetzt wird:

$$\sum_{i=1}^n \left(\frac{\partial\mu_i}{\partial\beta} \right)^t V_i^{-1} (y_i - \mu_i) = 0$$

(vergleiche mit Gleichung (86)) Damit ist Satz (6.4) bewiesen. □

Bemerkung 6.5

Für diesen Satz und speziell für Gleichung (85) wird angenommen, dass die Kovarianzmatrix V_i bekannt ist, ansonsten kann die Gleichung nicht einfach gelöst werden. Leider entspricht das nicht immer der Realität, in der die Kovarianzmatrizen meist unbekannt sind. Die Kovarianzmatrix hängt beim GLMM meist neben dem Parameter β noch von dem Vektor der Varianzkomponenten θ ab, es ist folglich $V_i = V_i(\beta, \theta)$. Falls die V_i s unbekannt sind, gibt es eine Methode, vorgestellt von [(Liang und Zeger 1986)], in der die V_i s durch sogenannte „working covariance matrices“ ersetzt werden, um den GEE Schätzer

zu erhalten. Die „working covariance matrices“ haben folgende Form

$$\tilde{V}_i = A_i^{\frac{1}{2}} R(\alpha) A_i^{\frac{1}{2}} \quad (89)$$

und die dazugehörige verallgemeinerte Schätzgleichung ist gegeben durch:

$$\sum_{i=1}^n \left(\frac{\partial \mu_i}{\partial \beta} \right)^t A_i^{\frac{1}{2}} R(\alpha) A_i^{\frac{1}{2}} (y_i - \mu_i) = 0 \quad (90)$$

mit ϕ aus der Dichtefunktion der 1-parametrischen Exponentialfamilie (24), $A_i = \text{diag} \{b''(\theta) \frac{\phi}{w}\}$ (Nach Satz (4.2) ist $b''(\theta) \frac{\phi}{w}$ gleich der Varianz von y_i) und $R(\alpha)$ stellt die „working covariance matrix“ dar. $R(\alpha)$ ist eine $n \times n$ Matrix, die die Bedingungen einer Korrelationsmatrix erfüllt. Gleichung (90) ist somit die verallgemeinerte Schätzgleichung für korrelierte Beobachtungen. \tilde{V}_i (89) ist dann äquivalent zur wahren Kovarianzmatrix V_i und es kann damit weiter gerechnet werden. Es stellt sich noch die Frage, ob diese „working covariance matrices“ den GEE Schätzer nicht verzerren.

In [(Liang und Zeger 1986)] wird gezeigt, dass der GEE Schätzer unter bestimmten Bedingungen konsistent ist, obwohl die „working covariance matrices“ nicht mit den wahren V_i s übereinstimmen. Im Falle von unabhängigen Beobachtungen kann Gleichung (85) auch folgendermaßen formuliert werden:

$$\sum_{i=1}^n \left(\frac{\partial \mu_i}{\partial \beta} \right)^t A_i^{\frac{1}{2}} I_{n_i} A_i^{\frac{1}{2}} (y_i - \mu_i) = 0 \quad (91)$$

wobei I_{n_i} die $n_i \times n_i$ Identitätsmatrix darstellt und $V_i := A_i^{\frac{1}{2}} I_{n_i} A_i^{\frac{1}{2}}$ ist.

Wie oben schon erwähnt, bieten die GEE gute Möglichkeiten den festen Parameter β zu schätzen. In der Analyse von longitudinalen Daten wird diese Methode häufig angewandt, da es dort oft von Interesse ist, den Parameter β zu schätzen. Natürlich gibt es immer noch die Alternative der Likelihood-Methoden⁸⁴ um den Parameter β zu schätzen, jedoch ist die Likelihood-Methode oftmals sehr aufwendig anzuwenden und es kann zu Verzerrungen der Schätzungen kommen, sofern das Modell, welches gewählt wurde, nicht das passende ist. Dieses ist oft bei der Analyse von longitudinalen Daten der Fall.

⁸⁴siehe: (Jiang 2007) S. 185

Die GEE-Methode ist somit im direkten Vergleich mit der Likelihood-Methode meist einfacher zu berechnen und wird deshalb auch gerne verwendet. Bedeutender ist hier aber die Tatsache, dass bei der GEE-Methode nicht die komplette Ausprägung der Verteilung der Daten benötigt wird, was häufig von Vorteil ist. Natürlich bedarf es bei den GEE-Schätzern auch einiger Voraussetzungen, wie z.B. die korrekte Vorgabe der Erwartungswerte μ_i für $1 \leq i \leq n$ und der Kovarianzmatrizen V_i , welche für die optimale Schätzgleichung gemäß Satz (6.2), von Bedeutung sind. Allerdings sind diese Voraussetzungen schwächer, als wenn die gesamte Verteilung der Daten bekannt sein muss, wie es bei der Likelihood-Methode der Fall ist.

Es gibt noch eine weitere Methode der verallgemeinerten Schätzgleichungen, die „*iterativen verallgemeinerten Schätzgleichungen*“. Wie schon erwähnt, benötigt man für die verallgemeinerten Schätzgleichungen die „wahren“ Kovarianzmatrizen V_i . Diese iterative Methode um einen Schätzer zu erhalten, die asymptotisch genauso effizient ist, wie die GEE-Schätzer, wird von [(Jiang 2007)] vorgestellt, wird im Rahmen dieser Arbeit allerdings nicht weiter behandelt.

7 Analyse von longitudinalen Daten

7.1 Longitudinale Daten

Es wurden in dieser Arbeit bereits mehrfach die longitudinalen Daten erwähnt. Im Folgenden werden diese Daten vorgestellt und deren Besonderheiten herausgearbeitet. Der genaue Ablauf von Studien mit longitudinalen Daten wird anschließend noch detaillierter dargestellt. Jedoch geht es zunächst um die Erklärung was longitudinale Daten sind und durch welche Eigenschaften solche Daten geprägt sind.

Bemerkung 7.1

Es gibt zwei Datentypen, die zu unterscheiden sind:

1. *„repeated measures Daten“*
2. *„longitudinale Daten“*

Definition 7.1 (Repeated measures Daten)

Repeated measures Daten sind Daten, bei denen die Zielgröße mehrfach unter verschiedenen experimentellen Bedingungen für die Subjekte (Beobachtungseinheiten) gemessen werden, wie es z.B. bei der Pulsmessung an mehreren Versuchspersonen nach unterschiedlichen sportlichen Aktivitäten der Fall ist.

Definition 7.2 (Longitudinale Daten)

Longitudinale Daten sind spezielle repeated measures Daten. Bei diesem Datentyp wird die Zielgröße für die Subjekte mehrfach in zeitlicher Abfolge gemessen. Solche gruppierten Daten benötigen spezielle Analysemethoden.

Ein Beispiel für longitudinale sind Blutdruckmessungen, wobei die Messungen mehrfach an mehreren Patienten im Zeitraum eines Jahres gemessen werden.

Besonderheiten der longitudinalen Daten:

1. Daten vom gleichen Subjekt sind sich meistens ähnlicher als Daten verschiedener Subjekte; Hier handelt es sich folglich um korrelierte, zeitlich oder räumlich zusammenhängende, Daten.

2. Oft ist eine spezielle Korrelationsstruktur vorhanden, z.B. eine abnehmende Korrelation über die Zeit. Hier sind Messungen, die kurz hintereinander gemacht wurden, stärker korreliert als Messungen mit großem zeitlichen Abstand.
3. Es gibt eine interindividuelle⁸⁵ und eine intraindividuelle⁸⁶ Variabilität. Die intraindividuellen Variabilität lässt sich noch einmal unterscheiden zwischen einer Variabilität, die auf die longitudinale Struktur zurückgeht, und einer Variabilität durch zusätzliche Störgrößen, wie z.B. Messfehler.

Laut [(Diggle 2002)] ist die Analyse von longitudinalen Daten einfacher gestaltet, als z.B. die Zeitreihenanalyse, da die Daten als stochastisch unabhängig angenommen werden.⁸⁷

7.2 Die Analyse und ein Beispiel

Die Definition der longitudinalen Daten wurde gerade erläutert. Im Folgenden wird gezeigt, wie die Analyse von longitudinalen Daten abläuft. Es ist zu erwähnen, dass bei solch einer Analyse meist ein Regressionsmodell wie das lineare Modell verwendet wird.⁸⁸

Bemerkung 7.2 (Analyse von longitudinalen Daten)

Sei Y_{ij} eine Antwortvariable und x_{ij} ein Vektor der Länge p von erklärenden Variablen, die zur Zeit t_{ij} für $i = 1, \dots, n; j = 1, \dots, n_i$ betrachtet werden. Der Erwartungswert von Y_{ij} sei μ_{ij} und die Varianz v_{ij} . Sei weiter $Y_i = (Y_{i1}, \dots, Y_{in_i}), Y = (Y_1, \dots, Y_n)$ mit Erwartungswert $E(Y_i) = \mu_i$ und Kovarianzmatrix $Cov(Y_i) = V_i$, wobei $cov(Y_{ij}, Y_{ik}) = v_{ijk}$, d.h. das ij -te Element von V_i ist die Kovarianz zwischen Y_{ij} und Y_{ik} . Damit hat die Kovarianzmatrix von Y_i folgende Form:

$$Cov(Y_i) = V_i = cov(Y_{ij}, Y_{ik})_{j,k=1,\dots,n_i} = v_{ijk}$$

⁸⁵also zwischen zwei oder mehreren Individuen

⁸⁶innerhalb eines Individuums

⁸⁷siehe: [(Diggle 2002)] S.2

⁸⁸siehe: [(Diggle 2002)] S.15

Damit ist:

$$V_i = \text{Cov}(X) = \begin{pmatrix} \text{cov}(Y_{i1}, Y_{i1}) & \cdots & \text{cov}(Y_{i1}, Y_{in_i}) \\ \vdots & \ddots & \vdots \\ \text{cov}(Y_{in_i}, Y_{i1}) & \cdots & \text{cov}(Y_{in_i}, Y_{in_i}) \end{pmatrix}$$

Folglich ist V_i definiert durch:

$$V_i = \begin{pmatrix} v_{i11} & \cdots & v_{i1n_i} \\ \vdots & \ddots & \vdots \\ v_{in_i1} & \cdots & v_{in_in_i} \end{pmatrix}$$

Die longitudinale Datenanalyse basiere auf folgendem linearen Modell (man vergleiche mit (9)):

$$\begin{aligned} Y_{ij} &= \beta_1 x_{ij1} + \beta_2 x_{ij2} + \dots + \beta_p x_{ijp} + \epsilon_{ij} \\ &= x_{ij}^t \beta + \epsilon_{ij} \end{aligned} \quad (92)$$

wobei $\beta = (\beta_1, \dots, \beta_p)$ der Vektor von unbekanntem Regressionskoeffizienten ist. Die jeweilige Gleichung für den i -ten Wert von (92) sieht wie folgt aus:

$$Y_i = X_i \beta + \epsilon_i \quad (93)$$

wobei X_i eine $n_i \times p$ Designmatrix ist mit x_{ij} in der j -ten Spalte, es ist somit:

$$X_i = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n_i1} & \cdots & x_{n_ip} \end{pmatrix}$$

und

$$\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{in_i}).$$

Bei longitudinalen Studien muss berücksichtigt werden, dass die experimentelle Einheit nicht die individuelle Größe Y_{ij} ist, sondern das ganze Y_i , also die Größe der einzelnen Subjekte.⁸⁹

Bevor die Vorteile von Longitudinalstudien herausgestellt werden, geht es

⁸⁹siehe: [(Diggle 2002)]

zunächst darum, longitudinale Daten näher zu betrachten, indem unter anderem ein Beispiel verwendet wird.

Die Analyse von longitudinalen Daten basiert auf bestimmten Versuchsdesigns, in der jede Beobachtungseinheit zum selben Zeitpunkt gemessen wird und das bestenfalls ohne fehlende Daten. Die Analyse mit fehlenden Daten wird in einem späteren Unterkapitel näher erläutert. Für den Fall, dass die Anzahl der Subjekte groß ist in Relation zur Anzahl der Beobachtungen pro Subjekt, wird in den meisten Fällen eine multivariate Analysemethode verwendet.

Innerhalb der letzten Jahre wurden viele Fortschritte in Bezug auf longitudinale Studien gemacht. Zum Beispiel wurde herausgefunden, dass zunächst alle Messungen für jedes Subjekt zusammengefasst werden sollte, um im folgenden Schritt diese „Zusammenfassung“ zu analysieren.

Um sich diesen Sachverhalt zu verdeutlichen, wird z.B. angenommen, dass es zwei Gruppen von Subjekten gibt, wobei jede Beobachtung aus einer Beobachtungseinheit approximativ durch eine Regressionsgerade angepasst werden kann. Eine der beiden Gruppen erhält eine Behandlung, also im Falle einer Krankheit ein Medikament, die andere Gruppe bekommt lediglich ein Placebo verabreicht. Die Behandlung, also das Medikament, verursacht das Gefälle der Gerade. Im Falle der Krankheit kann man sich zum besseren Verständnis vorstellen, dass eine Hypertonie, also Bluthochdruck, vorliegt. Bei solch einem Versuch werden die Subjekte über einen bestimmten Zeitraum verfolgt, wobei der Blutdruck in bestimmten Abständen gemessen wird. Die Frage, die man sich bei solch einem Versuch stellt ist, ob sich die Erwartungswerte der beiden Gruppen signifikant unterscheiden. Ein Modell für die erste Gruppe ist gegeben durch:

$$Y_{1j} = \beta_1 + \alpha_1 + \epsilon_{1j}$$

und das statistische Modell der zweiten Gruppe sieht folgendermaßen aus:

$$Y_{2j} = \beta_1 + \beta_2 + \alpha_2 + \epsilon_{2j} \tag{94}$$

Bei beiden Gleichungen stehen die Zahlen 1 und 2 für die jeweilige Beobachtungseinheit der verschiedenen Gruppen und j steht jeweils für eine Beobachtung in einer Beobachtungseinheit, also in einem Subjekt. Man sieht, dass die Modelle jeweils feste Effekte (β_1 : das ist der Erwartungswert des Blutdrucks von den Subjekten in Gruppe 1 und β_2 ist die Differenz zwischen den Er-

wartungswerten von den Gruppen 1 und 2), sowie Zufallseffekte (α_1 und α_2) enthalten. Bei einem Modell, wie es in Gleichung (94) gegeben ist, ist es von Interesse eine statistische Hypothese aufzustellen und etwas zu testen. In diesem Fall ist eine mögliche Hypothese gegeben durch:

$$H_0 : \beta_2 = 0$$

Das bedeutet, dass die Erwartungswerte für beide Gruppen gleich sind, da β_2 die Differenz der beiden Erwartungswerte ist, mit dem Ergebnis, dass beide Behandlungen den gleichen Effekt im Bezug auf den Blutdruck aufweisen.

Zurück zur Gerade, welche die Beobachtungen beschreibt. Diese Gerade kann durch eine lineare Regression für die Daten der einzelnen Subjekte angepasst werden. Das Gefälle der Gerade kann zwischen den beiden Gruppen durch einen *2-Stichproben t-Test* oder im nichtparametrischen Fall durch den *Wilcoxon-Rangsummentest* verglichen werden.

Seien nun die Daten der unterschiedlichen Subjekte als unabhängig vorausgesetzt, dann sind auch die Gefälle der Geraden voneinander unabhängig.

Werden die „zusammengefassten“ Messungen verwendet, muss als Voraussetzung gegeben sein, dass die Daten ausgewogen sind, so dass die zusammengefassten Daten als nur eine Zufallsstichprobe mit einer Verteilungsfunktion angenommen werden kann. Sind die Daten nicht ausgewogen, kann es vorkommen, dass die Varianzen der unterschiedlichen Subjekte nicht übereinstimmen. Bei der Analyse von longitudinalen Daten können einige Komplikationen auftreten. In erster Linie sind die Missing Values zu nennen. Eine Konsequenz daraus ist, dass der 2-Stichproben t-Test nicht mehr angewendet werden kann. Die verschiedenen Subjekte werden zu verschiedenen Zeitpunkten gemessen und der Erwartungswert im Blutdruckbeispiel würde für die unterschiedlichen Subjekte auch unterschiedliche Varianzen haben, was jedoch den Voraussetzungen eines t-Tests widerspräche.

Während die Zeit nicht willkürlich gewählt werden kann ist das bei Folge der applizierten Behandlung möglich. In dem Beispiel ist die normale Annahme, dass die wiederholten Beobachtungen in jedem Subjekt um den jeweiligen Erwartungswert zufällig verteilt sind. Für den Fall, dass die Zeitpunkte, an dem die Beobachtungen gemessen werden, zu nah beieinander liegen, wird eine Beobachtung, die oberhalb des Erwartungswertes liegt wahrscheinlicher, sofern die vorausgegangene Beobachtung ebenfalls oberhalb des Erwartungswertes

liegt. Das wird Autokorrelation⁹⁰ genannt und wird schwächer, sofern die Zeitintervalle sich vergrößern. Das Negative an autokorrelierten Daten ist, dass die meisten statistischen Tests nicht zur Anwendung kommen können.

Eine weitere Komplikation, die bei der Analyse von longitudinalen Daten auftreten kann sind ungleiche Beobachtungen. Dieses kann zum Beispiel durch ungleiche Zeitintervalle verursacht werden, wo es nicht möglich ist noch eine neue Messung vorzunehmen. Diese Komplikation spielt allerdings keine relevante Rolle, sofern eine einfache Situation vorliegt, in der es keine Autokorrelation gibt. So sind die Daten statistisch unabhängig und jedes Subjekt kann zur gleichen Anzahl von Zeitpunkten gemessen werden.

Im Folgenden wird ein Beispiel mit longitudinalen Daten aufgeführt. Es ist zu beachten, dass genau ein fehlender Wert vorliegt.

Beispiel 7.1 (Longitudinale Daten)

Es wurde ein Experiment durchgeführt, wo zwei Gruppen betrachtet wurden. Die Tabelle (2) zeigt die Ergebnisse des Experiments. Dabei enthält jede Zeile

| Zeit in Minuten | 0 | 5 | 10 | 20 | 30 | 45 | 60 |
|-----------------|-----|----|-----|----|----|----|-----|
| Gruppe 1 | 100 | 90 | 100 | 70 | 36 | 50 | 28 |
| | 100 | 83 | 97 | 60 | 83 | 83 | 97 |
| | 70 | 83 | 77 | 21 | 36 | 53 | 53 |
| | 77 | 36 | 36 | 36 | 36 | 36 | 28 |
| | 100 | 61 | 99 | 83 | 97 | 83 | 100 |
| Gruppe 2 | 90 | 28 | 44 | 33 | 36 | 44 | 36 |
| | 83 | 14 | 28 | 28 | 21 | 28 | 36 |
| | 69 | 36 | 28 | 21 | 14 | 21 | 21 |
| | 100 | 77 | 44 | 28 | 44 | 14 | |
| | 61 | 53 | 53 | 44 | 44 | 28 | 44 |

Tabelle 2: Zwei Gruppen Experiment

die Beobachtungen, die zu bestimmten Zeitpunkten (nach 0, 5, 10, 20, 30, 45, 60 Minuten) gemessen wurden. Den beiden Gruppen wurde jeweils unterschiedliche Behandlungen verabreicht.

