

Statistische Modellierung III

-Gemischte Modelle-

Dr. Martin Scharpenberg

MSc Medical Biometry/Biostatistics

WiSe 2019/2020

Setup

Setup

- Betrachten nun wiederholte Messungen
- Wiederholte Messungen entstehen, wenn eine Response wiederholt an einer Menge von Messeinheiten erhoben wird
- Messeinheiten in diesem Sinne können sein:
 - Personen, Patienten, etc.
 - Tiere, Pflanzen, etc.
 - Cluster: Familien, Schulklassen, Städte, etc.
 - etc.
- Spezialfall: Longitudinaldaten

Setup

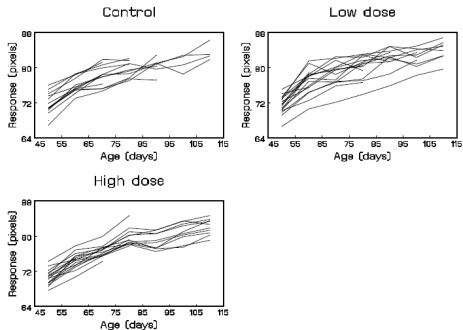
- Wiederholte Messungen an den gleichen Einheiten sind nicht unabhängig
- In der Praxis häufig ungleich viele Messungen pro Einheit und unterschiedliche Messzeitpunkte zwischen den Einheiten
- Herkömmliche Regressionsmethoden sind häufig nicht anwendbar
- Benötigen Modelle, die diese Unterschiede und die spezifischen Abhängigkeitsstrukturen berücksichtigen
- Werden dies mit sog. „random effects“ tun
- Dieser Foliensatz beruht wesentlich auf dem Buch: Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer Series in Statistics. New-York: Springer.

Beispieldaten

Beispieldaten – Ratten

- Untersucht werden soll der Einfluss von Testosteron auf das kraniofaziale Wachstum bei männlichen Wistar-Ratten
- 50 männliche Wistar-Ratten werden wie folgt randomisiert:
 - Kontrolle (15 Ratten)
 - Niedrige Dosis Decapeptyl (18 Ratten)
 - Hohe Dosis Decapeptyl (17 Ratten)
- Die Behandlung beginnt bei einem Lebensalter von 45 Tagen; ab dem 50ten Lebenstag werden alle 10 Tage Messwerte erhoben
- Zielvariable sind die Abstände (in Pixeln) von vordefinierten Punkten des Schädels auf einem Röntgenbild

Beispieldaten – Ratten



Individuelle Profile der Schädelhöhe – aus Verbeke & Molenberghs

(2000)

Alter (Tage)	Kontrolle	# Beobachtungen		Total
		Niedrig	Hoch	
50	15	18	17	50
60	13	17	16	46
70	13	15	15	43
80	10	15	13	38
90	7	12	10	29
100	4	10	10	24
110	4	8	10	22

Anzahl der Beobachtungen in den Gruppen pro Zeitpunkt

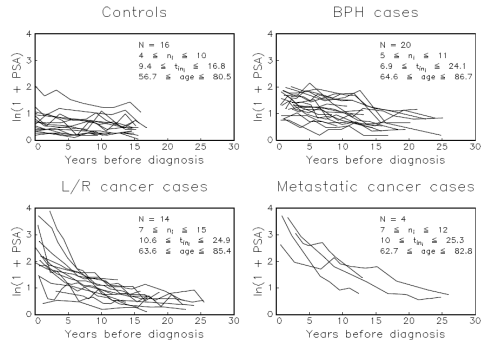
- Drop-out wegen Anästhesie
- Große Variabilität zwischen den Ratten, geringe Variabilität innerhalb der Ratten

Beispieldaten – Prostataerkrankungen

- Versuchen Marker zur Früherkennung von Prostatakrebs zu finden
- Zielvariable: PSA (Prostata-spezifisches Antigen), ein Enzym, welches von normalen und krebsartigen Prostatazellen gebildet wird
- PSA-Wert hängt vom Volumen des Prostatagewebes ab
- Problem: Patienten mit gutartiger Prostatahyperplasie (engl: Benign Prostatic Hyperplasia – BPH) haben auch einen erhöhten PSA-Wert
- Können wir trotzdem aus den longitudinalen PSA-Profilen frühzeitig Prostatakrebs erkennen?