Bei diesem Experiment stellt sich die Frage, ob beide Gruppen dieselben oder unterschiedliche Reaktionen über den Lauf der Zeit zeigen. Um sich die Daten

⁹⁰Die Autokorrelation bezeichnet eine Korrelation zwischen aufeinander folgenden Werten innerhalb einer Serie von Messwerten. Sie tritt zum Beispiel dann auf, wenn die Messwerte zeitabhängig aufgenommen werden und die Werte nicht voneinander unabhängig sind.

ein zu veranschaulichen, werden in den Abbildungen (3)(von Gruppe 1) und (4)(von Gruppe 2) die Datenkurven dargestellt. In Abbildung (4) ist deutlich zu erkennen, dass es einen fehlenden Wert gibt. (türkise Linie endet früher)

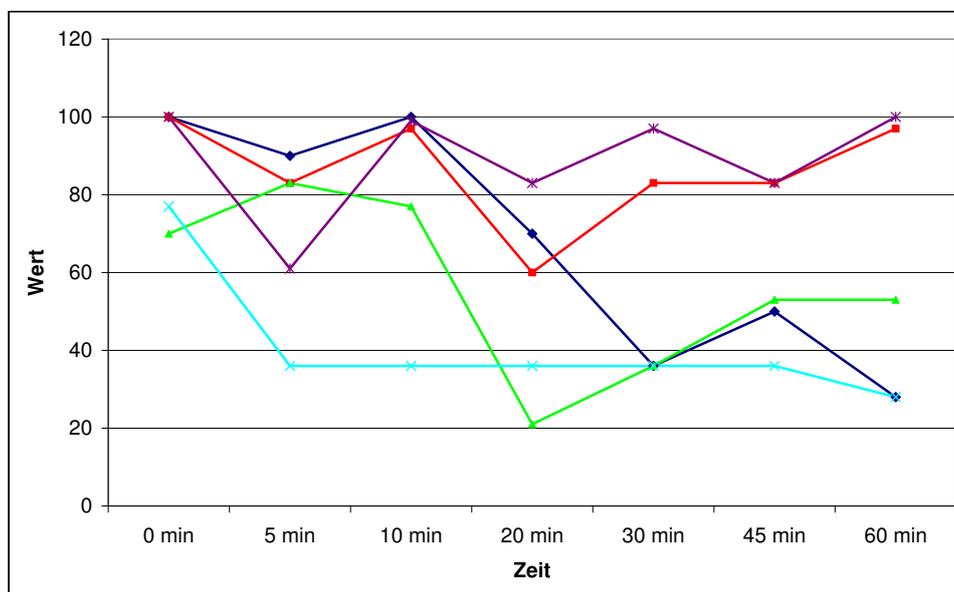


Abbildung 3: Datenkurve von Gruppe 1

Als folgenden Schritt werden nun auch die Mittelwerte und die Standardabweichungen für die verschiedenen Zeitpunkte (das sind jeweils fünf und in einem Fall vier Beobachtungen) der beiden Gruppen berechnet. Es fällt auf, dass die zweite Gruppe eine höhere Standardabweichung als die erste und diese wiederum größere Standardfehler als die zweite Gruppe hat. In Anhang (A) sind die SAS-Anweisungen, sowie die Outputs für die Means-Prozedur zu finden.

Als statistisch geeignetes Modell für die Daten ist das folgende aufzuführen:

$$y_{ij} = \beta_{ji} + \alpha_i + \epsilon_{ij} \quad (95)$$

wobei anstelle des Indizes i immer die Gruppe 1 oder 2 eingesetzt wird. Das Indize j steht für die jeweilige Zeit, d.h. es können sieben Werte eingesetzt werden. Damit gibt es 14 feste Effekte (β_{ji}). Dieses Modell beinhaltet einen Zeiteffekt, einen Gruppeneffekt und die Gruppe bei einer Zeitüberlagerung. Ein weiteres Modell für die Analyse der Daten, welches reduziert ist, ist für Gruppe 1 gegeben durch:

$$y_{1j} = \beta_{j1} + \alpha_1 + \epsilon_{1j}$$

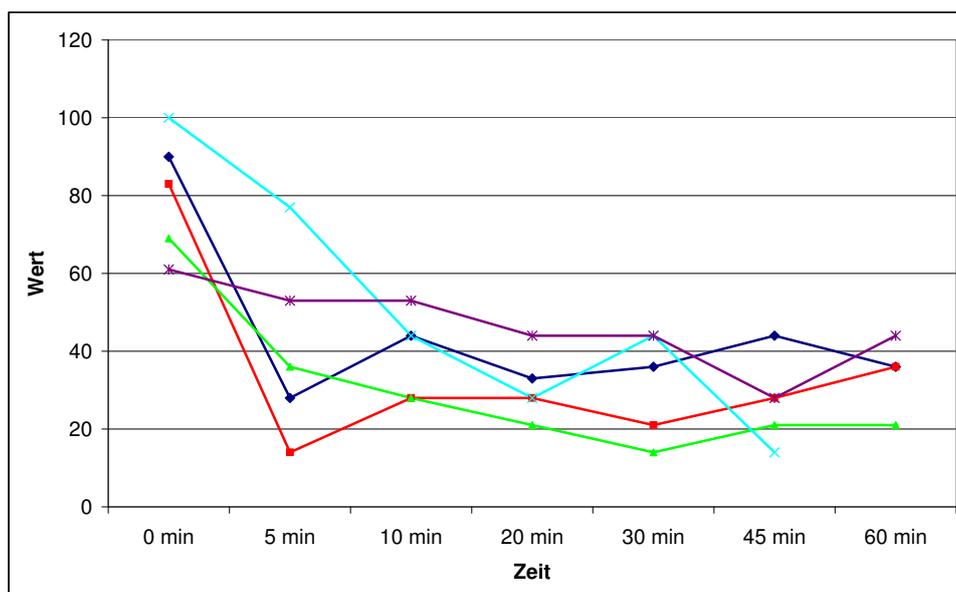


Abbildung 4: Datenkurve von Gruppe 1

In Matrixform sieht diese Gleichung so aus:

$$y_1 = \begin{pmatrix} I & 0 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + 1\alpha_1 + \epsilon_1 \quad (96)$$

und für Gruppe 2:

$$y_{2j} = \beta_{j1} + \beta_2 + \alpha_2 + \epsilon_{2j}$$

Wieder in Matrixform geschrieben, sieht diese Gleichung so aus:

$$y_2 = \begin{pmatrix} I & I \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + 1\alpha_2 + \epsilon_2 \quad (97)$$

Bei den Matrixgleichungen stellt I jeweils eine 7×7 Identitätsmatrix dar, 0 ist eine 7×7 Matrix bestehend aus Nullen und β ist jeweils ein 7×1 Spaltenvektor. Nun kann zum Beispiel auf einen Zeiteffekt getestet werden. Die Nullhypothese ist in diesem Fall gegeben durch:

$$H_0 : \beta_{11} = \beta_{21} = \dots = \beta_{71}$$

Um den Hypothesentest ausführen zu können, müssen die beiden Varianzen $\text{Var}(\alpha) = \sigma_\alpha^2$ und $\text{Var}(\epsilon) = \sigma_\epsilon^2$ geschätzt werden. Im Allgemeinen kommt bei

der Analyse heraus, dass die Daten keine Autokorrelation aufweisen, sofern die Subjekteffekte in das Modell eingebunden werden. Die Analyse ergibt weiter, dass ein signifikanter Zeiteffekt und ein signifikanter Gruppeneffekt vorliegt, es existiert jedoch kein signifikanter Gruppeneffekt im Bezug auf eine Zeitüberlagerung. (Genauer kann in [(Jones 1993)] nachgelesen werden) Da in Kapitel (8) für die Daten der Forschungsstudie Multiple Mittelwertvergleiche, sowie Signifikanztests durchgeführt werden, werden die Tests auf Signifikanzen hier nicht explizit durchgeführt.

7.3 Verschiedene Ansätze der Analyse longitudinaler Daten

Wenn repeated measures Daten⁹¹ vorliegen, gibt es mehrere Vorgehensweisen bei der Analyse.

Eine Strategie, welche meist gut anzuwenden ist, ist Folgende:

1. Als ersten Schritt dieser Strategie müssen die repeated measures Daten soweit reduziert werden, dass sie sich in zwei Gruppen zusammenfassen lassen.
2. Als zweiten Schritt wird jede Variable x_i der beiden „Zusammenfassungen“ analysiert, als eine Funktion der Kovarianzen.

Diese Art der Analyse bezeichnet man als „Zwei-Phasen-Analyse“, da sie in zwei Punkten abgearbeitet werden muss. Die Zwei-Phasen-Analyse kann immer dann angewendet werden, wenn $x_{ij} = x_i$ für alle i und j , die in Phase (1) der Analyse herausgekommen sind, und nur mit x_i in Phase (2) regressiert werden kann.

Diese Analyse ist leider unvorteilhaft, wenn eine wichtige erklärende Variable sich über die Zeit verändert.

Anstelle davon, dass die wiederholten Daten in Statistiken reduziert werden, in denen die Daten zunächst zusammenfasst werden, können auch andere Strategien benutzt werden, wo jedes einzelne y_{ij} im Bezug auf x_{ij} modelliert wird.

Darunter fallen drei bestimmte Strategien, die aussagen, wie in solch einem

⁹¹Was repeated measures Daten sind, wurde bereits im zweiten Kapitel erläutert

Fall analysetechnisch vorgegangen werden kann. Diese drei Modelle, vorgestellt von [(Diggle 2002)], werden nun näher dargelegt.

Beim ersten Modell wird, wie auch bei einer Querschnittsstudie,⁹² der marginalen Erwartungswert modelliert. Die marginale Analyse beinhaltet bestimmte Annahmen über die Gestalt einer Korrelation, sofern die „repeated values“ nicht unabhängig sind. Im linearen Modell bedeutet das, dass man folgende Annahmen trifft:

$$E(Y_i) = X_i\beta$$

und

$$Var(Y_i) = V_i(\alpha)$$

Hier sind α und β die zu schätzenden Parameter.

Dieses Modell, das auf einem marginalen Ansatz beruht, ist vorteilhaft, da der Erwartungswert und die Kovarianz getrennt voneinander modelliert werden. Selbst wenn eine nicht korrekte Form der Varianz $V(\alpha)$ bei dem Modell angenommen wird, können dennoch in manchen Fällen allgemein gültige Folgerungen für den Parameter β aufgestellt werden.

Im zweiten Ansatz den [(Diggle 2002)] beschreibt, einem Modell für die Zufallseffekte, wird angenommen, dass eine Korrelation zwischen den wiederholten Antworten besteht. Dieses geht darauf zurück, dass die Regressionskoeffizienten zwischen den Individuen unterschiedlich sind. In diesem Ansatz wird der bedingte Erwartungswert von der Zufallsvariable Y_{ij} , welche durch den Koeffizienten β_i gegeben ist, modelliert. Das sieht folgendermaßen aus:

$$E(Y_{ij}|\beta_i) = x_{ij}^t\beta_i$$

Wichtig zu beachten ist, dass genügend Daten gegeben sind, denn im Falle einer zu geringen Datenmenge würde es Komplikationen geben den festen Parameter β_i aus (Y_i, X_i) zu schätzen. Daher wird weiter angenommen, dass die β_i alle unabhängig sind mit einer Verteilung dessen Erwartungswert gleich β ist. Sei nun:

$$\beta_i = \beta + A_i$$

⁹²Querschnittsstudien sind Studien, bei denen nicht wie bei Longitudinalstudien mehrere Zeitpunkte gemessen werden, sondern die Daten lediglich zu einem bestimmten Zeitpunkt erhoben werden.

wobei β den festen Effekt darstellt und A_i eine Zufallsvariable mit dem Erwartungswert 0 ist. Das heisst, dass es unbeobachtete Faktoren gibt, welche durch A_i dargestellt werden und welche gleich der Antworten von beispielsweise Personen sind. Da die Beobachtungen zwischen den verschiedenen Personen unterschiedlich ausfallen können, führen sie daher die Korrelation herbei. Diese Zufallseffektmodelle, wie dieses eins ist, sind immer dann nützlich, wenn Rückschlüsse über Individuen gemacht werden.

Der dritte und letzte Ansatz, der von [(Diggle 2002)] diskutiert wird, beschäftigt sich mit dem bedingten Erwartungswert⁹³ von Y_{ij} . Bei diesem Modell geht es darum ein Regressionsmodell für den bedingten Erwartungswert $E(Y_{ij}|Y_{ij-1}, Y_{ij-2}, \dots, Y_{i1}, x_{ij})$ zu finden. Das Regressionsmodell ist eine eindeutige Funktion von x_{ij} und den vergangenen Beobachtungen. Bei solch einem Modell werden die Annahmen an die Abhängigkeitsstruktur von Y und x und die Korrelation zwischen den wiederholten Beobachten Y in eine einzige Gleichung gebracht, indem sie miteinander kombiniert werden.

Alle drei Ansätze haben eines gemeinsam: Bei allen Modellen wird sowohl die Abhängigkeit der Beobachtung Y mit der erklärenden Variable x , als auch die Autokorrelation zwischen den Beobachtungen modelliert. In einer Querschnittsstudie ist das anders, dort ist lediglich die Abhängigkeit von Y und x von Bedeutung. Es gibt in solch einer Studie keine Korrelation, die modelliert werden muss.

Bemerkung 7.3

Sofern die Korrelation bei longitudinalen Daten nicht beachtet wird, kann es zu einigen Fehlinterpretationen kommen, welche durch die folgenden drei Auswirkungen zusammengefasst werden:

- 1. Es kommt zu falschen Schlüssen über den Regressionskoeffizienten β_i .*
- 2. Die Schätzungen von β sind nicht effizient und weniger genau.*
- 3. Die Daten können aufgrund von Missing Values verzerrt werden.*

Was genau passiert, wenn die Korrelation einfach ignoriert wird, wird im Rahmen dieser Arbeit nicht behandelt.⁹⁴

⁹³siehe: Definition 3.11

⁹⁴siehe: [(Diggle 2002)]

Bemerkung 7.4

Im Allgemeinen können die Probleme, welche bei der Analyse von longitudinalen Daten möglicherweise auftreten, in zwei Gruppen eingeteilt werden:

- 1. Probleme, wenn die Regression von Y bezüglich x im Mittelpunkt steht und zudem die Anzahl der experimentellen Einheiten sehr viel größer ist als die Anzahl der Beobachtungen pro Einheit.*
- 2. Probleme die auftreten, wenn die Korrelation für die Analyse wichtig ist und wenn die Anzahl der experimentellen Einheiten relativ gering ist.*

In den meisten Fällen treffen eins der beiden Kriterien nicht genau auf die gegebenen Daten zu, jedoch können diese Klassifikationen bei der Analyse im Großen und Ganzen als guter Leitfaden betrachtet werden. Für beide Gruppen gibt es bestimmte Vorgehensweisen. Im ersten Fall (1.), sofern die Anzahl der experimentellen Einheiten groß ist relativ zur Anzahl der Beobachtungen pro Einheit ist es vorteilhaft eine robuste Varianzschätzung zu verwenden, um auch trotz der fehlenden Korrelation die richtigen Schlüsse über den Regressionsparameter zu ziehen. Für den zweiten Fall (2.) ist es wichtig, dass der Erwartungswert und die Kovarianz annäherungsweise korrekt ist, sonst kann es zu falschen Folgerungen kommen.

7.4 Maximum-Likelihood-Schätzung bei longitudinalen Daten

In Kapitel (5) wurde bereits die Maximum-Likelihood-Schätzung für das verallgemeinerte lineare gemischte Modell vorgestellt. Nun werden für den Fall, dass longitudinale Daten vorliegen, Details zur Maximum-Likelihood-Schätzung erläutert.

Gegeben sei ein statistisches Modell für eine einfache Anwendung, mit Beobachtungen aus zwei verschiedenen Gruppen. (siehe Beispiel (7.1)) Wichtig für die weiteren Berechnungen ist das Modell für die Subjekte der zweiten Gruppe, gegeben durch:

$$y_{i(2)j} = \beta_1 + \beta_2 + \alpha_{i(2)} + \epsilon_{i(2)j} \quad (98)$$

α_i stellen wie immer die Zufallseffekte, sowie β die festen Effekte und ϵ die Fehlerterme dar. Der Indize $i(k)$ bedeutet, dass ein Subjekt in Gruppe k gemeint ist (hier in Gruppe 2) und j kennzeichnet jeweils die Beobachtung in

dem Subjekt. Das Modell (98) kann geschrieben werden als:

$$y_{i(k)j} = \beta_1 + \beta_2 x_{i(k)j} + \alpha_{i(k)} + \epsilon_{i(k)j} \quad (99)$$

mit

$$x_{i(k)j} \begin{cases} 0 & \text{wenn } k = 1 \\ 1 & \text{wenn } k = 2 \end{cases}$$

und $i(k) = 1, \dots, m_k$ mit m_k Beobachtungseinheiten in Gruppe k , sowie $j = 1, \dots, n_i$ mit n_i Beobachtungen in Beobachtungseinheit i .

Unter den beiden Voraussetzungen des folgenden Satzes kann die Maximum-Likelihood-Schätzung bestimmt werden.⁹⁵

Satz 7.1

Für das Modell:

$$y_{i(k)j} = \beta_1 + \beta_2 x_{i(k)j} + \alpha_{i(k)} + \epsilon_{i(k)j}$$

mit den Indizes, wie oben angegeben, kann die Maximum-Likelihood-Schätzung unter folgenden Voraussetzungen erhalten werden:

1. Die Zufallseffekte $\alpha_{i(k)}$ sind unabhängig und normalverteilt mit Erwartungswert 0 und Varianz σ_α^2 . (d.h. für das Bluthochdruckbeispiel am Anfang dieses Kapitels, dass die verschiedenen Subjekte unabhängig voneinander sind und dass die Verteilung der Subjekte die Normalverteilung ist.)
2. Die Fehlerterme $\epsilon_{i(k)j}$ sind ebenfalls unabhängig und normalverteilt mit Erwartungswert 0 und Varianz σ_ϵ^2 . (d.h. die Daten in jeder Beobachtungseinheit sind ebenfalls unabhängig und normalverteilt.)