Beispieldaten – Prostataerkrankungen

- Retrospektive Fall-Kontroll-Studie
- 16 Kontrollen, 20 BPH, 14 mit lokalem Krebs, 4 mit metastasierendem Krebs
- Alter bei Diagnose und Follow-up Zeit nicht gematcht; Dies muss berücksichtigt werden
- Wieder große Variabilität zwischen Patienten, wenig Variabilität innerhalb eines Patienten
- Große Imbalance



Individuelle Profile des transformierten PSA-Wertes – aus Verbeke
& Molenberghs (2000)

Zweistufiges Modell

Stufe 1

- Die Zufallsvariable Y_{ij} ist die (möglicherweise transformierte) Zielvariable für das i -te Individuum, gemessen zum Zeitpunkt t_{ij} , $i = 1, \dots, N$, $j = 1, \dots, n_i$
- Fassen die Beobachtungen für das i -te Individuum im n_i -dimensionalen Vektor $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})$ zusammen
- In der ersten Stufe des zweistufigen Modells, nehmen wir an dass

$$\mathbf{Y}_i = \mathbf{Z}_i \beta_i + \epsilon_i$$

gilt, wobei \mathbf{Z}_i eine $(n_i \times q)$ -Matrix mit bekannten Kovariablen, β_i ein individuenspezifischer Vektor von Regressionskoeffizienten und $\epsilon_i \sim N(\mathbf{0}, \Sigma)$ ein Residualvektor ist

- Häufig wird angenommen, dass die $\epsilon_i \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_{n_i})$ und unabhängig voneinander sind

Stufe 1

- Mit dem Modell wird also der Einfluss der Kovariablen in Z_i auf die Zielvariable Y_i für jedes Individuum einzeln modelliert
- Die Koeffizienten in β_i geben also den Einfluss der Kovariablen auf die Zielvariable für das i -te Individuum an
- Es wird also die Variabilität innerhalb eines Individuums beschrieben
- Gewünscht ist Aussage über alle Individuen hinweg

Stufe 2

- Die Variabilität zwischen den Individuen wird in der zweiten Stufe modelliert
- Das Modell der zweiten Stufe ist:

$$\beta_i = K_i \beta + \mathbf{b}_i,$$

wobei K_i eine $(q \times k)$ -Matrix mit bekannten Kovariablen, β ein Vektor von Regressionskoeffizienten und $\mathbf{b}_i \sim N(\mathbf{0}, D)$ ein Residualvektor ist

Zweistufiges Modell – Beispiel: Ratten

- Im ersten Schritt wurde die Zeitachse transformiert, um die individuellen Profile zu linearisieren:

$$t_{ij} = \ln[1 + (\text{Alter}_{ij} - 45)/10]$$

- Damit sind die Profile annähernd linear und $t = 0$ entspricht dem Start der Behandlung
- Die Matrix Z_i hat 2 Spalten: Eine enthält nur 1en und die andere die Zeitvariablen $t_{ij}, j = 1, \dots, n_i$
- Das Modell der ersten Stufe ist damit:

$$Y_{ij} = \beta_{1i} + \beta_{2i}t_{ij} + \epsilon_{ij}, \quad j = 1, \dots, n_i$$

Zweistufiges Modell – Beispiel: Ratten

- In der zweiten Stufe werden die Individuenspezifischen Intercepts und Zeit-Effekte nach der Behandlung regressiert
- Das Modell der zweiten Stufe ist dann:

$$\beta_{1i} = \beta_0 + b_{1i}$$

$$\beta_{2i} = \beta_1 L_i + \beta_2 H_i + \beta_3 C_i + b_{2i},$$

- L_i , H_i und C_i sind Indikatorvariablen:

$$L_i = \begin{cases} 1 & \text{falls niedrige Dosis} \\ 0 & \text{sonst} \end{cases} \quad H_i = \begin{cases} 1 & \text{falls hohe Dosis} \\ 0 & \text{sonst} \end{cases} \quad C_i = \begin{cases} 1 & \text{falls Kontrolle} \\ 0 & \text{sonst} \end{cases}$$

Zweistufiges Modell – Beispiel: Ratten

- Wegen der Randomisierung und der Transformation der Zeitvariable, ist die Annahme, dass die Individuenspezifischen Intercepts nicht von der Behandlung abhängen gerechtfertigt
- β_0 kann somit als mittlere Response zu Beginn der Behandlung interpretiert werden
- β_1 , β_2 und β_3 sind die mittleren Zeit-Effekte für jede einzelne Behandlung

Zweistufiges Modell – Beispiel: Prostatakrebs

- Wir nehmen an, dass jedes der Profile gut durch eine quadratische Funktion (der Zeit vor der Diagnose) approximiert werden kann
- Das Modell der ersten Stufe ist daher:

$$Y_{ij} = \ln(PSA_{ij} + 1) = \beta_{1i} + \beta_{2i}t_{ij} + \beta_{3i}t_{ij}^2 + \epsilon_{ij} \quad j = 1, \dots, n_i$$

- Im Modell der zweiten Stufe werden die so erhaltenen Koeffizienten nach der Behandlung und dem Alter bei Diagnose regressiert

Zweistufiges Modell – Beispiel: Prostatakrebs

- Wir erhalten das Modell:

$$\beta_{1i} = \beta_1 \text{Age}_i + \beta_2 C_i + \beta_3 B_i + \beta_4 L_i + \beta_5 M_i + b_{1i},$$

$$\beta_{2i} = \beta_6 \text{Age}_i + \beta_7 C_i + \beta_8 B_i + \beta_9 L_i + \beta_{10} M_i + b_{2i},$$

$$\beta_{3i} = \beta_{11} \text{Age}_i + \beta_{12} C_i + \beta_{13} B_i + \beta_{14} L_i + \beta_{15} M_i + b_{3i},$$

wobei C_i , B_i , L_i und M_i Indikatorvariablen für Kontroll-, BPH-, lokale Krebs- bzw. metastasierender Krebs-Patienten sind

- β_2 , β_3 , β_4 , β_5 sind die mittleren Intercepts für Kontroll-, BPH-, ... Patienten, wenn auf Alter korrigiert wird
- Die restlichen Koeffizienten haben analoge Interpretationen

Zweistufige Datenanalyse

- In der Praxis kann man das Modell der ersten Stufe rechnen und erhält die geschätzten Koeffizienten $\hat{\beta}_i$
- Diese können dann in die Modellgleichungen der zweiten Stufe eingesetzt werden
- Dann wird auch das Modell der zweiten Stufe an die Schätzer aus der ersten Stufe angepasst und man erhält Schätzer für β

Zweistufige Datenanalyse – Nachteile

- Die Zweistufige Analyse von Daten birgt Nachteile
- Man wertet im zweiten Schritt zusammengefasste Daten der ersten Stufe aus
- Beim Übergang von \mathbf{Y}_i zu $\hat{\beta}_i$ gehen Informationen verloren
- Durch das einsetzen von $\hat{\beta}_i$ für β_i wird zufällige Variabilität erzeugt
- Die Kovarianzmatrix von $\hat{\beta}_i$ hängt stark von der Zahl der Beobachtungen und den Beobachtungszeitpunkten für das i -te Individuum ab
- Die Kovarianzmatrix von $\hat{\beta}_i$ wird in der Analyse des zweiten Schrittes nicht berücksichtigt

Das allgemeine gemischte lineare Modell

Allgemeines gemischtes Modell

- Die Modelle des 2-Stufigen Ansatz lauteten:

$$\mathbf{Y}_i = \mathbf{Z}_i \beta_i + \epsilon_i \quad \text{und} \quad \beta_i = \mathbf{K}_i \beta + \mathbf{b}_i$$