Beweis. Mit diesen beiden Voraussetzungen ist die Varianz von der Beobachtung $y_{i(k)j}$ die Summe der beiden Varianzen von Fehlerterm und Zufallseffekt:

$$\begin{aligned} \text{var}(y_{i(k)j}) &= \text{var}(\alpha_{i(k)} + \epsilon_{i(k)j}) \\ &= \sigma_\alpha^2 + \sigma_\epsilon^2 \end{aligned}$$

Das gilt deshalb, da die beiden Parameter β_1 und β_2 feste Parameter sind und das $x_{i(k)j}$ eine nicht zufällig erklärende Variable darstellt. Die entsprechende

⁹⁵siehe [(Jones 1993)]

Kovarianz zwischen zwei verschiedenen Beobachtungen (hier j und j') einer Beobachtungseinheit ist gegeben durch:

$$\text{cov}((\alpha_{i(k)} + \epsilon_{i(k)j})(\alpha_{i(k)} + \epsilon_{i(k)j'})) = \sigma_\alpha^2$$

Dies gilt, da die beiden Fehlerterme $\epsilon_{i(k)j}$ und $\epsilon_{i(k)j'}$ korreliert sind und somit die Kovarianz nicht gleich $\sigma_\alpha^2 + \sigma_\epsilon^2$, sondern nur gleich σ_α^2 ist. Natürlich muss die Voraussetzung gegeben sein, dass σ_α^2 für alle Gruppen gleich ist. Die daraus folgende Kovarianzmatrix für die Beobachtungen jedes Subjektes ist gegeben durch:

$$V = \text{Cov}((\alpha_{i(k)} + \epsilon_{i(k)j})(\alpha_{i(k)} + \epsilon_{i(k)j'})) = \begin{pmatrix} \sigma_\alpha^2 + \sigma_\epsilon^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \ddots & \vdots & \\ \vdots & \cdots & \sigma_\alpha^2 + \sigma_\epsilon^2 & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_\epsilon^2 \end{pmatrix}$$

Die Diagonaleinträge kommen dadurch zustande, dass

$$\text{cov}((\alpha_{i(k)} + \epsilon_{i(k)j})(\alpha_{i(k)} + \epsilon_{i(k)j})) = \text{var}(y_i(k)j) = \sigma_\alpha^2 + \sigma_\epsilon^2$$

gilt. Die zugehörige Korrelationsmatrix⁹⁶ ist mit $\rho = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\epsilon^2}$, welcher den Korrelationskoeffizienten, der zwischen 0 und 1 liegt⁹⁷ darstellt, (wird hier auch als klasseninterner Korrelationskoeffizient bezeichnet) gegeben durch:

$$V = \text{Kor}((\alpha_{i(k)} + \epsilon_{i(k)j})(\alpha_{i(k)} + \epsilon_{i(k)j'})) = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & \ddots & \vdots & \\ \vdots & \cdots & 1 & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix}$$

Für den Fall, dass $\rho = 0$ ist, ist die Korrelationsmatrix eine Diagonalmatrix mit Einsen in der Diagonale. Des Weiteren gilt: $\sigma_\alpha^2 = 0$ Das bedeutet, dass

⁹⁶Der Korrelationskoeffizient für zwei jeweils quadratisch integrierbare Zufallsvariablen X und Y sieht so aus:

$$\text{Kor}(X, Y) = \rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$$

⁹⁷siehe: [(Jones 1993)] S.14

die Zufallseffekte $\alpha_{i(k)}$ im Modell (99) nicht auftreten, da $\alpha_{i(k)} \sim N(0, 0) \Rightarrow$ es gibt keine Zufallseffekte. Um die Likelihood-Funktion zu erhalten, sieht man sich zunächst die Dichtefunktion der Beobachtungen von einem Subjekt an. Sofern die Beobachtungen unabhängig sind, ist die Dichtefunktion für alle Beobachtungen gegeben durch das Produkt der einzelnen Dichtefunktionen. Der n -dimensionale Zufallsvektor Y mit der Realisierung $y_{i(k)}$ besitzt die multivariate Normalverteilung $N_n(\mu_k, V)$, wobei μ_k der Erwartungswert der festen Effekte, die 1 ein Spaltenvektor bestehend aus Einsen und V die eben aufgestellte Kovarianzmatrix mit $V \in \Re^{n \times n}$ ist:

$$f(y_{i(k)}) = \frac{1}{(\sqrt{2\pi})^n \det(V)^{\frac{1}{2}}} \exp\left(\frac{-(y_{i(k)} - \mu_k 1)^t V^{-1} (y_{i(k)} - \mu_k 1)}{2}\right)$$

Die Likelihood-Funktion ist, aufgrund der Unabhängigkeit der Daten, das Produkt der einzelnen Dichtefunktionen. Sie wird betrachtet als eine Funktion von den unbekanntem Parametern $\mu_1, \mu_2, \sigma_\alpha^2$ und σ_ϵ^2 :

$$L(\mu_k, \sigma_\alpha^2, \sigma_\epsilon^2) = \prod_{i(k)} \frac{1}{(\sqrt{2\pi})^n \det(V)^{\frac{1}{2}}} \exp\left(\frac{-(y_{i(k)} - \mu_k 1)^t V^{-1} (y_{i(k)} - \mu_k 1)}{2}\right)$$

Die Maximum-Likelihood-Schätzung wird erhalten, indem man Werte der Parameter findet, welche die Funktion (7.4) maximiert. Seien nun $\sigma_\alpha^2 > 0$ und $\sigma_\epsilon^2 > 0$. Da es häufig einfacher ist das Maximum von dem Logarithmus einer Funktion, die aus dem Produkt vieler Funktionen besteht, zu berechnen, sei nun die logarithmierte Likelihood-Funktion gegeben durch: (es wird nun mit der $-2 \ln$ -fachen Likelihood-Funktion gearbeitet)

$$l(\mu_k, \sigma_\alpha^2, \sigma_\epsilon^2) = \sum_{i(k)} [n \ln(2\pi) + \ln \det(V) + (y_{i(k)} - \mu_k 1)^t V^{-1} (y_{i(k)} - \mu_k 1)] \tag{100}$$

In einigen einfachen Fällen, ist es möglich die Loglikelihood-Funktion mit allen Parametern zu differenzieren und dann das Maximum zu ermitteln. In numerischen Berechnungen ist es oft einfacher ein Minimum zu berechnen als ein Maximum, deshalb gehen die folgenden Berechnungen auf das Minimum ein.⁹⁸

Die Minimierung von $-2 \ln$ der Loglikelihood-Funktion ist einfacher zu berechnen, wenn man sich die einzelnen Parameter separat anschaut. Man dif-

⁹⁸siehe: [(Jones 1993)] S.16

ferenziert l somit bzgl. einem Parameter und setzt dann das Ergebnis gleich 0. Die Gleichung wird anschließend für den einen Parameter gelöst und das Ergebnis wieder in die Loglikelihood-Funktion eingesetzt. Es ergibt sich ein neues l , jedoch mit einem Parameter weniger. Die Ableitung von (100) bzgl. des Parameters μ_1 ist gegeben durch:

$$\frac{\partial l}{\partial \mu_1} = \sum_{i(1)} \frac{\partial}{\partial \mu_1} (y_{i(1)} - \mu_1 \mathbf{1})^t V^{-1} (y_{i(1)} - \mu_1 \mathbf{1}) = 0 \quad (101)$$

Die Ableitung wird, wie eben beschrieben, gleich Null gesetzt. Die Summe in Gleichung (101) läuft von 1, ..., m_1 , man summiert folglich nur über den Subjekten in der zugehörigen Gruppe (hier Gruppe 1), da die anderen Parameter aus Gruppe 2 hier noch nicht von Interesse sind.

Dieselbe Gleichung ergibt auch für μ_2 , mit Ersetzen von μ_1 durch μ_2 und $i(1)$ durch $i(2)$.

Durch Umformung der Gleichung (101) (Multiplizieren der Terme) ergibt sich:

$$\sum_{i(1)} \frac{\partial}{\partial \mu_1} (y_{i(1)}^t V^{-1} y_{i(1)} - \mu_1 y_{i(1)}^t V^{-1} \mathbf{1} - \mu_1 \mathbf{1}^t V^{-1} y_{i(1)} + \mu_1^2 \mathbf{1}^t V^{-1} \mathbf{1}) = 0 \quad (102)$$

Die Ableitung sieht wie folgt aus:

$$\sum_{i(1)} (-y_{i(1)}^t V^{-1} \mathbf{1} - \mathbf{1}^t V^{-1} y_{i(1)} + 2\mu_1 \mathbf{1}^t V^{-1} \mathbf{1}) = 0 \quad (103)$$

Die Gleichung wird nach μ_1 aufgelöst. Dabei ist zu beachten, dass aufgrund der Matrixrechnung $y_{i(1)}^t V^{-1} \mathbf{1}$ scalarwertig ist, d.h. dass es gleich dem scalarwertigen Ergebnis von $\mathbf{1}^t V^{-1} y_{i(1)}$ ist. Man bedenke, dass es in Gruppe 1 m_1 Subjekte gibt:

$$\begin{aligned} \sum_{i(1)} (-y_{i(1)}^t V^{-1} \mathbf{1} - \mathbf{1}^t V^{-1} y_{i(1)} + 2\mu_1 \mathbf{1}^t V^{-1} \mathbf{1}) &= 0 \\ \Leftrightarrow 2 \sum_{i(1)} \mu_1 \mathbf{1}^t V^{-1} \mathbf{1} &= 2 \sum_{i(1)} \mathbf{1}^t V^{-1} y_{i(1)} \\ \Leftrightarrow \hat{\mu}_1 &= \frac{2 \mathbf{1}^t V^{-1} \sum_{i(1)} y_{i(1)}}{2 m_1 \mathbf{1}^t V^{-1} \mathbf{1}} \\ \Leftrightarrow \hat{\mu}_1 &= \frac{\mathbf{1}^t V^{-1} \sum_{i(1)} y_{i(1)}}{m_1 \mathbf{1}^t V^{-1} \mathbf{1}} \end{aligned}$$

$\hat{\mu}_1$ in den letzten beiden Gleichungen bedeutet, dass dies eine Schätzung ist. Die Schätzung von μ_1 ist der BLUE („Best linear unbiased estimator“) und ist genau dann die Maximum-Likelihood-Schätzung für μ_1 , wenn die Kovarianz-

matrix $V = \begin{pmatrix} 1 & \cdots & \rho \\ \vdots & \ddots & \vdots \\ \rho & \cdots & 1 \end{pmatrix}$ bekannt ist. Des Weiteren gilt:

$$\begin{aligned} V &= \sigma_\epsilon^2 I + \sigma_\alpha^2 11^t \\ \Leftrightarrow V^{-1} &= \frac{1}{\sigma_\epsilon^2} \left(I - \frac{\sigma_\alpha^2 11^t}{n\sigma_\alpha^2 + \sigma_\epsilon^2} \right) \end{aligned}$$

Nun ist die Schätzung unter Beachten, dass $1^t y_{i(1)} = \sum_{j=1}^n y_{i(1)j}$ und $I1 = 1$, wobei I die $n \times n$ Einheitsmatrix ist, $1^t I = 1^t$ und $1^t 1 = n$:

$$\begin{aligned} \hat{\mu}_1 &= \frac{1^t V^{-1} \sum_{i(1)} y_{i(1)}}{m_1 1^t V^{-1} 1} \\ &= \frac{1^t \frac{1}{\sigma_\epsilon^2} \left(I - \frac{\sigma_\alpha^2 11^t}{n\sigma_\alpha^2 + \sigma_\epsilon^2} \right) \sum_{i(1)} y_{i(1)}}{m_1 1^t \frac{1}{\sigma_\epsilon^2} \left(I - \frac{\sigma_\alpha^2 11^t}{n\sigma_\alpha^2 + \sigma_\epsilon^2} \right) 1} \\ &= \frac{\sum_{i(1)} 1^t y_{i(1)j}}{m_1 1^t 1} \\ &= \frac{\sum_{i=1}^{m_1} \sum_{j=1}^n y_{i(1)j}}{m_1 n} \\ &= \frac{1}{m_1 n} \sum_{i=1}^{m_1} \sum_{j=1}^n y_{i(1)j} \\ &= \bar{y}_{\cdot(1)}. \end{aligned}$$

$\bar{y}_{\cdot(1)}$ stellt das arithmetische Mittel dar. Für μ_2 ergibt sich die gleiche Schätzung, nur dass anstelle des Indizes 1 eine 2 steht.

Um die $-2 \ln$ Likelihood zu erhalten sei:

$$\tilde{y}_{i(k)} = y_{i(k)} - \mu_k 1$$

Damit kann Gleichung (100) folgendermaßen vereinfacht dargestellt werden:

$$l(\sigma_\alpha^2, \sigma_\epsilon^2) = \sum_{i(k)} [n \ln(2\pi) + \ln \det(V) + (\tilde{y}_{i(k)} V^{-1} (\tilde{y}_{i(k)}))] \quad (104)$$

$$= \sum_{i(k)} \left[n \ln(2\pi) + \ln(n\sigma_\alpha^2 + \sigma_\epsilon^2)(\sigma_\epsilon^2)^{n-1} + (\tilde{y}_{i(k)} \frac{1}{\sigma_\epsilon^2} \left(I - \frac{\sigma_\alpha^2 \mathbf{1}\mathbf{1}^t}{n\sigma_\alpha^2 + \sigma_\epsilon^2} \right) (\tilde{y}_{i(k)}) \right]$$

wobei $(n\sigma_\alpha^2 + \sigma_\epsilon^2)(\sigma_\epsilon^2)^{n-1} = \det(V)$. Sei im Folgenden $m := m_1 + m_2$ die komplette Anzahl der Subjekte. Dann kann die $-2 \ln$ Likelihood, unter Auslassen des konstanten Terms $mn \ln(2\pi)$, geschrieben werden als:

$$\begin{aligned} l(\sigma_\alpha^2, \sigma_\epsilon^2) &= m(n-1) \ln \sigma_\epsilon^2 + m \ln(n\sigma_\alpha^2 + \sigma_\epsilon^2) \\ &\quad + \tilde{y}_{i(k)}^t \frac{1}{\sigma_\epsilon^2} I \tilde{y}_{i(k)} - \tilde{y}_{i(k)}^t \frac{\mathbf{1}\mathbf{1}^t \sigma_\alpha^2}{\sigma_\epsilon^2 (n\sigma_\alpha^2 + \sigma_\epsilon^2)} \tilde{y}_{i(k)} \\ &= m(n-1) \ln \sigma_\epsilon^2 + m \ln(n\sigma_\alpha^2 + \sigma_\epsilon^2) + \frac{1}{\sigma_\epsilon^2} \sum_{i(k)} \tilde{y}_{i(k)}^t \tilde{y}_{i(k)} \\ &\quad - \frac{n\sigma_\alpha^2}{\sigma_\epsilon^2 (n\sigma_\alpha^2 + \sigma_\epsilon^2)} \frac{1}{n} \sum_{i(k)} (\tilde{y}_{i(k)}^t \mathbf{1})(\mathbf{1}^t \tilde{y}_{i(k)}) \\ &= m(n-1) \ln \sigma_\epsilon^2 + m \ln(n\sigma_\alpha^2 + \sigma_\epsilon^2) + \frac{1}{\sigma_\epsilon^2} \sum_{i(k)} \tilde{y}_{i(k)}^t \tilde{y}_{i(k)} \\ &\quad - \frac{n\sigma_\alpha^2}{\sigma_\epsilon^2 (n\sigma_\alpha^2 + \sigma_\epsilon^2)} \frac{1}{n} \sum_{i(k)} \tilde{y}_{i(k)}^t \tilde{y}_{i(k)} \\ &= m(n-1) \ln \sigma_\epsilon^2 + m \ln(n\sigma_\alpha^2 + \sigma_\epsilon^2) + \frac{1}{\sigma_\epsilon^2} \sum_{i(k)} \tilde{y}_{i(k)}^t \tilde{y}_{i(k)} \\ &\quad - \frac{n\sigma_\alpha^2}{\sigma_\epsilon^2 (n\sigma_\alpha^2 + \sigma_\epsilon^2)} \frac{1}{n} \sum_{i(k)} (\tilde{y}_{i(k)}^t \mathbf{1})(\mathbf{1}^t \tilde{y}_{i(k)}) \\ &= m(n-1) \ln \sigma_\epsilon^2 + m \ln(n\sigma_\alpha^2 + \sigma_\epsilon^2) + \frac{1}{\sigma_\epsilon^2} T_T - \frac{n\sigma_\alpha^2}{\sigma_\epsilon^2 (n\sigma_\alpha^2 + \sigma_\epsilon^2)} T_B \end{aligned}$$

mit

$$T_T := \sum_{i(k)} \tilde{y}_{i(k)}^t \tilde{y}_{i(k)} = \sum_{i(k)} \tilde{y}_{i(k),j}^2$$

und

$$T_B := \frac{1}{n} \sum_{i(k)} (\tilde{y}_{i(k)}^t \mathbf{1})(\mathbf{1}^t \tilde{y}_{i(k)}) = \frac{1}{n} \sum_{i(k)} (\tilde{y}_{i(k),j})^2$$

Die Ableitungen nach σ_α^2 und σ_ϵ^2 und anschließend gleich Null gesetzt sind

gegeben durch:

$$\frac{\partial l(\sigma_\alpha^2, \sigma_\epsilon^2)}{\partial \sigma_\alpha^2} = n\sigma_\alpha^2 + \sigma_\epsilon^2 - \frac{1}{m}T_B = 0 \quad (105)$$

$$\Leftrightarrow n\sigma_\alpha^2 + \sigma_\epsilon^2 = \frac{1}{m}T_B \quad (106)$$

und

$$\frac{\partial l(\sigma_\alpha^2, \sigma_\epsilon^2)}{\partial \sigma_\epsilon^2} = \sigma_\epsilon^2 - \frac{1}{m(n-1)}T_W = 0 \quad (107)$$

$$\Leftrightarrow \sigma_\epsilon^2 = \frac{1}{m(n-1)}T_W \quad (108)$$

mit $T_W := T_T - T_B$. Mit (106) und (108) ergibt sich:

$$\begin{aligned} n\sigma_\alpha^2 + \sigma_\epsilon^2 &= \frac{1}{m}T_B \\ \Leftrightarrow \sigma_\alpha^2 &= \frac{\frac{1}{m}T_B - \sigma_\epsilon^2}{n} \end{aligned}$$

$$\begin{aligned} \frac{\frac{1}{m}T_B - \sigma_\epsilon^2}{n} &= \frac{T_B}{mn} - \frac{\sigma_\epsilon^2}{n} \\ &= \frac{T_B}{mn} - \frac{T_W}{n(m(n-1))} \\ &= \frac{T_B}{mn} - \frac{T_W}{nm(n-1)} \\ &= \frac{1}{mn} \left(T_B - \frac{T_W}{n-1} \right) \end{aligned}$$

Damit sind die Maximum-Likelihood-Schätzungen für das im Satz genannte Modell gefunden. Allerdings muss beachtet werden, dass dies nur die Schätzungen sind, sofern sie positiv sind. Für den Fall, dass die Schätzungen negativ sind, sind die Ergebnisse keine Maximum-Likelihood-Schätzungen mehr. Dann können die Schätzungen auf folgendem Weg gefunden werden: Man muss den

Wert von σ_ϵ^2 finden, welcher l minimiert. Es wird $n\sigma_\alpha^2 = 0$ gesetzt. Dann ist:

$$\begin{aligned} l(\sigma_\epsilon^2, \sigma_\alpha^2 = 0) &= m(n-1) \ln \sigma_\epsilon^2 + m \ln(n \cdot 0 + \sigma_\epsilon^2) \\ &\quad + \frac{1}{\sigma_\epsilon^2} T_T - \frac{0}{\sigma_\epsilon^2(n \cdot 0 + \sigma_\epsilon^2)} n T_B \\ &= mn \ln \sigma_\epsilon^2 - m \ln \sigma_\epsilon^2 + m \ln \sigma_\epsilon^2 + \frac{1}{\sigma_\epsilon^2} T_T \\ &= mn \ln \sigma_\epsilon^2 + \frac{1}{\sigma_\epsilon^2} T_T \end{aligned}$$

Wenn l abgeleitet und gleich Null gesetzt wird, ergibt sich:

$$\begin{aligned} \frac{\partial l(\sigma_\epsilon^2, \sigma_\alpha^2 = 0)}{\partial \sigma_\epsilon^2} &= mn \frac{1}{\sigma_\epsilon^2} - \frac{1}{\sigma_\epsilon^4} T_T = 0 \\ \Leftrightarrow mn \frac{1}{\sigma_\epsilon^2} &= \frac{1}{\sigma_\epsilon^4} T_T \\ \Leftrightarrow mn &= \frac{\sigma_\epsilon^2}{\sigma_\epsilon^4} T_T \\ \Leftrightarrow mn &= \frac{1}{\sigma_\epsilon^2} T_T \\ \Leftrightarrow \sigma_\epsilon^2 &= \frac{1}{mn} T_T \end{aligned}$$

Es ist für das besagte Modell und den im Satz gegebenen Voraussetzungen somit eine Maximum-Likelihood-Schätzungen gefunden und die Schätzungen (für die Varianzen) sind gegeben durch:

Für $(n-1)T_B \geq T_W$:

$$\begin{aligned} \hat{\sigma}_\epsilon^2 &= \frac{1}{m(n-1)} T_W \\ \hat{\sigma}_\alpha^2 &= \frac{1}{mn} \left(T_B - \frac{T_W}{n-1} \right) \end{aligned}$$

und für $(n-1)T_B < T_W$:

$$\begin{aligned} \hat{\sigma}_\epsilon^2 &= \frac{1}{mn} T_T \\ \hat{\sigma}_\alpha^2 &= 0 \end{aligned}$$

Damit ist der Satz (7.1) bewiesen. □

Die klassische Analyse von longitudinalen Daten beruht auf einem ausge-

gleichen Aufbau, bei dem jedes Subjekt zu gleichen Zeitpunkten gemessen wird und das ohne fehlende Werte. Die Vorgehensweise bei nicht vollständig gegebenen Daten (Daten mit Missing Values) wird im folgenden Unterkapitel dargelegt.

7.5 Missing Values bei longitudinalen Daten

Wie bereits erwähnt, spielen fehlende Daten, bzw. „Missing Values“ bei der Analyse von longitudinalen Daten oft eine bedeutende Rolle.