- Kombiniert man die beiden Stufen, erhält man

$$\mathbf{Y}_i = \mathbf{Z}_i \mathbf{K}_i \beta + \mathbf{Z}_i \mathbf{b}_i + \epsilon_i = \mathbf{X}_i \beta + \mathbf{Z}_i \mathbf{b}_i + \epsilon_i,$$

wobei $\mathbf{X}_i = \mathbf{Z}_i \mathbf{K}_i$ die $(n_i \times k)$ -Matrix der bekannten Kovariablen ist

- Dieses Modell wird gemischtes lineares Modell mit fixen Effekten β zufälligen Effekten (random effects) \mathbf{b}_i genannt

Allgemeines gemischtes Modell

- Das allgemeine lineare gemischte Modell lässt sich also schreiben als:

$$\mathbf{Y}_i = X_i\beta + Z_i\mathbf{b}_i + \epsilon_i,$$

wobei

$$\mathbf{b}_i \sim N(\mathbf{0}, D) \quad \text{und} \quad \epsilon_i \sim N(0, \Sigma_i)$$

und $\mathbf{b}_1, \dots, \mathbf{b}_N, \epsilon_1, \dots, \epsilon_N$ stoch. unabhängig sind

- Dies kann umgeschrieben werden als

$$\mathbf{Y}_i | \mathbf{b}_i = N(X_i\beta + Z_i\mathbf{b}_i, \Sigma_i), \quad \mathbf{b}_i \sim N(\mathbf{0}, D)$$

- Manchmal wird das Modell direkt so definiert (ohne den Weg über das 2-stufige Verfahren)

Allgemeines gemischtes Modell

- Man nennt dieses Modell ein **hierarchisches Modell**
- Häufig nimmt man an, dass Σ_i nur über die Dimension von i abhängt, d.h. dass die unbekannten Parameter in Σ_i von i unabhängig sind
- Aus dem hierarchischen Modell folgt folgendes **marginale Modell**:

$$\mathbf{Y}_i \sim N(\mathbf{X}_i\beta, \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \Sigma_i)$$

- Das hierarchische Modell impliziert das marginale Modell, die Umkehrung gilt **nicht** allgemein
- Notation: $\mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^T + \Sigma_i$

Allgemeines gemischtes Modell – Beispiel: Ratten I

- Modell der ersten Stufe:

$$Y_{ij} = \beta_{1i} + \beta_{2i}t_{ij} + \epsilon_{ij}, \quad j = 1, \dots, n_i$$

- Modell der zweiten Stufe:

$$\beta_{1i} = \beta_0 + b_{1i}$$

$$\beta_{2i} = \beta_1 L_i + \beta_2 H_i + \beta_3 C_i + b_{2i},$$

- Zusammen ergibt sich:

$$Y_{ij} = (\beta_0 + b_{1i}) + (\beta_1 L_i + \beta_2 H_i + \beta_3 C_i + b_{2i})t_{ij} + \epsilon_{ij}, \quad j = 1, \dots, n_i$$

Allgemeines gemischtes Modell – Beispiel: Ratten I

- Dieses Modell lässt sich umformulieren zu:

$$Y_{ij} = \begin{cases} \beta_0 + b_{1i} + (\beta_1 + b_{2i})t_{ij} + \epsilon_{ij} & \text{falls niedrige Dosis} \\ \beta_0 + b_{1i} + (\beta_2 + b_{2i})t_{ij} + \epsilon_{ij} & \text{falls hohe Dosis} \\ \beta_0 + b_{1i} + (\beta_3 + b_{2i})t_{ij} + \epsilon_{ij} & \text{falls Kontrolle} \end{cases}$$

- Die Erwartungswertstruktur des durch das hierarchische Modell implizierte marginale Modell ist also:
 - Linearer mittlerer Verlauf in jeder Gruppe
 - Gleiche Intercepts in jeder Gruppe
 - Unterschiedliche mittlere Steigungen in den Gruppen