Bemerkung 7.5

Fehlende Daten können verschiedene Ursachen haben, die folgenden fünf sind die Häufigsten:

- *Fehlerhaftes oder mangelhaftes Untersuchungsdesign*
- *Unaufmerksamkeit des Beobachters*
- *Antwortverweigerung bei einer Befragung*
- *Nicht alle Sekundärdaten sind vorhanden*
- *Codierungs- und Übertragungsfehler bei den Daten*

„missing values“ können die Analyse erheblich erschweren und verfälschen. Beim Auftreten von solchen fehlenden Werten können die normalen, auf vollständigen Datenmaterial beruhenden Analysemethoden nicht mehr ohne Weiteres angewandt werden. Die Auswirkungen der „missing values“ müssen also bei der statistischen Analyse berücksichtigt werden.

Gerade bei biologischen Studien, wie es die Forschungsstudie aus Kapitel (2) eine ist, sind fehlende Werte keinesfalls eine Seltenheit. Leider ist es bei einigen Computerprogrammen der Fall, dass diese nicht mit solchen fehlenden Daten umgehen können und die Daten einfach aus der Analyse eliminiert werden.

Falls fehlende Daten vorliegen, ist es wichtig vor allem zwei Punkte zu beachten.

Zum Einen müssen die Formeln abgeändert werden und zum Anderen muss die Interpretation der Ergebnisse bei solchen unvollständigen Daten berücksichtigt werden.⁹⁹

⁹⁹[(Brunner und Langer 1999)]

Falls die Daten zufällig fehlen, ist es somit wichtig die Formel abzuändern. Es gibt zwei Arten von Ausfallmechanismen, die zu unterscheiden sind. (siehe: [(Runte)]) Einerseits gibt es die zufällig fehlenden Daten, welche auch als „*unsystematisch fehlende Daten*“ bezeichnet werden und andererseits können nicht zufällig fehlende Werte vorliegen, die sogenannten „*systematisch fehlenden Daten*“. Liegen nur unvollständige Daten bei einem nichtparametrischen Modell vor, so ist es notwendig weitere Auswertungen vorzunehmen, um diese am Ende nebeneinander zu stellen und die Resultate der verschiedenen Auswertungen zu vergleichen. Es gibt zwei Vorgehensweisen: Zum Einen können die fehlenden Daten als zufällig fehlend angesehen werden oder diese fehlenden Werte werden durch sogenannte „vernünftige“¹⁰⁰ Werte ersetzt. Das sind Werte, welche im bestmöglichen oder schlechtmöglichen Fall vorliegen würden.

Bemerkung 7.6

Es gibt folgende drei Wege die fehlenden Daten zu ergänzen:

- *Ergänzung der fehlenden Werte durch die entsprechenden Zeilen- bzw. (Spalten-)Durchschnittswerte der Datenmatrizen.*
- *Ergänzung der fehlenden Werte durch Werte, welche mit der Regression geschätzt wurden.*
- *Die fehlenden Daten werden aufgrund bekannter Eigenschaften der Datenquelle geschätzt.*

Für die Behandlung von fehlenden Daten sind die zufällig und nicht zufälligen Daten noch weiter zu erläutern:

Die zwei Ausfallmechanismen, also die systematisch und die unsystematisch fehlenden Daten wurden von [(Rubin 1976)] eingeführt.

Definition 7.3 (Eigenschaften von Missing Values)

Nach [(Rubin 1976)] sind die fehlenden Daten der Eigenschaft MAR=“Missing At Random“ zuzuordnen, wenn das Fehlen der Daten nicht mit der Ausprägung des Merkmals selbst korreliert.

Ist das Fehlen der Daten unabhängig von den Ausprägungen anderer Merkmale im selben Objekt, haben die Daten die Eigenschaft OAR=“Observed At Random“

¹⁰⁰[(Brunner und Langer 1999)]

Wenn sowohl MAR, als auch OAR zutreffen, gilt MCAR=“Missing Completely At Random“.

Der MCAR bedeutet auch, dass die Existenz der Daten weder von fehlenden noch von existierenden Merkmalen abhängt und somit das Fehlen der Daten keinerlei Relation zur Existenz oder zu den Ausprägungen anderer Daten hat. In Tabelle (3) sind die Voraussetzungen der unsystematisch fehlenden Ausfallmechanismen, also der zufällig fehlenden Daten dargestellt. Im Folgenden wird

| | |
|------|--|
| MAR | Antwortrate unabhängig von der Ausprägung des Merkmals |
| OAR | Antwortrate unabhängig von der Ausprägung anderer Merkmale |
| MCAR | MAR und OAR treffen zu |

Tabelle 3: Unsystematisch fehlende Daten

kurz erklärt, wie die Vorgehensweise bei solchen fehlenden Daten aussieht: Es gibt genau fünf verschiedene Verfahren zur Behandlung von fehlenden Daten, welche im Folgenden kurz aufgezählt und in [(Runte)] nachgelesen und vertieft werden können.

Als allerersten Schritt, beim Auftreten von fehlenden Daten, muss eine Strukturanalyse der Daten vorgenommen werden, um Erkenntnisse über die Abhängigkeit der Struktur der Missing Values von anderen Merkmalen zu erhalten. Es gibt drei verschiedene Ansätze:¹⁰¹

1. **Die deskriptive Analyse:** Bei dieser Analyse lassen sich die Maße der fehlenden Daten berechnen. Es steht hier im Mittelpunkt möglichst aussagekräftige Kennzahlen für die fehlenden Daten zu ermitteln. Die Aufgabe der deskriptiven Analyse ist es das Verhältnis von fehlenden zu existierenden Werten innerhalb einer Matrix, der sogenannten Indikatormatrix, darzustellen.
2. **Die explorative Analyse:** Hier wird nach Zusammenhängen innerhalb der unvollständigen Datenmatrix gesucht. Das Ziel der Untersuchung ist es Abhängigkeitsstrukturen innerhalb der Daten aufzudecken und somit fehlende Werte zu entdecken. Auch bei dieser Analysemethode spielt die Indikatormatrix eine wichtige Rolle.
3. **Die induktive Analyse:** Hier ist die Vorgehensweise etwas anders gestaltet als bei den beiden ersten Analysemethoden, es werden statistische

¹⁰¹siehe[(Bankhofer 1995)]

Tests angesetzt. Es werden im Bezug auf die Missing Values verschiedene Hypothesen aufgestellt und anschließend getestet. Außerdem besteht bei diesem Analyseverfahren auch die Möglichkeit eine Abhängigkeit der Missing Values von den Ausprägungen eben dieser Daten zu berechnen.

Wenn eine Strukturanalyse durchgeführt wurde, gibt es folgende verschiedene Verfahren zur Behandlung von Missing Values: Eliminierungsverfahren, Imputationsverfahren, Parameterschätzverfahren, Multivariate Verfahren oder Sensitivitätsanalysen.¹⁰² Diese werden hier allerdings nicht weiter von Bedeutung sein, da explizit in der Analyse im folgenden Kapitel nicht auf Missing Values eingegangen wird. Die Prozedur, die für die Auswertung verwendet wird, berücksichtigt automatisch Missing Values, diese gehen also mit in die Analyse ein, wodurch es nicht zu einer Verfälschung der Ergebnisse kommt.

¹⁰²siehe [(Bankhofer 1995)]

8 Auswertung der Forschungsstudie

Nachdem in der Arbeit dargestellt wurde, wie die Analyse mit longitudinalen Daten theoretisch abläuft, kann nun mit der Auswertung der Forschungsstudie des visuellen Systems, welche in Kapitel (2) vorgestellt wurde, begonnen werden. Für die Analyse wird im Folgenden das Statistikprogramm SAS verwendet, welches bereits im vorhergehenden Kapitel (7) bei dem Beispiel (7.1) der Analyse von longitudinalen Daten angewandt wurde. Für die Analyse wird das verallgemeinerte lineare gemischte Modell verwendet, welches bereits in den Kapiteln (5) und (6) genau behandelt wurde. Das GLMM ist das am besten geeignete Modell für die Daten der Forschungsstudie. Zu diesem Schluss kommt man unter Anderem deshalb, weil das verallgemeinerte lineare gemischte Modell, wie im fünften Kapitel bereits beschrieben wurde, nicht nur feste Effekte, sondern auch Zufallseffekte berücksichtigt. Des Weiteren ist das Modell sehr robust gegenüber Missing Values.

Die Daten aus der Studie wurden bereits anhand des GLMMs mit SPSS ausgewertet. Nun besteht die Aufgabe, die Auswertung erneut, dieses Mal allerdings mit dem Statistikprogramm SAS durchzuführen. Bevor die Analyse durchgeführt werden kann, muss zunächst überlegt werden, welche feste und zufällige Effekte, sowie abhängige Variablen bei der Analyse der Forschungsstudie vorliegen. Ein möglicher Zufallseffekt ist durch die Stichprobenauswahl, die hier die gemessenen Ganglienzellen der Netzhäute der Mäuse sind, gegeben.

In der folgenden Analyse werden die SAR-Werte, die Lichtintensität, und die Temperatur als feste Effekte definiert. Die Messreihenfolge, die in der Studie als weiterer fester Effekt gewählt wurde, wird in dieser Analyse außer Acht gelassen. Es wird folgedessen nicht zwischen den Messreihen, bzw Intensitätsreihen K, T3 und T6 unterschieden. Aus diesem Grund müssen die Ergebnisse, die in den folgenden Abschnitten dargelegt werden, mit den Ergebnissen der Forschungsstudie nicht komplett übereinstimmen.

Die Unit-Nummer¹⁰³ ist, wie oben schon erwähnt, der ausgewählte Zufallseffekt. Als abhängige Variablen, also diejenigen Variablen, die während des Versuches gemessen wurden, sind die Latenzen und Antwortraten, bzw. deren

¹⁰³Die Unit Nummer ist die gemessene Ganglienzelle

Quotienten und Differenzen mit der Kontrollmessreihe, gegeben.

In einigen Fällen kommt es vor, dass nur das Zusammenspiel von mehreren Faktoren bestimmte Effekte hervorruft. Aus diesem Grund sind bei der Analyse auch die sogenannten Interaktionen von großer Bedeutung. Eine mögliche Interaktion zwischen den festen Effekten ist die Interaktion zwischen dem SAR-Wert und der Lichtintensität. Diese Interaktion, sowie die Interaktion von Intensität und Temperatur werden in der Analyse berücksichtigt.

Wie bereits am Anfang des Kapitels (5) erwähnt, können die beiden Prozeduren „GLIMMIX“ und „NLMIXED“ für die Auswertung der Daten verwendet werden. Die Prozedur „GLIMMIX“ muss gesondert in SAS eingebunden werden. Diese Prozedur ist noch sehr neu und erst seit der SAS Version 9.1 verfügbar. Die Prozedur „NLMIXED“ hingegen ist in SAS bereits vorinstalliert. Beide Prozeduren werten verallgemeinerte lineare gemischte Modelle aus, jedoch ist die NLMIXED-Prozedur, was die Klasse der Modelle angeht, etwas beschränkter als die GLIMMIX-Prozedur. Im Folgenden werden daher die Daten der Forschungsstudie anhand der „GLIMMIX“-Prozedur ausgewertet. Diese Prozedur beschäftigt sich unter anderem mit der Approximation der Daten, welche in Kapitel (5) bereits erläutert wurde. Die beiden Quasi-Likelihood-Schätzungen (MQL und PQL) sind in der „GLIMMIX“-Prozedur implementiert. Es können verschiedene Typen der Quasi-Likelihood, z.B. der Quasi-Likelihood-Typ PQL mit der Maximum-Likelihood oder mit der Restricted-Maximum-Likelihood Methode (auf diese wurde in der Arbeit nicht eingegangen) durch die Anweisung „method=“¹⁰⁴ gewählt werden. Ein Vorteil, den die GLIMMIX-Prozedur mit sich bringt ist, dass Missing Values beachtet werden, was nicht in allen Prozeduren von SAS der Fall ist. Die Daten aus der Forschungsstudie enthalten viele fehlenden Daten, daher ist es wichtig, dass diese richtig berücksichtigt werden.

Die Prozedur weist große Ähnlichkeiten zu der Prozedur „MIXED“ auf, sie enthält ebenfalls die Statements „CLASS, MODEL und RANDOM“. Ein einfacher SAS Code für die „GLIMMIX“-Prozedur sieht folgendermaßen aus:

```
PROC GLIMMIX DATA=Name des Datensatzes;  
CLASS Klassifikationsvariable;  
MODEL Antwortvariable=erklärende Variable;  
RANDOM INTERCEPT / SUBJECT= Zufallsvariable;
```

¹⁰⁴siehe auch Abschnitt (8.1.1)

RUN;

Dabei gibt die „CLASS“-Anweisung die Klassifikationsvariablen, also die qualitativen Variablen der Analyse an. In der „MODEL“-Anweisung werden die Beobachtungsvariable, sowie die erklärenden Variablen mit all ihren festen Effekten definiert und die Zufallseffekte werden durch den „RANDOM“-Befehl spezifiziert.

Im folgenden Abschnitt wird die Analyse der Daten aus der Forschungsstudie des visuellen Systems, getrennt nach den verschiedenen Frequenzen, erfolgen.

8.1 Auswertung anhand der SAS-Prozedur „GLIMMIX“

Wie bereits erwähnt, werden die Auswertungen jeweils einzeln für die Frequenzen UMTS Standard, GSM 900 MHz und GSM 1800 MHz durchgeführt. Es wurden jeweils mehrere Versuche angesetzt, bei UMTS Standard waren es 80 Versuche, von denen nach dem Verwerfen von einigen (hier waren es 38 Experimente, die nicht mit in die Analyse gingen) Experimenten, aufgrund zu schlechter Messbedingungen noch 42 Versuche übrig blieben, bei GSM 900 MHz waren es nach Elimination einiger Versuche noch 88, die mit in die Auswertung gingen und bei GSM 1800 MHz blieben 49 Versuche übrig.

Bevor die Auswertungen für UMTS Standard, GSM 900 und GSM 1800 anhand des GLMMs stattfinden, werden zunächst die Mittelwerte und Standardabweichungen für UMTS Standard in einer Tabelle, sowie die Temperaturkurve dargestellt. Bei GSM 900 und GSM 1800 wird lediglich die Temperaturkurve aufgezeichnet.

8.1.1 Auswertung der UMTS Standard Versuche

Wie gerade schon erwähnt werden zunächst die Mittelwerte und Standardabweichungen der Temperatur für den UMTS Standard dargestellt. Wie in der Beschreibung der Forschungsstudie am Anfang dieser Arbeit dargelegt, ist es wichtig die Temperatur auf einem möglichst konstanten Level zu halten. Es ist von Bedeutung direkte Effekte der SAR-Werte von indirekten Temperatureffekten zu unterscheiden. Aus diesem Grund ist es wichtig sich zunächst einen Überblick über den Verlauf der Temperatur vor der Befeldung, sowie während und auch nach der Befeldung zu verschaffen. In Tabelle (4) werden die Mittelwerte \pm der Standardabweichungen (diese wurden mit der SAS-Prozedur

„MEANS“ berechnet) für den UMTS Standard zu den verschiedenen Zeitpunkten vor (C), während (B1 und B2) und nach (N1-N4) der Befeldung aufgelistet. Die Temperaturen sind in Grad Celsius angegeben, SD ist die Standardabweichung. Es ist zu erkennen, dass die Temperatur immer nahe bei 35 °C liegt. Bei genauerem Hinsehen, ist jedoch zu erkennen, dass bei einem SAR-Wert von 20 W/kg, also einem sehr großer SAR-Wert, die Temperatur bei der ersten und zweiten Messung während der Befeldung (Temp B1 und Temp B2) und bei der ersten Messung nach Befeldung (Temp N1) die Temperatur etwas absinkt, bevor sie dann wieder leicht ansteigt. Die Schwankungen lagen um $\pm 0,3^{\circ}\text{C}$. Da die Messungen während der Befeldungspause vorgenommen wurden, kann man sagen, dass jeweils nach Abschalten der Befeldung die Temperatur vorübergehend gesunken ist. Da eine konstante Temperatur wichtig ist, kann man hier vermuten, dass große SAR-Werte schädlich sein können, da sich die Temperatur bei diesem Wert am meisten verändert. Allerdings ist zu erwähnen, dass ein Handy nicht annäherungsweise SAR-Werte von 20 W/kg besitzt.

Wie gerade schon erwähnt, wird die Prozedur MEANS verwendet, die folgendermaßen aussieht:

```
PROC SORT DATA=WORK.TEMPUMTS;  
BY SAR;  
RUN;  
TITLE 'Mittelwerte und Standardabweichung der Temperatur bei UMTS  
Standard';  
PROC MEANS DATA=WORK.TEMPUMTS;  
VAR Temp_C Temp_B1 Temp_B2 Temp_N1 Temp_N2 Temp_N3  
Temp_N4;  
CLASS SAR;  
RUN;
```

In der folgenden Abbildung (5) ist die Temperaturkurve bei UMTS Standard zu sehen. Es ist deutlich zu erkennen, dass bei einem SAR-Wert von 20 W/kg die Temperatur bei den Messungen TempB1 leicht und bei der Messung TempB2 um so stärker (folglich bei den beiden Messungen, die während der Pausen der Befeldung gemacht wurden) abfällt.

Nachdem die Temperaturkurve für UMTS Standard bestimmt wurde, ist es nun noch von Interesse, wie sich die Quotienten und Differenzen der Latenzen und Antwortraten der Forschungsstudie im Mittel verhalten und ob sich die

| | | | | | |
|--------|-----------|-----------|-----------|-----------|-----------|
| SAR | 0 W/kg | 0,02 W/kg | 0,2 W/kg | 2 W/kg | 20 W/kg |
| TempC | 34.907725 | 34.946479 | 34.916887 | 34.833878 | 34.913835 |
| SD | 0.110419 | 0.102346 | 0.210533 | 0.173392 | 0.127998 |
| TempB1 | 34.971711 | 34.946903 | 34.957009 | 34.931327 | 34.904189 |
| SD | 0.135975 | 0.078786 | 0.112189 | 0.078497 | 0.080381 |
| TempB2 | 34.951098 | 34.917917 | 34.927649 | 34.937170 | 34.861595 |
| SD | 0.100509 | 0.047195 | 0.076112 | 0.022564 | 0.097687 |
| TempN1 | 34.915799 | 34.941883 | 34.938590 | 34.924982 | 34.891133 |
| SD | 0.048612 | 0.088807 | 0.081051 | 0.035442 | 0.071403 |
| TempN2 | 34.938283 | 34.930798 | 34.934263 | 34.937574 | 34.925652 |
| SD | 0.055820 | 0.073695 | 0.077351 | 0.027474 | 0.056180 |
| TempN3 | 34.912198 | 34.931113 | 34.898994 | 34.932323 | 34.941876 |
| SD | 0.033231 | 0.042430 | 0.125028 | 0.033049 | 0.059778 |
| TempN4 | 34.912198 | 34.931113 | 34.898994 | 34.932323 | 34.941876 |
| SD | 0.033231 | 0.042430 | 0.125028 | 0.033049 | 0.059778 |

Tabelle 4: Temperaturen bei UMTS Standard

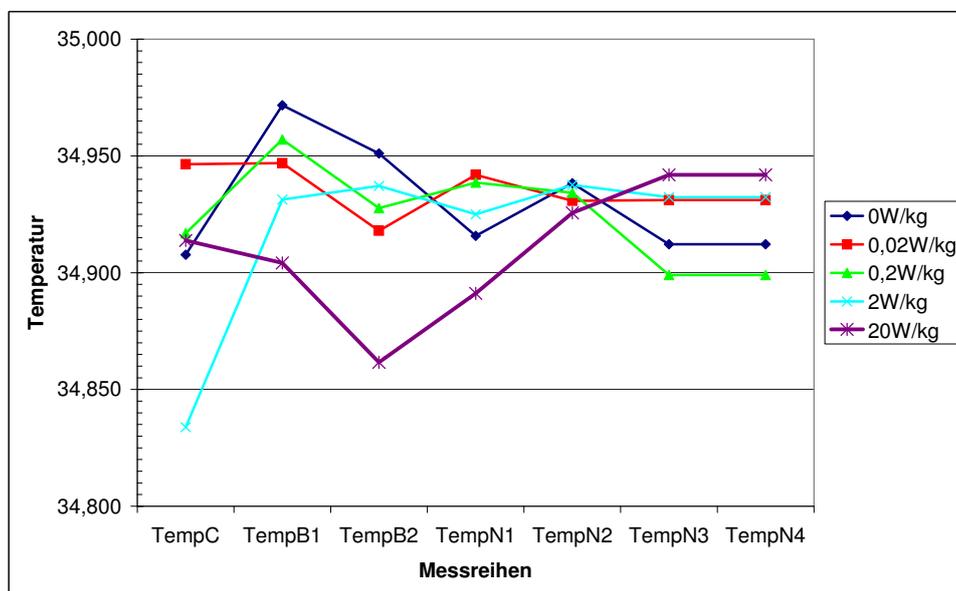


Abbildung 5: Temperaturkurve bei UMTS Standard

Mittelwerte von den Mittelwerten der Scheinbefeldung (SAR=0) signifikant unterscheiden. Dazu wurde der Dunnet-Test, der in der SAS-Prozedur GLM implementiert ist, mit einem Alpha-Niveau von 0,05 durchgeführt. Da es im Rahmen dieser Arbeit aufgrund der großen Anzahl nicht möglich ist, für jede On und Off Zelle die Kurven der Quotienten und Differenzen für die verschiedenen SAR-Werte (0, 0,02, 0,2, 2, 20) darzustellen, sowie die Signifikanz der Befeldungsgruppen (SAR=0,02; 0,2; 2 und 20) zur Kontrollgruppe SAR=0 anhand der unterschiedlichen Intensitäten aufzuführen, wird dieser Punkt sehr minimal dargestellt.

In der folgenden Abbildung (6)

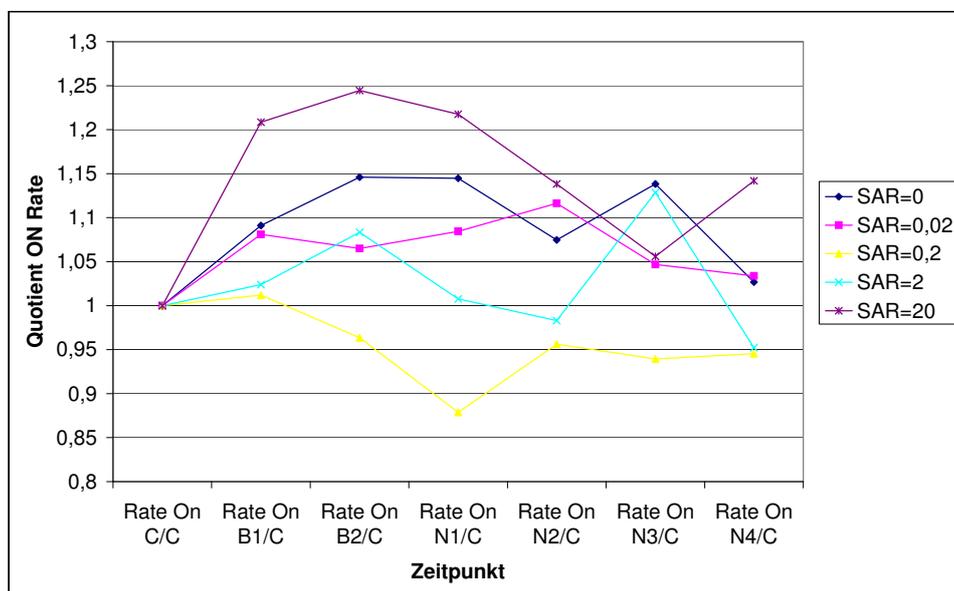


Abbildung 6: ON Antwortrate des Quotienten bei UMTS Standard bei 16lx

wird der durchschnittliche Quotient der On Antwortrate bei einer Intensität von 16 lx dargestellt. Dabei ist anzumerken, dass anhand der Auswertung in der Forschungsstudie herausgefunden wurde, dass bei dieser Intensität ein signifikanter Unterschied zur Kontrollgruppe bei einem SAR-Wert von 0,2 W/kg und dem Quotienten N1 (erste Messung nach der Befeldung) herauskam. Der Dunnet-Test, der mit SAS durchgeführt wurde, zeigte bei einem SAR-Wert von 0,2 W/kg und dem Quotienten N1 keine Signifikanzen.¹⁰⁵ Es ist möglich, dass diese Unterschiede aus dem Grund auftreten, weil die beiden Messreihen

¹⁰⁵Ein Ausschnitt des Outputs zum Quellcode des Dunnet-Tests ist in Anhang (C) dargelegt.

T3 und T6 in dieser Arbeit nicht getrennt voneinander ausgewertet wurden.

Die SAS Anweisungen sahen folgendermaßen aus:

```
PROC GLM DATA=SASUSER.QUOUMTS16lx;
CLASS SAR;
MODEL Rate_ON_C Rate_ON_B1_C Rate_ON_B2_C
Rate_ON_N1_C Rate_ON_N2_C
Rate_ON_N3_C Rate_ON_N4_C=SAR;
MEANS SAR/DUNNETT ('0') ALPHA=0.05;
RUN; QUIT;
```

('0') nach der Anweisung Dunnett bedeutet, dass der Vergleich zur Kontrolle SAR=0 durchgeführt wurde. Bei den multiplen Mittelwertvergleichen, welche mit SAS bei UMTS Standard durchgeführt wurden, zeigten sich keine bestimmten Muster, so dass der Dunnett-Test zeigt, dass keine großen Veränderungen bei Antwortraten und Latenz der unterschiedlichen SAR-Werten vorliegen. Ausserdem waren nur schwach signifikante Unterschiede zu erkennen, bei einem Niveau von 0,01 gab es keine signifikanten Unterschiede mehr.

Nachdem zunächst die Temperaturkurve und der Quotient der ON Rate dargestellt, sowie ein Signifikanztest von den UMTS Standard Versuchen durchgeführt wurde, wird nun die eigentliche Analyse erfolgen. Es wird die folgende GLIMMIX-Prozedur verwendet:

```
PROC GLIMMIX DATA=SASUSER.UMTS METHOD=MSPL IC=Q;
CLASS UNIT#;
MODEL Quotienten und Differenzen der Beobachtungen der verschiedenen
Latenzen und Antwortraten=SAR Intensität Temperatur SAR*Intensität In-
tensität*Temperatur Temperatur*SAR;
RANDOM INTERCEPT / SUBJECT= UNIT#;
RUN;
```

In der Prozedur bedeutet **METHOD=MSPL**, dass die PQL (mit der Maximum -Likelihood-Schätzung) verwendet wird. Diese wurde bereits in Kapitel (5) bearbeitet. Das **IC=Q** Statement bedeutet, wie bereits erwähnt, dass die Berechnung der verschiedenen Informationskriterien, wie z.B. Akaike's und Bayesian's, mit in die Analyse eingeht. Diese Prozedur wertet, wie schon erwähnt, GLMMs aus und ist relativ resistent gegenüber den fehlenden Werten, die in den auszuwertenden Datensätzen der Studie zu genüge vorhanden waren. Eine konkrete SAS Anweisung für die Daten der Studie, sowie dessen

Output ist in Anhang (D) zu finden. In den folgenden Abbildungen (5), (6), (7) und (8) sind die Ergebnisse der Signifikanztests der festen Effekte SAR, Intensität, Temperatur, sowie der beiden Interaktionen SAR*Intensität und Intensität*Temperatur getrennt nach Quotient und Differenz dargestellt. In der GLIMMIX-Prozedur wird zum Testen der festen Effekte die F-Statistik verwendet. Die Werte in den Tabellen geben jeweils den p-Wert $Pr > F$ an. Es wurden auch die nicht signifikanten p-Werte aufgenommen. In den Tabellen sind signifikante Werte bis 0,05 dick gedruckt. Die Auswertungen beziehen sich auf die beiden Messungen B1 und B2, die während der Befeldung in den Befeldungspausen aufgenommen wurden. Auf die Nacheffekte (N1-N4) wird bei UMTS Standard, wie auch in den folgenden Abschnitten bei GSM 900 und GSM 1800 nicht eingegangen.

| QUOTIENT B1 | Rate ON | Latenz ON | Rate OFF | Latenz OFF |
|-----------------------|---------|-----------|----------|------------|
| SAR | 0,4036 | 0,9203 | 0,9031 | 0,0612 |
| Intensität | 0,3964 | 0,8902 | 0,6738 | 0,7707 |
| Temperatur | 0,1034 | 0,9031 | 0,9179 | 0,3570 |
| SAR*Intensität | 0,2320 | 0,4388 | 0,0821 | 0,2035 |
| Intensität*Temperatur | 0,3951 | 0,8898 | 0,6815 | 0,7707 |

Tabelle 5: Signifikanzen des Quotienten B1 bei UMTS Standard

| QUOTIENT B2 | Rate ON | Latenz ON | Rate OFF | Latenz OFF |
|-----------------------|---------------|-----------|---------------|------------|
| SAR | 0,5641 | 0,4782 | 0,7133 | 0,2106 |
| Intensität | 0,0819 | 0,8948 | 0,1838 | 0,7297 |
| Temperatur | 0,5299 | 0,9783 | 0,2312 | 0,6556 |
| SAR*Intensität | 0,0231 | 0,8663 | 0,0368 | 0,9489 |
| Intensität*Temperatur | 0,0820 | 0,8957 | 0,1860 | 0,7310 |

Tabelle 6: Signifikanzen des Quotienten B2 bei UMTS Standard

| DIFFERENZ B1 | Rate ON | Latenz ON | Rate OFF | Latenz OFF |
|-----------------------|---------|-----------|----------|------------|
| SAR | 0,8571 | 0,9868 | 0,3892 | 0,8431 |
| Intensität | 0,9330 | 0,4784 | 0,5772 | 0,7119 |
| Temperatur | 0,1668 | 0,7255 | 0,8978 | 0,8019 |
| SAR*Intensität | 0,1857 | 0,4335 | 0,5813 | 0,4947 |
| Intensität*Temperatur | 0,9305 | 0,4769 | 0,5831 | 0,7117 |

Tabelle 7: Signifikanzen der Differenzen B1 bei UMTS Standard

| DIFFERENZ B2 | Rate ON | Latenz ON | Rate OFF | Latenz OFF |
|-----------------------|---------------|-----------|----------|------------|
| SAR | 0,2107 | 0,6363 | 0,3944 | 0,5047 |
| Intensität | 0,2575 | 0,6978 | 0,9530 | 0,8792 |
| Temperatur | 0,5964 | 0,4897 | 0,8582 | 0,1462 |
| SAR*Intensität | 0,0094 | 0,9511 | 0,9805 | 0,1647 |
| Intensität*Temperatur | 0,2573 | 0,6981 | 0,9497 | 0,8769 |

Tabelle 8: Signifikanzen der Differenzen B2 bei UMTS Standard

Bei der Analyse kam bei den Quotienten und Differenzen B1 und B2 kein signifikanter Effekt für den SAR-Wert heraus. Lediglich bei 2 Werten gab es einen schwach signifikanten Unterschied, (bei der Interaktion SAR*Intensität in der Tabelle Quotient B2) sowie ein signifikanter Unterschied bei der Differenz B2 (auch wieder bei der Interaktion SAR*Intensität). Auffällig bei den Ergebnissen der Analyse ist folglich, dass nur die Interaktion SAR*Intensität zu einem signifikanten Effekt, jedoch nur in seltenen Fällen, führt.

8.1.2 Auswertung der GSM 900 MHz Versuche

Vor der Analyse wird zunächst in Abbildung (7) wieder die Temperaturkurve,

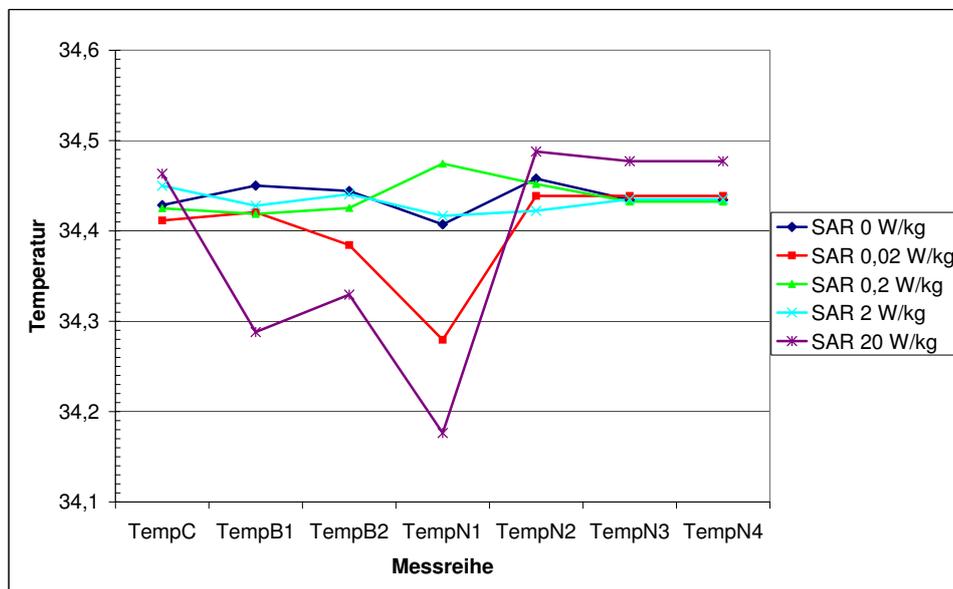


Abbildung 7: Temperaturkurve bei GSM 900

die über der gesamten Versuchsdauer gemessen wurde, dargestellt. Es erweisen

sich ähnliche Schwankungen (geringer Temperaturabfall nach der Befeldung zum Zeitpunkt (N1)) von $\pm 0,3^\circ\text{C}$ wie bei UMTS Standard. Zur Berechnung der Mittelwerte wurde die folgende Prozedur verwendet:

```
PROC SORT DATA=WORK.TEMPGSM900;  
BY SAR;  
RUN;  
PROC MEANS DATA=WORK.TEMPGSM900;  
VAR Temp_C Temp_B1 Temp_B2 Temp_N1 Temp_N2 Temp_N3  
Temp_N4;  
CLASS SAR;  
RUN;
```

Bei GSM 900 wurde ebenfalls ein multipler Mittelwertvergleich anhand des Dunnett-Tests (mit SAS) durchgeführt. Dabei ergab sich z.B. anhand der Daten für die OFF Antwortrate der Differenz bei einer Intensität von 16 lx folgendes Ergebnis: Es wurde ein signifikanter Unterschied bei einem Alpha-Niveau von 0,05 beim SAR-Wert von 0,02 und zum Zeitpunkt B1 herausgefunden.¹⁰⁶ Ein Niveau von 0,01 führte zu keinen signifikanten Unterschieden. Auf die anderen Ergebnisse des Dunnett-Tests wird nicht weiter eingegangen, denn auch bei GSM 900 zeigten sich keine bestimmten Muster, z.B. dass immer bei einem hohen SAR-Wert signifikante Unterschiede bei allen Intensitäten vorhanden waren.

Auch bei GSM 900 wurden die Daten anhand der GLIMMIX-Prozedur (siehe Abschnitt (8.1.1)) ausgewertet. Der Output für die Glimmix-Prozedur ist anhand des Beispiels des Quotienten der On Antwortrate zum Zeitpunkt B1 in Anhang (D) aufgeführt. In den folgenden vier Tabellen (9), (10), (11) und (12) werden, wie auch schon bei UMTS Standard die P-Werte für die F-Statistiken für GSM 900 angegeben. Wieder sind Werte kleiner 0,05 zur besseren Übersicht für signifikante und hochsignifikante Werte dick gedruckt.

Wenn die dick gedruckten, signifikanten p-Werte der Tabellen (9) und (10) mit (11) und (12) verglichen werden, fällt als erstes auf, dass diese für Latenzen und Raten der Quotienten und Differenzen nicht immer übereinstimmen. Das liegt schlicht daran, dass mit den Quotienten und Differenzen unterschiedliche Aspekte der Ganglienzellantworten gemessen werden. Wie schon erwähnt ist der Quotient die relative Änderung und die Differenz die absolute

¹⁰⁶siehe. Anhang(C)

| QUOTIENT B1 | Rate ON | Latenz ON | Rate OFF | Latenz OFF |
|-----------------------|---------------|-----------|---------------|------------|
| SAR | <,0001 | 0,8192 | <,0001 | 0,7756 |
| Intensität | 0,0375 | 0,6838 | 0,0356 | 0,0984 |
| Temperatur | 0,0005 | 0,7285 | 0,0007 | 0,5908 |
| SAR*Intensität | 0,0218 | 0,5891 | 0,0640 | 0,2368 |
| Intensität*Temperatur | 0,0372 | 0,6810 | 0,0347 | 0,0985 |

Tabelle 9: Signifikanzen des Quotienten B1 bei GSM 900

| QUOTIENT B2 | Rate ON | Latenz ON | Rate OFF | Latenz OFF |
|-----------------------|---------------|---------------|---------------|------------|
| SAR | <,0001 | 0,8561 | 0,0079 | 0,8356 |
| Intensität | 0,2290 | 0,8564 | 0,2766 | 0,7013 |
| Temperatur | 0,2090 | 0,8681 | 0,5235 | 0,4601 |
| SAR*Intensität | 0,0056 | 0,0147 | 0,0096 | 0,8721 |
| Intensität*Temperatur | 0,2277 | 0,8560 | 0,2728 | 0,7011 |

Tabelle 10: Signifikanzen des Quotienten B2 bei GSM 900

| DIFFERENZ B1 | Rate ON | Latenz ON | Rate OFF | Latenz OFF |
|-----------------------|---------------|-----------|---------------|------------|
| SAR | <,0001 | 0,7413 | 0,0006 | 0,3844 |
| Intensität | 0,0664 | 0,6508 | 0,2106 | 0,9765 |
| Temperatur | 0,0222 | 0,4678 | 0,1042 | 0,5885 |
| SAR*Intensität | 0,0039 | 0,6274 | 0,0092 | 0,8550 |
| Intensität*Temperatur | 0,0659 | 0,6441 | 0,2056 | 0,9856 |

Tabelle 11: Signifikanzen der Differenzen B1 bei GSM 900

| DIFFERENZ B2 | Rate ON | Latenz ON | Rate OFF | Latenz OFF |
|-----------------------|---------------|---------------|---------------|------------|
| SAR | <,0001 | 0,2670 | 0,0004 | 0,1956 |
| Intensität | 0,0664 | 0,9784 | 0,4678 | 0,7834 |
| Temperatur | 0,0222 | 0,9464 | 0,0079 | 0,5013 |
| SAR*Intensität | 0,0039 | 0,0133 | 0,0920 | 0,7422 |
| Intensität*Temperatur | 0,0659 | 0,9805 | 0,4621 | 0,7801 |

Tabelle 12: Signifikanzen der Differenzen B2 bei GSM 900

Änderung zur Kontrolle. Auffällig bei GSM 900 sind die teilweise hochsignifikanten p-Werte beim Haupteffekt SAR bei Quotienten und Differenzen der Antwortrate, während bei den Latenzen keine signifikanten Unterschiede vorliegen. Außerdem wurden einige signifikanten bis hochsignifikanten Werte bei der Temperatur (bei der Antwortrate), sowie 2 schwach signifikante p-Werte bei der Intensität (wieder nur bei der Antwortrate) gefunden. Ein signifikanter Effekt, der sowohl Rate, wie auch Latenz trifft ist die Interaktion SAR*Intensität.