Allgemeines gemischtes Modell – Beispiel: Ratten I

- Wollen auch die implizierte marginale Kovarianzstruktur betrachten
- Nehmen dazu $\Sigma_i = \sigma^2 \mathbf{I}_{n_i}$ an
- Dann ist

$$\begin{aligned} \text{Cov}(\mathbf{Y}_i(t_1), \mathbf{Y}_i(t_2)) &= \begin{pmatrix} 1 & t_1 \end{pmatrix} D \begin{pmatrix} 1 \\ t_2 \end{pmatrix} + \sigma^2 \mathbb{1}_{\{t_1=t_2\}} \\ &= d_{22} t_1 t_2 + d_{12}(t_1 + t_2) + d_{11} + \sigma^2 \mathbb{1}_{\{t_1=t_2\}} \end{aligned}$$

- Das Modell impliziert also eine Varianzfunktion die quadratisch in der Zeit ist, wobei $d_{22} > 0$ ist (d.h. die Varianz wächst mit der Zeit)

Allgemeines gemischtes Modell – Beispiel: Ratten II

- Werden nun ein Modell betrachten, welches eine andere Kovarianzstruktur induziert
- Nehmen an, dass die Unterschiede in den Individuenspezifischen Steigungen den Behandlungen zugeschrieben werden können, können die Zufälligen Steigungen b_{i2} des obigen Modells weggelassen werden. Wir erhalten

$$\begin{aligned} Y_{ij} &= (\beta_0 + b_{i1}) + (\beta_1 L_i + \beta_2 H_i + \beta_3 C_i) t_{ij} + \epsilon_{ij} \\ &= \begin{cases} \beta_0 + b_{1i} + \beta_1 t_{ij} + \epsilon_{ij} & \text{falls niedrige Dosis} \\ \beta_0 + b_{1i} + \beta_2 t_{ij} + \epsilon_{ij} & \text{falls hohe Dosis} \\ \beta_0 + b_{1i} + \beta_3 t_{ij} + \epsilon_{ij} & \text{falls Kontrolle} \end{cases} \end{aligned}$$

- Man nennt dies das random-intercepts Modell

Allgemeines gemischtes Modell – Beispiel: Ratten II

- Die durch das Modell induzierte marginale Erwartungswertstruktur ist dieselbe wie zuvor
- Mit $\Sigma_i = \sigma^2 \mathbf{I}_{n_i}$ gilt:

$$\begin{aligned}\text{Cov}(\mathbf{Y}_i(t_1), \mathbf{Y}_i(t_2)) &= \begin{pmatrix} 1 \end{pmatrix} D \begin{pmatrix} 1 \end{pmatrix} + \sigma^2 \mathbb{1}_{\{t_1=t_2\}} \\ &= d_{11} + \sigma^2 \mathbb{1}_{\{t_1=t_2\}}\end{aligned}$$

- Man nennt diese Kovarianzstruktur *compound symmetry*:
 - Die Varianz ist konstant $d_{11} + \sigma^2$
 - Die Korrelation zwischen zwei Messungen am gleichen Individuum ist unabhängig von der Zeit $\rho = d_{11}/(d_{11} + \sigma^2)$
 - ρ wird auch *intraclass correlation coefficient* genannt

Hierarchisches vs. marginales Modell

Hierarchisches vs. marginales Modell

- Haben bereits argumentiert, dass das hierarchische Modell das marginale impliziert und nicht umgekehrt
- Betrachten nun ein Beispiel dafür, in dem 2 hierarchische Modelle dasselbe marginale Modell implizieren
- Nehmen an, dass wir binäre Daten haben und je 2 Beobachtungen pro Individuum ($n_i = 2$)
- Je ein Erfolg und ein Misserfolg beobachtet