8.1.3 Auswertung der GSM 1800 MHz Versuche

Auch bei GSM 1800 wurde die Means-Prozedur für die Berechnung der Mittelwerte der Temperatur verwendet:

```
PROC SORT DATA=WORK.TEMPGSM1800;  
BY SAR;  
RUN;  
PROC MEANS DATA=WORK.TEMPGSM1800;  
VAR Temp_C Temp_B1 Temp_B2 Temp_N1 Temp_N2 Temp_N3  
Temp_N4;  
CLASS SAR;  
RUN;
```

Es gab wieder nur geringe Temperaturschwankungen von $\pm 0,3^{\circ}\text{C}$. Auch hier ist allerdings der Effekt, wie bereits bei UMTS und GSM 900 zu erkennen, dass die Temperatur während der Befeldung bei einem SAR-Wert von 20 W/kg abfällt. Ebenfalls sehr auffällig ist, dass bei der Scheinbefeldung (SAR=0) die Temperatur relativ stark abfällt. Dies ist in Abbildung (8) zu erkennen.

Wie schon bei UMTS Standard und GSM 900 wurde auch bei GSM 1800 ein Dunnett-Test mit SAS durchgeführt. Bei der ON Rate der Differenz bei 16 lx ergaben sich 2 signifikante Unterschiede zum Alpha-Niveau 0,05 bei einem SAR-Wert von 0,02 W/kg und zu den Zeitpunkten N1 und N4.¹⁰⁷ Bei Verringerung des Alpha-Niveaus auf 0,01 lagen keine signifikanten Unterschiede mehr vor. Genauso wie bei UMTS Standard und GSM 900 lassen sich bei GSM 1800 anhand des multiplen Mittelwertvergleiches ebenfalls keine Regelmäßigkeiten, die mit einem bestimmten SAR-Wert eingehen, feststellen.

Auch bei GSM 1800 wurde die entgeltige Analyse der Daten anhand der GLIMMIX-prozedur vorgenommen. In den Tabellen (13), (14), (15) und (16)

¹⁰⁷siehe. Anhang(C)

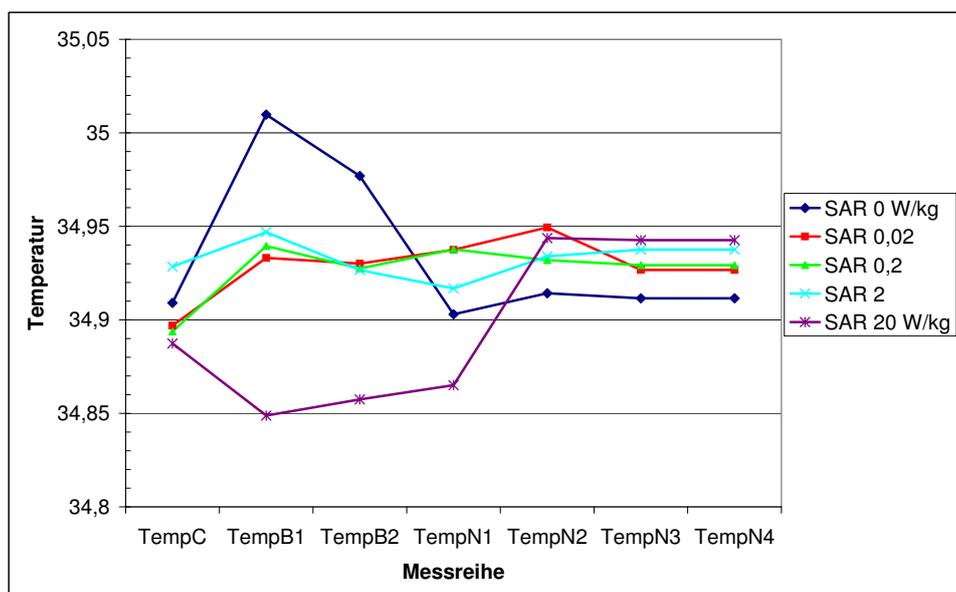


Abbildung 8: Temperaturkurve bei GSM 1800

werden wieder die p-Werte für die F-Statistiken für GSM 1800 angegeben. Alle signifikanten und hochsignifikanten Werte kleiner 0,05 sind dick gedruckt. Der SAS Output für die Glimmix-Prozedur der Differenz für die OFF Rate zum Zeitpunkt B2 ist in Anhang (D) zu finden.

| QUOTIENT B1 | Rate ON | Latenz ON | Rate OFF | Latenz OFF |
|-----------------------|---------|-----------|------------------|------------|
| SAR | 0,0564 | 0,3788 | <,0001 | 0,4337 |
| Intensität | 0,5528 | 0,7359 | 0,9236 | 0,7483 |
| Temperatur | 0,0777 | 0,8825 | 0,0016 | 0,8876 |
| SAR*Intensität | 0,5586 | 0,7568 | 0,0002 | 0,9414 |
| Intensität*Temperatur | 0,5430 | 0,7371 | 0,9227 | 0,7493 |

Tabelle 13: Signifikanzen des Quotienten B1 bei GSM 1800

Auch bei GSM 1800 unterscheiden sich teilweise die signifikanten p-Werte der Quotienten von den Differenzen.

Beim ersten Blick auf die Tabellen (13), (14), (15) und (16) fällt auf, dass signifikante p-Werte ausschließlich bei der Antwortrate vorliegen. Es liegen bis auf 2 Ausnahmen hochsignifikante p-Werte beim Haupteffekt SAR bei Quotienten und Differenzen der Antwortrate vor, während es bei den Latenzen zu keinen signifikanten Werten kam. Außerdem sind signifikante bis hochsignifikante Werte beim Effekt „Temperatur“, sowie 2 signifikante p-Werte beim Effekt der

| QUOTIENT B2 | Rate ON | Latenz ON | Rate OFF | Latenz OFF |
|-----------------------|------------------|-----------|------------------|------------|
| SAR | <,0001 | 0,3336 | <,0001 | 0,5191 |
| Intensität | 0,1847 | 0,3918 | 0,6733 | 0,6980 |
| Temperatur | 0,0480 | 0,3438 | 0,0239 | 0,6137 |
| SAR*Intensität | 0,0305 | 0,3643 | 0,0003 | 0,6605 |
| Intensität*Temperatur | 0,1833 | 0,3911 | 0,6740 | 0,7012 |

Tabelle 14: Signifikanzen des Quotienten B2 bei GSM 1800

| DIFFERENZ B1 | Rate ON | Latenz ON | Rate OFF | Latenz OFF |
|-----------------------|---------------|-----------|------------------|------------|
| SAR | 0,0069 | 0,2448 | <,0001 | 0,8311 |
| Intensität | 0,0045 | 0,9831 | 0,0386 | 0,5806 |
| Temperatur | 0,0002 | 0,8129 | 0,0003 | 0,3928 |
| SAR*Intensität | 0,3068 | 0,4129 | 0,0119 | 0,4414 |
| Intensität*Temperatur | 0,0046 | 0,9889 | 0,0390 | 0,5820 |

Tabelle 15: Signifikanzen der Differenzen B1 bei GSM 1800

| DIFFERENZ B2 | Rate ON | Latenz ON | Rate OFF | Latenz OFF |
|-----------------------|---------------|-----------|------------------|------------|
| SAR | 0,0002 | 0,6194 | <,0001 | 0,6444 |
| Intensität | 0,3092 | 0,5559 | 0,2340 | 0,8388 |
| Temperatur | 0,0175 | 0,5806 | 0,0097 | 0,4474 |
| SAR*Intensität | 0,0677 | 0,9072 | 0,0422 | 0,6346 |
| Intensität*Temperatur | 0,3083 | 0,5578 | 0,2342 | 0,8413 |

Tabelle 16: Signifikanzen der Differenzen B2 bei GSM 1800

Intensität, zu erkennen. Die Interaktion SAR*Intensität weist 5 Signifikanzen auf. Bei der Interaktion Intensität*Temperatur liegen zwei signifikante Unterschiede vor.

8.2 Diskussion der Ergebnisse

Wenn man die Ergebnisse der 3 Versuche UMTS Standard, GSM 900 und GSM 1800 zusammenfasst, liegen vor allem signifikante bis hochsignifikante Unterschiede beim SAR-Wert (in 15 von 48 Fällen), aber auch bei einigen Werten für Intensität und Temperatur, sowie bei der Interaktion zwischen SAR und Intensität bei den Quotienten und Differenzen der Zeitpunkte B1 und B2 (Messungen während der Befeldung) vor. Diese Effekte waren in erster Linie bei der Antwortrate zu finden. In lediglich zwei Fällen gab es signifikante Unterschiede bei der Latenz. (GSM 900 bei der Interaktion SAR*Intensität) Ebenfalls auffällig ist, dass es bei der Frequenz UMTS Standard zu lediglich 3 signifikanten Werten kommt, während bei GSM 900 25 und bei GSM 1800 24 Signifikanzen vorliegen.

In 12 von möglichen 48 Fällen tritt ein schwach oder sogar hoch signifikanter Haupteffekt der Temperatur auf, (davon allerdings keiner bei UMTS Standard) obwohl die Temperaturschwankungen über der gesamten Versuchsdauer lediglich bei $\pm 0,3^\circ\text{C}$ lagen. Das bedeutet, dass minimale Temperaturänderungen zu statistisch messbaren Veränderungen in der Aktivität der Neurone, bzw in den Ganglienzellen der Retina führen. Jedoch sind, wie bereits in Abschnitt (2.3.2) erwähnt, kleine Temperaturschwankungen dieser Art normal. Das Nervensystem hat sich an diese minimalen Temperaturschwankungen, sowie an die Änderungen der Aktivität von Nervenzellen aufgrund dieser Schwankungen nahezu gewöhnt. Deshalb sind die signifikanten Effekte der Temperatur schwer zu deuten. Es kann zum Einen der Fall sein, dass dies ein direkter Effekt der hochfrequenten Felder ist, zum Anderen ist es jedoch denkbar, dass dies von den minimalen Temperaturänderungen rührt.

Des Weiteren ist wichtig zu beachten, dass die Temperatur nicht direkt in der Retina gemessen werden konnte, sondern nur ausserhalb, direkt neben der Retina. Es ist somit möglich, dass die Temperatur in der Retina noch größeren Schwankungen ausgesetzt war.

Auffällig bei den Ergebnissen ist die vermehrte Signifikanz in 13 von 48 Fällen von dem Haupteffekt SAR, wovon die meisten sogar hochsignifikante Werte

waren. Dies widerspricht den Ergebnissen aus der Forschungsstudie. Dort gab es nur in 3 von 24 Fällen schwach signifikante Werte beim Haupteffekt SAR. Die Auswertung wurde in dieser Arbeit etwas anders durchgeführt. Die Nacheffekte N1-N4 gingen, ebenso wie die Messreihenfolge nicht mit in die Analyse ein. Es ist möglich, dass andere Ergebnisse erhalten werden, wenn die Messreihenfolge, sowie die anderen Zeitpunkte N1-N4 mit in die Analyse eingegangen wären.

Dass bei den signifikanten Werten von dem Effekt SAR von möglichen Fehlern 1. der 2. Art ausgegangen werden kann ist eher unwahrscheinlich, da hoch signifikante Werte vorliegen und die unberechtigte Ablehnung der Nullhypothese somit fast ausgeschlossen werden und es folglich kaum zu einem Fehler 1. Art kommen kann.

Auch beim Effekt der Temperatur liegen mehr hochsignifikante (7) als schwach signifikante (5) p-Werte vor, somit ist davon auszugehen, dass wirkliche Effekte vorliegen.

Bei den schwach signifikanten p-Werten (lediglich 4 von 48) bei dem Haupteffekt Intensität kann jedoch ein Fehler 1. Art vorliegen. Der Effekt der Intensität spiegelt die Antwortaktivität der retinalen Ganglienzellen wieder. Dieses Ergebnis war zu erwarten, da theoretisch angenommen wird, dass die Quotienten und Differenzen bei 1 oder 0 liegen.¹⁰⁸

Aufgrund der teilweise signifikanten Haupteffekte von SAR, Intensität und Temperatur kann man deuten, dass die Retina ihren Zustand (die Latenzen, vor allem aber die Antwortraten) durch die Faktoren verändert hat, im Vergleich mit der Kontrolle C, sowie auf Belichtung unterschiedlicher Intensität mit einer Veränderung reagiert.

Bei den Interaktionen ist vor allem SAR*Intensität ein signifikanter Effekt. In 16 von 48 Fällen (darunter sind auch die 3 signifikanten Werte bei UMTS Standard) liegt ein signifikanter bis hochsignifikanter p-Wert vor. Die Kombination von SAR und Intensität wirkt sich folglich signifikant auf die Latenzen und Antwortraten aus. Interessant ist, dass die Interaktion SAR*Intensität hier nicht mit den signifikanten Haupteffekten der Intensität zusammenhängt, es gibt keine Übereinstimmungen, in denen signifikante Werte bei der Interaktion und zugleich beim Haupteffekt vorliegen. Auch die Interaktion Intensität*Temperatur korreliert mit seinen 4 aufgetretenen schwach signifikanten

¹⁰⁸siehe: [(Abschlussbericht 2007)]

p-Werten nicht mit der Interaktion SAR*Intensität. In den meisten Fällen wo signifikante Werte bei der Interaktion SAR*Intensität vorliegen, sind jedoch Signifikanzen beim Haupteffekt SAR, sowie bei der Temperatur zu erkennen. Diese Zusammenhänge können der Beweis dafür sein, dass Temperaturveränderungen in der Retina, die durch den SAR-Wert hervorgerufen wurden, direkt oder indirekt von der Intensität der Befeldung abhängen. Dass die signifikanten Werte der Interaktion SAR*Temperatur nicht mit SAR*Intensität zusammenhängen, kann daran liegen, dass die Temperatur, wie eben erwähnt, nicht direkt in der Retina gemessen werden konnte. Es ist somit möglich, dass die Signifikanz von Temperatur*Intensität nur schwer nachweisbar ist.

Die Schwierigkeit, die Ergebnisse richtig zu deuten, liegt darin, dass nicht sicher ist, ob die Effekte von den Hochfrequenzfeldern rühren oder ob ein indirekter Temperatureffekt mit im Spiel ist. Das bleibt spekulativ.

Was allerdings zu beachten ist, ist dass eine geringe Temperaturerwärmung des Nervengewebes angestrebt werden sollte.

Damit ist auch klar, dass Befeldungen von einem SAR-Wert von 20 W/kg in keinem Fall gut für den Körper sind und man diesen nicht Werte von 20 W/kg aussetzen sollte. Allerdings geben Handys nicht annäherungsweise SAR-Werte von bis zu 20 W/kg ab.

9 Zusammenfassung und Schlussbemerkungen

Ziel dieser Arbeit war es die Modellwahl für die Analyse von longitudinalen Daten, speziell für Daten aus einer Forschungsstudie des visuellen Systems, zu treffen und anhand dieses Modells eine Analyse der longitudinalen Daten durchzuführen.

In der Arbeit wird deutlich, dass obwohl es viele Ansätze von Modellen gibt, um longitudinale Daten auszuwerten, hier sind z.B. die marginalen Modelle zu nennen, dennoch in vielen Fällen die subjektspezifischen Modelle die bessere Wahl sind. Es wird relativ schnell klar, dass das GLMM, welches unter die subjektspezifischen Modelle fällt, eine geeignete Wahl für die Datenauswertung ist. In der Forschungsstudie sind sowohl feste, als auch zufällige Effekte vorhanden, somit kommt nur ein gemischtes Modell in Frage. Die Wahl fällt unter Anderem auch deswegen auf das GLMM, da es sehr robust gegenüber Missing Values ist. Das ist für die Auswertung der Studie besonders von Bedeutung gewesen, da etliche Messungen nicht erhoben werden konnten und folglich sehr viele fehlende Werte vorlagen.

Bevor die Daten anhand des GLMMs ausgewertet wurden, wurden in den vorhergehenden Kapiteln die statistischen Grundlagen, welche für das Verständnis des GLMMs wichtig sind, geschaffen. Im Anschluss daran wurde dann das GLMM selbst dargestellt. Schätzungen wie die Maximum-Likelihood-Schätzung spielten dabei eine bedeutende Rolle. Da diese numerisch schwer zu berechnen sind, wurden in dem Zusammenhang verschiedene Aspekte, wie die beiden Typen der Quasi-Likelihood-Schätzungen PQL (*penalized quasi likelihood estimation*) und MQL (*marginal quasi likelihood estimation*), sowie die GEEs (*generalized estimating equations*) zum Schätzen der festen Effekte und die *small area estimation* im Bezug auf den BP (*best predictor*) zum Schätzen der Zufallseffekte charakterisiert. Nachdem kurz auf die theoretische Analyse von longitudinalen Daten eingegangen wurde, ist zum Schluss dieser Arbeit eine Analyse der Daten aus der Forschungsstudie durchgeführt worden. Hierbei ging es in erster Linie darum, anhand des GLMMs auf Signifikanzen der verschiedenen festen Effekte, sowie deren Interaktionen, zu testen. Dies wurde anhand der SAS-Prozedur GLIMMIX gemacht.

A Anhang 1: SAS Anweisungen für die MEANS-Prozedur

Die SAS Anweisungen für die Means-Prozedur für Beispiel (7.1) in Kapitel (7) sind gegeben durch:

```

DATA g_1;
INPUT Zeit$ Wert @@;
CARDS;
a 100 b 90 c 100 d 70 e 36 f 50 g 28
a 100 b 83 c 97 d 60 e 83 f 83 g 97
a 70 b 83 c 77 d 21 e 36 f 53 g 53
a 77 b 36 c 36 d 36 e 36 f 36 g 28
a 100 b 61 c 99 d 83 e 97 f 83 g 100
RUN;
PROC PRINT DATA=g_1;
RUN;
PROC SORT DATA=g_1;
BY Zeit;
RUN;
TITLE1 'Gruppe 1';
TITLE2 'Mittelwert und Standardabweichung';
FOOTNOTE1 'a:0 Minuten, b:5 Minuten, c:10 Minuten, d:20 Minuten,
e:30 Minuten, f:45 Minuten, g:60 Minuten';
PROC Print DATA=g_1;
RUN;
PROC MEANS DATA=g_1;
VAR Wert;
BY Zeit;
RUN;
DATA g_2;
INPUT Zeit$ Wert @@;
CARDS;
a 90 b 28 c 44 d 33 e 36 f 44 g 36

```

a 83 b 14 c 28 d 28 e 21 f 28 g 36

a 69 b 36 c 28 d 21 e 14 f 21 g 21

a 100 b 77 c 44 d 28 e 44 f 14

a 61 b 53 c 53 d 44 e 44 f 28 g 44

RUN;

PROC PRINT DATA=g_2;

RUN;

PROC SORT DATA=g_2;

BY Zeit;

RUN;

TITLE1 'GRUPPE 2';

TITLE2 'Mittelwert und Standardabweichung';

FOOTNOTE 'a:0 Minuten, b:5 Minuten, c:10 Minuten, d:20 Minuten,

e:30 Minuten, f:45 Minuten, g:60 Minuten';

PROC PRINT DATA=g_2;

RUN;

PROC MEANS DATA=g_2;

VAR Wert;

BY Zeit;

Auf den folgenden drei Seiten wird der Output für die gerade gegebene MEANS-Prozedur aufgeführt.

A ANHANG 1: SAS ANWEISUNGEN FÜR DIE MEANS-PROZEDUR

Hier wird für die Gruppe 1 der Mittelwert, Standardabweichung und jeweils das Minimum und das Maximum für die verschiedenen Zeiten mit dem Statistikprogramm SAS berechnet. Folgende Ergebnisse gab es dabei:

Gruppe 1 17:26 Monday, June 16, 2008 3
Mittelwert und Standardabweichung

----- Zeit=a -----

Die Prozedur MEANS

Analysis Variable : Wert

| N | Mittelwert | Std. abweichung | Minimum | Maximum |
|---|------------|--------------------|------------|-------------|
| 5 | 89.4000000 | 14.7241299 | 70.0000000 | 100.0000000 |

----- Zeit=b -----

Analysis Variable : Wert

| N | Mittelwert | Std. abweichung | Minimum | Maximum |
|---|------------|--------------------|------------|------------|
| 5 | 70.6000000 | 22.2103579 | 36.0000000 | 90.0000000 |

----- Zeit=c -----

Analysis Variable : Wert

| N | Mittelwert | Std. abweichung | Minimum | Maximum |
|---|------------|--------------------|------------|-------------|
| 5 | 81.8000000 | 27.2891920 | 36.0000000 | 100.0000000 |

----- Zeit=d -----

Analysis Variable : Wert

| N | Mittelwert | Std. abweichung | Minimum | Maximum |
|---|------------|--------------------|------------|------------|
| 5 | 54.0000000 | 25.2289516 | 21.0000000 | 83.0000000 |

----- Zeit=e -----

Die Prozedur MEANS

Analysis Variable : Wert

Std.

A ANHANG 1: SAS ANWEISUNGEN FÜR DIE MEANS-PROZEDUR

| N | Mittelwert | Std. abweichung | Minimum | Maximum |
|---|------------|--------------------|------------|------------|
| 5 | 57.6000000 | 29.9883311 | 36.0000000 | 97.0000000 |

----- Zeit=f -----

Analysis Variable : Wert

| N | Mittelwert | Std. abweichung | Minimum | Maximum |
|---|------------|--------------------|------------|------------|
| 5 | 61.0000000 | 21.0831686 | 36.0000000 | 83.0000000 |

----- Zeit=g -----

Analysis Variable : Wert

| N | Mittelwert | Std. abweichung | Minimum | Maximum |
|---|------------|--------------------|------------|-------------|
| 5 | 61.2000000 | 35.5626208 | 28.0000000 | 100.0000000 |

a:0 Minuten, b:5 Minuten, c:10 Minuten, d:20 Minuten, e:30 Minuten, f:45 Minuten, g:60 Minuten

Standardabweichung, Mittelwert, Minimum und Maximum werden nun auch für Gruppe 2 mit SAS berechnet.

GRUPPE 2 14:13 Tuesday, June 17, 2008 3
Mittelwert und Standardabweichung

----- Zeit=a -----

Die Prozedur MEANS

Analysis Variable : Wert

| N | Mittelwert | Std. abweichung | Minimum | Maximum |
|---|------------|--------------------|------------|-------------|
| 5 | 80.6000000 | 15.7257750 | 61.0000000 | 100.0000000 |

----- Zeit=b -----

Analysis Variable : Wert

| N | Mittelwert | Std. abweichung | Minimum | Maximum |
|---|------------|--------------------|------------|------------|
| 5 | 41.6000000 | 24.2960902 | 14.0000000 | 77.0000000 |

B Anhang 2: Versuchsdesign Forschungsstudie

Auf den folgenden beiden Seiten wird das Versuchsdesign der Forschungsstudie des visuellen Systems anhand einer Tabelle verdeutlicht. Die Tabelle zeigt einen kleinen Auszug aus einer Excel-Tabelle, in der die vollständigen Datensätze getrennt für die Frequenzen UMTS Standard, GSM 900 und GSM 1800 festgehalten wurden.

Hier werden lediglich die verschiedenen Merkmale aufgeführt, es fehlen, aufgrund seiner Komplexität, die kompletten Messungen, die bei der Studie vorgenommen wurden, wie z.B. die Temperatur, Latenz und Antwortrate. Diese werden hier nicht weiter aufgeführt, da, wie eben schon erwähnt, der Umfang der Daten einfach zu groß ist.

Die Zahlen in der Spalte *Intensität* stehen jeweils für die verschiedenen Intensitäten. Dabei gilt Folgendes:

| | | | | | | |
|------------|---------|--------|---------|-------|-------|--------|
| Intensität | 1 | 2 | 3 | 4 | 5 | 6 |
| Luxwert | 0,07 lx | 0,5 lx | 2,66 lx | 16 lx | 83 lx | 445 lx |

In der Analyse wurden selbstverständlich die Luxwerte verwendet.

Die Zahlen in der Spalte *Intensitätsreihe* geben die verschiedenen Reize an, welche schon in Kapitel (2.3.2) näher erklärt wurden. Dabei steht die *1* für die konstanten Stimuli, die eine retinale Beleuchtungsstärke von 2,66 lx aufwiesen, die *2* steht für die Teststimuli T3 mit den jeweiligen aufsteigenden Intensitäten von 0,5 lx; 16 lx und 445 lx und die *3* steht für die Teststimuli T6 mit den jeweils 6 aufsteigenden Intensitäten von 0,07 lx; 0,5 lx; 2,66 lx; 16 lx; 83 lx und 445 lx.

B ANHANG 2: VERSUCHSDESIGN FORSCHUNGSSTUDIE

| Versuch # | Unit # | SAR | Intensität | Intensitätsreihe | Messreihe |
|-----------|--------|-----|------------|------------------|-----------|
| 1 | 1 | 1 | 0 | 3 | 1 |
| 1 | 1 | 1 | 0 | 2 | 2 |
| 1 | 1 | 1 | 0 | 4 | 3 |
| 1 | 1 | 1 | 0 | 6 | 4 |
| 1 | 1 | 1 | 0 | 3 | 5 |
| 1 | 1 | 1 | 0 | 1 | 6 |
| 1 | 1 | 1 | 0 | 2 | 7 |
| 1 | 1 | 1 | 0 | 3 | 8 |
| 1 | 1 | 1 | 0 | 4 | 9 |
| 1 | 1 | 1 | 0 | 5 | 10 |
| 1 | 1 | 1 | 0 | 6 | 11 |
| 1 | 2 | 0 | 3 | 1 | 1 |
| 1 | 2 | 0 | 2 | 2 | 2 |
| 1 | 2 | 0 | 4 | 2 | 3 |
| 1 | 2 | 0 | 6 | 2 | 4 |
| 1 | 2 | 0 | 3 | 1 | 5 |
| 1 | 2 | 0 | 1 | 3 | 6 |
| 1 | 2 | 0 | 2 | 3 | 7 |
| 1 | 2 | 0 | 3 | 3 | 8 |
| 1 | 2 | 0 | 4 | 3 | 9 |
| 1 | 2 | 0 | 5 | 3 | 10 |
| 1 | 2 | 0 | 6 | 3 | 11 |
| 1 | 3 | 0 | 3 | 1 | 1 |
| 1 | 3 | 0 | 2 | 2 | 2 |
| 1 | 3 | 0 | 4 | 2 | 3 |
| 1 | 3 | 0 | 6 | 2 | 4 |
| 1 | 3 | 0 | 3 | 1 | 5 |
| 1 | 3 | 0 | 1 | 3 | 6 |
| 1 | 3 | 0 | 2 | 3 | 7 |
| 1 | 3 | 0 | 3 | 3 | 8 |
| 1 | 3 | 0 | 4 | 3 | 9 |
| 1 | 3 | 0 | 5 | 3 | 10 |
| 1 | 3 | 0 | 6 | 3 | 11 |
| 1 | 4 | 0 | 3 | 1 | 1 |
| 1 | 4 | 0 | 2 | 2 | 2 |
| 1 | 4 | 0 | 4 | 2 | 3 |
| 1 | 4 | 0 | 6 | 2 | 4 |
| 1 | 4 | 0 | 3 | 1 | 5 |
| 1 | 4 | 0 | 1 | 3 | 6 |
| 1 | 4 | 0 | 2 | 3 | 7 |
| 1 | 4 | 0 | 3 | 3 | 8 |
| 1 | 4 | 0 | 4 | 3 | 9 |
| 1 | 4 | 0 | 5 | 3 | 10 |
| 1 | 4 | 0 | 6 | 3 | 11 |
| 1 | 5 | 0 | 3 | 1 | 1 |
| 1 | 5 | 0 | 2 | 2 | 2 |
| 1 | 5 | 0 | 4 | 2 | 3 |
| 1 | 5 | 0 | 6 | 2 | 4 |
| 1 | 5 | 0 | 3 | 1 | 5 |
| 1 | 5 | 0 | 1 | 3 | 6 |
| 1 | 5 | 0 | 2 | 3 | 7 |

| | | | | | |
|---|---|------|---|---|----|
| 1 | 5 | 0 | 3 | 3 | 8 |
| 1 | 5 | 0 | 4 | 3 | 9 |
| 1 | 5 | 0 | 5 | 3 | 10 |
| 1 | 5 | 0 | 6 | 3 | 11 |
| 2 | 1 | 0,02 | 3 | 1 | 1 |
| 2 | 1 | 0,02 | 2 | 2 | 2 |
| 2 | 1 | 0,02 | 4 | 2 | 3 |
| 2 | 1 | 0,02 | 6 | 2 | 4 |
| 2 | 1 | 0,02 | 3 | 1 | 5 |
| 2 | 1 | 0,02 | 1 | 3 | 6 |
| 2 | 1 | 0,02 | 2 | 3 | 7 |
| 2 | 1 | 0,02 | 3 | 3 | 8 |
| 2 | 1 | 0,02 | 4 | 3 | 9 |
| 2 | 1 | 0,02 | 5 | 3 | 10 |
| 2 | 1 | 0,02 | 6 | 3 | 11 |
| 2 | 2 | 0,02 | 3 | 1 | 1 |
| 2 | 2 | 0,02 | 2 | 2 | 2 |
| 2 | 2 | 0,02 | 4 | 2 | 3 |
| 2 | 2 | 0,02 | 6 | 2 | 4 |
| 2 | 2 | 0,02 | 3 | 1 | 5 |
| 2 | 2 | 0,02 | 1 | 3 | 6 |
| 2 | 2 | 0,02 | 2 | 3 | 7 |
| 2 | 2 | 0,02 | 3 | 3 | 8 |
| 2 | 2 | 0,02 | 4 | 3 | 9 |
| 2 | 2 | 0,02 | 5 | 3 | 10 |
| 2 | 2 | 0,02 | 6 | 3 | 11 |
| 2 | 3 | 0,02 | 3 | 1 | 1 |
| 2 | 3 | 0,02 | 2 | 2 | 2 |
| 2 | 3 | 0,02 | 4 | 2 | 3 |
| 2 | 3 | 0,02 | 6 | 2 | 4 |
| 2 | 3 | 0,02 | 3 | 1 | 5 |
| 2 | 3 | 0,02 | 1 | 3 | 6 |
| 2 | 3 | 0,02 | 2 | 3 | 7 |
| 2 | 3 | 0,02 | 3 | 3 | 8 |
| 2 | 3 | 0,02 | 4 | 3 | 9 |
| 2 | 3 | 0,02 | 5 | 3 | 10 |
| 2 | 3 | 0,02 | 6 | 3 | 11 |
| 2 | 4 | 0,02 | 3 | 1 | 1 |
| 2 | 4 | 0,02 | 2 | 2 | 2 |
| 2 | 4 | 0,02 | 4 | 2 | 3 |
| 2 | 4 | 0,02 | 6 | 2 | 4 |
| 2 | 4 | 0,02 | 3 | 1 | 5 |
| 2 | 4 | 0,02 | 1 | 3 | 6 |
| 2 | 4 | 0,02 | 2 | 3 | 7 |
| 2 | 4 | 0,02 | 3 | 3 | 8 |
| 2 | 4 | 0,02 | 4 | 3 | 9 |
| 2 | 4 | 0,02 | 5 | 3 | 10 |
| 2 | 4 | 0,02 | 6 | 3 | 11 |
| 2 | 5 | 0,02 | 3 | 1 | 1 |
| 2 | 5 | 0,02 | 2 | 2 | 2 |
| 2 | 5 | 0,02 | 4 | 2 | 3 |
| 2 | 5 | 0,02 | 6 | 2 | 4 |

C Anhang 3: SAS Output für den Dunnett-Test

Im Folgenden sind Ausschnitte des Outputs des Quellcodes:

```
PROC GLM DATA=SASUSER.QUOUMTS16lx;  
CLASS SAR;  
MODEL Rate_ON_C Rate_ON_B1_C Rate_ON_B2_C  
Rate_ON_N1_C Rate_ON_N2_C  
Rate_ON_N3_C Rate_ON_N4_C=SAR;  
MEANS SAR/DUNNETT ('0') ALPHA=0.05;  
RUN; QUIT;
```

zur Bestimmung der Signifikanzen der Mittelwerte bei UMTS Standard, GSM 900 und GSM 1800 für eine Intensität von 16 lx zur Kontrollgruppe SAR=0 aufgeführt. Der Quellcode kann für alle anderen Intensitäten sowie für die Differenzen bzw. Quotienten von Antwortrate und Latenz, unter Umänderung der Variablen verwendet werden, deshalb werden in der Arbeit auch keine weiteren Dunnett-Tests aufgeführt, da diese immer nach dem gleichen Schema ablaufen. Auf den folgenden 2 Seiten ist der Output für UMTS Standard dargestellt, auf der 3 Seite für GSM 900 und auf der 4 Seite für GSM 1800, dort sind signifikante Unterschiede bei bestimmten SAR-Werten zu erkennen.

C ANHANG 3: SAS OUTPUT FÜR DEN DUNNETT-TEST

Das SAS System

Die Prozedur GLM

Klassifizierungsausprägungsinformationen

| Klasse | Ausprägungen | Werte |
|--------|--------------|-----------------|
| SAR | 5 | 0 2 20 0.2 0.02 |

Daten für Analyse von Rate_ON_C
 Rate_ON_N1_C Rate_ON_N2_c
 Rate_ON_N3_C Rate_ON_N4_C

| | |
|-----------------------------|-----|
| Number of Observations Read | 286 |
| Number of Observations Used | 286 |

Daten für Analyse von Rate_ON_B1_C
 Rate_ON_B2_C

| | |
|-----------------------------|-----|
| Number of Observations Read | 286 |
| Number of Observations Used | 143 |

HINWEIS: Variables in each group are consistent with respect to the presence or absence of missing values.

Abhängige Variable: Rate_ON_N1_C

| Quelle | Freiheits- grade | Summe der Quadrate | Mittleres Quadrat | F-Statistik | Pr > |
|-------------------|---------------------|-----------------------|----------------------|-------------|------|
| F | | | | | |
| Modell | 4 | 3.9182067 | 0.9795517 | 2.78 | |
| 0.0271 | | | | | |
| Fehler | 281 | 98.9292774 | 0.3520615 | | |
| Korrigierte Summe | 285 | 102.8474841 | | | |

| | | | |
|-----------|------------|------------|-------------------------|
| R-Quadrat | Koeff. var | Wurzel MSE | Rate_ON_N1_C Mittelwert |
| 0.038097 | 55.10069 | 0.593348 | 1.076843 |

| Quelle | Freiheits- grade | Typ I SS | Mittleres Quadrat | F-Statistik | Pr > |
|--------|---------------------|------------|----------------------|-------------|------|
| F | | | | | |
| SAR | 4 | 3.91820669 | 0.97955167 | 2.78 | |
| 0.0271 | | | | | |

| Quelle | Freiheits- grade | Typ III SS | Mittleres Quadrat | F-Statistik | Pr > |
|--------|---------------------|------------|----------------------|-------------|------|
| F | | | | | |
| SAR | 4 | 3.91820669 | 0.97955167 | 2.78 | |
| 0.0271 | | | | | |

C ANHANG 3: SAS OUTPUT FÜR DEN DUNNETT-TEST

Die Prozedur GLM

Dunnetts t-Tests auf Rate_ON_N1_C

HINWEIS: Dieser Test kontrolliert den Fehler erster Art für das gesamte Experiment für
Vergleiche

aller Behandlungen gegen eine Kontrolle.

Alpha 0.05
Freiheitsgrade des Fehlers 281
Mittlerer quadratischer Fehler 0.352061
Kritischer Wert von Dunnetts t 2.44918

Vergleiche, die zum 0.05 Niveau signifikant sind, werden durch *** gekennzeichnet.

| SAR | Differenz | Simultan 95% |
|-----------|--------------|------------------|
| Vergleich | zwischen | Konfidenzgrenzen |
| | Mittelwerten | |
| 20 - 0 | 0.07273 | -0.19174 0.33719 |
| 0.02 - 0 | -0.06042 | -0.34542 0.22458 |
| 2 - 0 | -0.13708 | -0.41241 0.13826 |
| 0.2 - 0 | -0.26612 | -0.55396 0.02171 |

C ANHANG 3: SAS OUTPUT FÜR DEN DUNNETT-TEST

GSM 900
Das SAS System

Die Prozedur GLM

Dunnetts t-Tests auf Rate_OFF_B1_C

HINWEIS: Dieser Test kontrolliert den Fehler erster Art für das gesamte Experiment für
Vergleiche

aller Behandlungen gegen eine Kontrolle.

Alpha 0.05
Freiheitsgrade des Fehlers 180
Mittlerer quadratischer Fehler 148.4345
Kritischer Wert von Dunnetts t 2.48579

Vergleiche, die zum 0.05 Niveau signifikant sind, werden durch *** gekennzeichnet.

| SAR Vergleich | Differenz zwischen Mittelwerten | Simultan 95% Konfidenzgrenze n | |
|------------------|---------------------------------------|--------------------------------------|-----|
| 0.02 - 0 | 7.392 | 0.448 14.335 | *** |
| 20 - 0 | 2.490 | -3.925 8.906 | |
| 0.2 - 0 | 0.892 | -6.052 7.835 | |
| 2 - 0 | -0.919 | -7.599 5.760 | |

C ANHANG 3: SAS OUTPUT FÜR DEN DUNNETT-TEST

Das SAS System
Die Prozedur GLM

Dunnetts t-Tests auf Rate_ON_N1_C

HINWEIS: Dieser Test kontrolliert den Fehler erster Art für das gesamte Experiment für
Vergleiche

aller Behandlungen gegen eine Kontrolle.

Alpha 0.05
Freiheitsgrade des Fehlers 341
Mittlerer quadratischer Fehler 113.4594
Kritischer Wert von Dunnetts t 2.45874

Vergleiche, die zum 0.05 Niveau signifikant sind, werden durch *** gekennzeichnet.

| SAR Vergleich | Differenz zwischen Mittelwerten | Simultan Konfidenzgrenze | 95% n | |
|------------------|---------------------------------------|-----------------------------|----------|-----|
| 0.02 - 0 | 4.877 | 0.627 | 9.127 | *** |
| 20 - 0 | 3.194 | -1.277 | 7.665 | |
| 2 - 0 | 2.332 | -1.918 | 6.582 | |
| 0.2 - 0 | 0.916 | -3.823 | 5.655 | |

Die Prozedur GLM

Dunnetts t-Tests auf Rate_ON_N4_C

HINWEIS: Dieser Test kontrolliert den Fehler erster Art für das gesamte Experiment für
Vergleiche

aller Behandlungen gegen eine Kontrolle.

Alpha 0.05
Freiheitsgrade des Fehlers 341
Mittlerer quadratischer Fehler 197.4256
Kritischer Wert von Dunnetts t 2.45874

Vergleiche, die zum 0.05 Niveau signifikant sind, werden durch *** gekennzeichnet.

| SAR Vergleich | Differenz zwischen Mittelwerten | Simultan Konfidenzgrenze | 95% n | |
|------------------|---------------------------------------|-----------------------------|----------|-----|
| 0.02 - 0 | 5.716 | 0.110 | 11.322 | *** |
| 20 - 0 | 4.834 | -1.063 | 10.732 | |
| 2 - 0 | 2.780 | -2.826 | 8.386 | |
| 0.2 - 0 | 1.306 | -4.946 | 7.557 | |

D Anhang 4: SAS Output für die GLIMMIX-Prozedur

In Kapitel (8.1.1), sowie in den folgenden beiden Abschnitten, wurden die p-Werte für die Tests der festen Effekte in Tabellen nach Quotient und Differenz, sowie B1 bzw B2, angegeben. Die p-Werte wurden jeweils in einer separaten Glimmix-Prozedur berechnet. Damit dies etwas deutlicher wird, wird nun die Prozedur mitsamt dem Output für die ON Rate der Differenz bei UMTS Standard folgen. Die Prozedur ist gegeben durch:

```
proc glimmix data=SASUSER.UMTS method=MSPL IC=Q;  
class Unit_;  
model Rate_ON_B2C=SAR Intensitaet Temp_B2 SAR*Intensitaet Inten-  
sitaet*Temp_B2;  
random intercept / subject=Unit_;  
run;
```

Der SAS OUTPUT wird auf den beiden nächsten Seiten aufgeführt. Für alle anderen Differenzen und Quotienten wurde die gleiche Prozedur verwendet nur unter Austauschung der Zielvariablen (hier ist das Rate_ON_B2C). Deshalb werden diese, sowie die anderen SAS Outputs hier nicht weiter aufgeführt. Im SAS Output ist unter der Rubrik „Type III Tests of Fixed Effects“ deutlich die Signifikanz der Interaktion „SAR*Intensitaet“ zu erkennen.

Weiter oben des SAS Outputs gibt es die Rubrik „Fit Statistics“ wo die Werte der verschiedenen Informationskriterien zu sehen sind. Diese Kategorie wird aufgrund der Anweisung IC=Q aufgeführt. Die Methode, womit die Schätzungen bestimmt wurden, ist „method=MSPL“, das sagt aus, das die „penalized quasi likelihood“ (PQL) verwendet wurde.

Nach dem Output für UMTS-Standard wird der Output der Glimmix-Prozedur für GSM 900 für den Quotienten der On Antwortrate zum Zeitpunkt B1, sowie der Output für GSM 1800 der Differenz für die OFF Rate zum Zeitpunkt B2 dargestellt.

UMTS DIFFERENZ ON RATE B2: SAS OUTPUT

The GLIMMIX Procedure

Model Information

| | |
|----------------------------|--------------------|
| Data Set | SASUSER.UMTS |
| Response Variable | Rate_ON_B2C |
| Response Distribution | Gaussian |
| Link Function | Identity |
| Variance Function | Default |
| Variance Matrix Blocked By | Unit__ |
| Estimation Technique | Maximum Likelihood |
| Degrees of Freedom Method | Containment |

Class Level Information

| Class | Levels | Values |
|--------|--------|-----------|
| Unit__ | 5 | 1 2 3 4 5 |

| | |
|-----------------------------|------|
| Number of Observations Read | 1573 |
| Number of Observations Used | 715 |

Dimensions

| | |
|--------------------------|-----|
| G-side Cov. Parameters | 1 |
| R-side Cov. Parameters | 1 |
| Columns in X | 6 |
| Columns in Z per Subject | 1 |
| Subjects (Blocks in V) | 5 |
| Max Obs per Subject | 210 |

Optimization Information

| | |
|----------------------------|-------------------|
| Optimization Technique | Dual Quasi-Newton |
| Parameters in Optimization | 1 |
| Lower Boundaries | 1 |
| Upper Boundaries | 0 |
| Fixed Effects | Profiled |
| Residual Variance | Profiled |
| Starting From | Data |

Iteration History

| Iteration | Restarts | Evaluations | Objective Function | Change | Max Gradient |
|-----------|----------|-------------|--------------------|------------|--------------|
| 0 | 0 | 4 | 5591.1899818 | . | 51.91766 |
| 1 | 0 | 6 | 5591.1719466 | 0.01803519 | 26.46352 |

D ANHANG 4: SAS OUTPUT FÜR DIE GLIMMIX-PROZEDUR

The GLIMMIX Procedure

Iteration History

| Iteration | Restarts | Evaluations | Objective Function | Change | Max Gradient |
|-----------|----------|-------------|--------------------|------------|--------------|
| 2 | 0 | 2 | 5591.1571658 | 0.01478087 | 7.11602 |
| 3 | 0 | 2 | 5591.1562557 | 0.00091012 | 1.495642 |
| 4 | 0 | 2 | 5591.1562167 | 0.00003901 | 0.066333 |
| 5 | 0 | 2 | 5591.1562166 | 0.00000008 | 0.000587 |

HINWEIS: Convergence criterion (GCONV=1E-8) satisfied.

Fit Statistics

| | |
|--------------------------|----------|
| -2 Log Likelihood | 5591.16 |
| AIC (smaller is better) | 5607.16 |
| AICC (smaller is better) | 5607.36 |
| BIC (smaller is better) | 5604.03 |
| CAIC (smaller is better) | 5612.03 |
| HQIC (smaller is better) | 5598.77 |
| Generalized Chi-Square | 103833.3 |
| Gener. Chi-Square / DF | 145.22 |

Covariance Parameter Estimates

| Cov Parm | Subject | Estimate | Standard Error |
|-----------|---------|----------|----------------|
| Intercept | Unit__ | 0.7537 | 1.2186 |
| Residual | | 145.22 | 7.7068 |

Type III Tests of Fixed Effects

| Effect | Num DF | Den DF | F-Statistik | Pr > F |
|---------------------|--------|--------|-------------|--------|
| SAR | 1 | 705 | 1.57 | 0.2107 |
| Intensitaet | 1 | 705 | 1.28 | 0.2575 |
| Temp_B2 | 1 | 705 | 0.28 | 0.5964 |
| SAR*Intensitaet | 1 | 705 | 6.78 | 0.0094 |
| Intensitaet*Temp_B2 | 1 | 705 | 1.29 | 0.2573 |

SAS OUTPUT für GSM 900 für den Quotienten der On Antwortrate zum Zeitpunkt B1:

Das SAS System
The GLIMMIX Procedure

Model Information

| | |
|----------------------------|--------------------|
| Data Set | SASUSER.GSM900 |
| Response Variable | Rate_ON_B1_C |
| Response Distribution | Gaussian |
| Link Function | Identity |
| Variance Function | Default |
| Variance Matrix Blocked By | Unit__ |
| Estimation Technique | Maximum Likelihood |
| Degrees of Freedom Method | Containment |

Class Level Information

| Class | Levels | Values |
|--------|--------|-----------------|
| Unit__ | 8 | 1 2 3 4 5 6 7 8 |

| | |
|-----------------------------|------|
| Number of Observations Read | 2046 |
| Number of Observations Used | 736 |

Dimensions

| | |
|--------------------------|-----|
| G-side Cov. Parameters | 1 |
| R-side Cov. Parameters | 1 |
| Columns in X | 6 |
| Columns in Z per Subject | 1 |
| Subjects (Blocks in V) | 8 |
| Max Obs per Subject | 166 |

Optimization Information

| | |
|----------------------------|-------------------|
| Optimization Technique | Dual Quasi-Newton |
| Parameters in Optimization | 1 |
| Lower Boundaries | 1 |
| Upper Boundaries | 0 |
| Fixed Effects | Profiled |
| Residual Variance | Profiled |
| Starting From | Data |

Iteration History

| Iteration | Restarts | Evaluations | Objective Function | Change | Max Gradient |
|-----------|----------|-------------|--------------------|------------|--------------|
| 0 | 0 | 4 | 1777.4693985 | . | 20.90858 |
| 1 | 0 | 5 | 1777.4484526 | 0.02094596 | 8.784474 |

The GLIMMIX Procedure

Iteration History

| Iteration | Restarts | Evaluations | Objective Function | Change | Max Gradient |
|-----------|----------|-------------|--------------------|------------|--------------|
| 2 | 0 | 2 | 1777.4414256 | 0.00702693 | 1.646144 |
| 3 | 0 | 2 | 1777.4411986 | 0.00022701 | 0.169603 |
| 4 | 0 | 2 | 1777.4411963 | 0.00000237 | 0.0029 |

HINWEIS: Convergence criterion (GCONV=1E-8) satisfied.

Fit Statistics

| | |
|--------------------------|---------|
| -2 Log Likelihood | 1777.44 |
| AIC (smaller is better) | 1793.44 |
| AICC (smaller is better) | 1793.64 |
| BIC (smaller is better) | 1794.08 |
| CAIC (smaller is better) | 1802.08 |
| HQIC (smaller is better) | 1789.15 |
| Generalized Chi-Square | 477.01 |
| Gener. Chi-Square / DF | 0.65 |

Covariance Parameter Estimates

| Cov Parm | Subject | Estimate | Standard Error |
|-----------|---------|----------|----------------|
| Intercept | Unit__ | 0.01434 | 0.01174 |
| Residual | | 0.6481 | 0.03395 |

Type III Tests of Fixed Effects

| Effect | Num DF | Den DF | F-Statistik | Pr > F |
|---------------------|--------|--------|-------------|--------|
| SAR | 1 | 723 | 21.64 | <.0001 |
| Intensitaet | 1 | 723 | 4.34 | 0.0375 |
| Temp_B1 | 1 | 723 | 12.13 | 0.0005 |
| SAR*Intensitaet | 1 | 723 | 5.29 | 0.0218 |
| Intensitaet*Temp_B1 | 1 | 723 | 4.36 | 0.0372 |

SAS OUTPUT für GSM 1800 der Differenz für die OFF Rate zum Zeitpunkt B2:

Das SAS System
The GLIMMIX Procedure

Model Information

| | |
|----------------------------|--------------------|
| Data Set | SASUSER.GSM1800 |
| Response Variable | Rate_OFF_B2C |
| Response Distribution | Gaussian |
| Link Function | Identity |
| Variance Function | Default |
| Variance Matrix Blocked By | Unit__ |
| Estimation Technique | Maximum Likelihood |
| Degrees of Freedom Method | Containment |

Class Level Information

| Class | Levels | Values |
|--------|--------|---------------|
| Unit__ | 7 | 1 2 3 4 5 6 7 |

| | |
|-----------------------------|------|
| Number of Observations Read | 1925 |
| Number of Observations Used | 854 |

Dimensions

| | |
|--------------------------|-----|
| G-side Cov. Parameters | 1 |
| R-side Cov. Parameters | 1 |
| Columns in X | 6 |
| Columns in Z per Subject | 1 |
| Subjects (Blocks in V) | 7 |
| Max Obs per Subject | 233 |

Optimization Information

| | |
|----------------------------|-------------------|
| Optimization Technique | Dual Quasi-Newton |
| Parameters in Optimization | 1 |
| Lower Boundaries | 1 |
| Upper Boundaries | 0 |
| Fixed Effects | Profiled |
| Residual Variance | Profiled |
| Starting From | Data |

Iteration History

| Iteration | Restarts | Evaluations | Objective Function | Change | Max Gradient |
|-----------|----------|-------------|-----------------------|--------|-----------------|
| 0 | 0 | 4 | 6883.9058871 | . | 0 |

The GLIMMIX Procedure

HINWEIS: Convergence criterion (ABSGCONV=0.00001) satisfied.

Estimated G matrix is not positive definite.

Fit Statistics

| | |
|--------------------------|----------|
| -2 Log Likelihood | 6883.91 |
| AIC (smaller is better) | 6897.91 |
| AICC (smaller is better) | 6898.04 |
| BIC (smaller is better) | 6897.53 |
| CAIC (smaller is better) | 6904.53 |
| HQIC (smaller is better) | 6893.23 |
| Generalized Chi-Square | 158392.9 |
| Gener. Chi-Square / DF | 185.47 |

Covariance Parameter Estimates

| Cov Parm | Subject | Estimate | Standard Error |
|-----------|---------|----------|----------------|
| Intercept | Unit__ | 0 | . |
| Residual | | 185.47 | 8.9756 |

Type III Tests of Fixed Effects

| Effect | Num DF | Den DF | F-Statistik | Pr > F |
|---------------------|--------|--------|-------------|--------|
| SAR | 1 | 842 | 54.63 | <.0001 |
| Intensitaet | 1 | 842 | 1.42 | 0.2340 |
| Temp_B2 | 1 | 842 | 6.71 | 0.0097 |
| SAR*Intensitaet | 1 | 842 | 4.14 | 0.0422 |
| Intensitaet*Temp_B2 | 1 | 842 | 1.42 | 0.2342 |

Literaturverzeichnis

- [Abschlussbericht 2007] Apl. Prof. Dr. Ammermüller, Josef: *Abschlussbericht zum Forschungsvorhaben: „Einfluss hochfrequenter elektromagnetischer Felder der Mobilfunkkommunikation auf Sinnesorgane. B. Das visuelle System.“*;
Carl-von-Ossietzky Universität Oldenburg
Fakultät V, Mathematik und Naturwissenschaften
Institut für Biologie und Umweltwissenschaften; 2007
- [Akaike 1974] Akaike, H.: *A New Look at the Statistical Model Identification* ;
In: IEEE Transaction on Automatic Control, Vol.19, Seiten 716- 723; 1974
- [Avital,Laird,Slasor 1997] Avital, Cnaan; Laird, Nan M. ; Slasor, Peter : *Tutorial in Biostatistics: Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data*;
In: Statistics in Medicine Vol.16, Seiten 2349 - 2380; 1997
- [Bankhofer 1995] Bankhofer, Udo: *Unvollständige Daten- und Distanzmatrizen in der Multivariaten Datenanalyse*; Bergisch-Gladbach, Eul; 1995
- [Breslow und Clayton 1993] Breslow, N.E.; Clayton, D.G.: *Approximate Inference in generalized linear mixed models*; In: Journal of the American Statistical Association,Vol. 88, No. 421 Seiten 9 - 25; 1993
- [Brunner und Langer 1999] Brunner, E.; Langer, F: *Nichtparametrische Analyse Longitudinaler Daten*; München: Oldenbourg Verlag; 1999
- [Diggle 2002] Diggle, Peter J.; Heagerty, Patrick; Liang, Kung-Yee; Zeger, Scott L.: *Analysis of longitudinal Data*; Oxford University Press; 2002
- [Dufner,Jensen,Schumacher 2004] Dufner, Julius; Jensen, Uwe; Schumacher, Erich: *Statistik mit SAS; 3.Auflage*; B.G. Teubner Verlag/ GWV Fachverlage Gmbh, Wiesbaden; Juni 2004
- [Fahrmeier 2003] Fahrmeier, Ludwig; Künstler, Rita; Pigeot, Iris; Tutz, Gerhard 2003: *Statistik, Der Weg zur Datenanalyse; 4.Auflage*; Springer-Verlag Berlin Heidelberg; 2003

- [Fox 2002] Fox, John: *Linear Mixed Models (Appendix to An R and S-PLUS Companion to Applied Regression)*; May 2002
(<http://cran.r-project.org/doc/contrib/Fox-Companion/appendix-mixed-models.pdf>; Stand vom 22.04.2008)
- [Green 1987] Green, Peter: *Penalized Likelihood for General Semi-Parametric Regression Models*; In: International Statistical Review, Revue Internationale de Statistique, Vol. 55, No. 3 Seiten 245 - 259; 1987
- [Greven 2005] Greven, Sonja: *Lineare gemischte Modelle für longitudinale Daten*; 10. Januar 2005
(<http://www.stat.uni-muenchen.de/~greven/presentations/LMMsforlongitudinaldata.pdf>, Stand vom 9.11.2007)
- [Grunau und Riemer 2005] Grunau, Janne; Riemer, Alexander: *Clustering*; 5. Januar 2005
(<http://www.zib.de/andrzejak/lehre/dm0405/data/A8.pdf>, Stand vom 3.7.2008)
- [Hatzinger 1991] Hatzinger, Reinhold: *Quasi-Likelihood Methoden zur Analyse von unabhängigen und abhängigen Beobachtungen*; Institut für Statistik; Wirtschaftsuniversität Wien; Forschungsberichte; Bericht 13; 1991
(<http://statmath.wu-wien.ac.at/>
Stand vom 30.9.2008)
- [Hedeker] Hedeker, Donald: *Generalized Linear Mixed Models*;
In: Encyclopedia of Statistics in Behavioral Science. John Wiley & Sons, Ltd
- [Heuser 2001] Heuser, Harro: *Lehrbuch der Analysis, Teil 1, 14. Auflage*; BG Teubner GMBH Stuttgart, Leipzig, Wiesbaden; 2001
- [Heuser 2002] Heuser, Harro: *Lehrbuch der Analysis, Teil 2, 12. Auflage*; BG Teubner GMBH Stuttgart, Leipzig, Wiesbaden; 2002
- [Heyde 1997] Heyde, Christopher, C.: *Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation* ; Springer, New York; 1997

- [Jiang 2007] Jiang, Jiming: *Linear and Generalized Linear Mixed Models and Their Applications*; Springer (Series in Statistics); 2007
- [Jiang et al. 2008] Jiang, Jiming; Rao, J. Sunil; Gu, Zhonghua; Nguyen, Thuan : *Fence methods for mixed model selection*; In: *Annals of Statistics*, Volume 36, Number 4 , Seiten 1669-1692; 2008
- [Jones 1993] Jones, R.H.: *Longitudinal Data with Serial Correlation: A State-Space*; Approach. London: Chapman and Hall; 1993
- [Kachmann] Kachman, Stephen D.: *An introduction to generalized linear mixed models*; Department of Biometry, University of Nebraska Lincoln
<http://statistics.unl.edu/faculty/steve/glmm/paper.pdf>
Stand vom 24.04.2008
- [Krueger und Tian 2004] Krueger, Charlene; Tian, Lili : *A Comparison of the General Linear Mixed Model and Repeated Measures ANOVA Using a Dataset with Multiple Missing Data Poin*;
In: *Biological Research For Nursing*, Vol. 6, No. 2, Seiten 151-157; 2004
- [Liang und Zeger 1986] Liang, Kung-Yee ; Zeger, Scott L.: *Longitudinal data analysis using generalized linear models*;
In: *Biometrika* 73-1, Vol. 73, No. 1, Seiten 13-22; 1986
- [Liard und Ware 1982] Liard, N.M.; Ware, J.H.: *Random effects models for longitudinal data*;
In: *Biometrics*, 38, Seiten 963- 974; 1982
- [Lin 1997] Lin, Xihong: *Variance Component Testing in Generalised Linear Models with Random Effects*;
In: *Biometrika*, Vol. 84, No. 2, Seiten 309-326; 1997
- [McCullagh und Nelder 1989] McCullagh, P. ; Nelder, J. A. ; Cox, D. R. (Hrsg.) ; Hinkley, D. V. (Hrsg.) ; Reid, N. (Hrsg.) ; Rubin, D. R. (Hrsg.) ; Silverman, B.W. (Hrsg.): *Monographs on Statistics and Applied Probability. Bd. 37: Generalized Linear Models. zweite Ausgabe*; Chapman and Hall; 1989

- [Molenberghs und Verbeke 2004] Molenberghs, Geert ; Verbeke, Geert: *Longitudinal and Incomplete Data Analysis*;
In: International Biometric Conference, Cairns, Australia; 10 July 2004
- [Molenberghs und Verbeke 2005] Molenberghs, Geert ; Verbeke, Geert: *Models for Discrete Longitudinal Data*; Springer; 2005
- [Rubin 1976] Rubin,, Donald B.: *Inference and missing data*; In: Biometrika, Vol. 63, Seiten 581-592; 1976
- [Runte] Runte, Matthias : *Missing Values, Konzepte und statistische Literatur*; (<http://www.runte.de/matthias/publications/missingvalues.pdf>
Stand vom 22.4.2008)
- [SAS/STAT 2004] *SAS/STAT 9.1 User's Guide*; SAS Institute Inc., Cary, NC: USA; 2004
- [SAS/STAT 2006] *The Glimmix Procedure*; SAS Institute Inc., Cary, NC: USA; 2006
- [Schmidt,Lang,Thews 2005] Schmidt, Robert F. ; Lang, Florian; Thews, Gerhard: *Physiologie des Menschen mit Pathophysiologie; 29. Auflage*; Springer Medizin Verlag Heidelberg; 2005
- [Schwarz 1978] Schwarz, Gideon : *Estimating the Dimension of a Model*; In: The Annals of Statistics, Vol.6, No.2, Seiten 461-464; 1978
- [Verbeke und Molenberghs 2000] Verbeke, Geert; Molenberghs, Geert: *Linear Mixed Models for Longitudinal Data*; Springer Verlag New York, Inc.; 2000
- [Voss 2000] Voss, Werner u.a.: *Taschenbuch der Statistik*; Fachbuchverlag Leipzig im Carl Hanser Verlag; 2000
- [Wedderburn 1974] Wedderburn, R. W. M.: *Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method*; In: Biometrika, Vol. 61, No. 3 Seiten 439-447; 1974

Erklärung

Hiermit versichere ich, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt habe.

Oldenburg, 25. November 2008