Hierarchisches vs. marginales Modell

- Modell 1: Zufällige Effekte für Intercept und Heteroskedasdität der ϵ_{ij} (d.h. $\Sigma_i \neq \sigma^2 \mathbf{I}_{n_i}$):

$$V_i = \begin{pmatrix} 1 \\ 1 \end{pmatrix} (d) \begin{pmatrix} 1 & 1 \end{pmatrix} + \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} d + \sigma_1^2 & d \\ d & d + \sigma_2^2 \end{pmatrix}$$

- Modell 2: Zufällige Effekte für Intercept und Steigung (unkorreliert) und Homoskedasdität der ϵ_{ij} (d.h. $\Sigma_i = \sigma^2 \mathbf{I}_{n_i}$):

$$V_i = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} d_1 & 0 \\ 0 & d_2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} = \begin{pmatrix} d_1 + \sigma^2 & d_1 \\ d_1 & d_1 + d_2 + \sigma^2 \end{pmatrix}$$

- Diese Modelle sind äquivalent ($d \hat{=} d_1$, $\sigma_1^2 \hat{=} \sigma^2$, $\sigma_2^2 \hat{=} d_2 + \sigma_1^2$)

Hierarchisches vs. marginales Modell

- Die beiden hierarchischen Modelle führen zum selben marginalen Modell
- In der Praxis wird üblicherweise mit den marginalen Modellen gerechnet (z.B. in SAS)
- Ein guter Fit des marginalen Modells kann nicht als Evidenz für eines der hierarchischen Modelle interpretiert werden
- Können die hierarchischen Modelle nur mit Methoden der Bayes-Statistik gut behandeln (nicht in dieser VL)

Schätzen im marginalen Modell

ML Schätzung

- Wollen nun die Parameter des marginalen Modells schätzen
- Notation:

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_N \end{pmatrix}, \quad \mathbf{V} = \text{diag}(V_1, \dots, V_N), \quad \mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix}$$

- Sei α der Vektor der Vektor der $q(q+1)/2$ verschiedenen Elemente in D und aller Elemente in den Σ_i
- Dann ist die Matrix \mathbf{V} abhängig von α und somit schreiben wir $\mathbf{V}(\alpha)$
- Damit lässt sich das marginale Modell schreiben als

$$\mathbf{Y} \sim N(\mathbf{X}\beta, \mathbf{V}(\alpha))$$

ML Schätzung

- Die Log-Likelihood-Kern ist also gegeben als:

$$l(\beta, \alpha) = -\frac{1}{2} \left\{ |\mathbf{V}(\alpha)| + (\mathbf{Y} - \mathbf{X}\beta)^T \mathbf{V}(\alpha)^{-1} (\mathbf{Y} - \mathbf{X}\beta) \right\}$$

- Wenn α gegeben ist, wissen wir (da es sich um ein allgemeines lineares Modell handelt), dass der ML-Schätzer für β gegeben ist als der Aitken-Schätzer:

$$\hat{\beta}(\alpha) = (\mathbf{X}^T \mathbf{V}(\alpha)^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}(\alpha)^{-1} \mathbf{Y}$$

- Können den MLE $\hat{\alpha}$ für α bestimmen, indem wir die Formel für $\hat{\beta}$ in den Log-Likelihood-Kern einsetzen und die so erhaltene Funktion in α maximieren (Nullstelle des Scores z.B. via Fisher-Scoring)
- Der MLE für β ergibt sich dann durch einsetzen von $\hat{\alpha}$ in obige Formel ($\hat{\beta}(\hat{\alpha})$)

REML Schätzung

- Wie der MLE von σ^2 im linearen Modell, ist auch der MLE $\mathbf{V}(\hat{\alpha})$ kein unverzerrter Schätzer für $\mathbf{V}(\alpha)$
- Können den Bias der Schätzung verringern, indem wir die sog. *restricted maximum likelihood* Methode benutzen
- Notation: $\sum_{i=1}^N n_i = n$, Gesamtzahl der Beobachtungen
- Betrachten nichtmehr \mathbf{Y} selbst, sondern $\mathbf{U} = \mathbf{A}^T \mathbf{Y}$, wobei \mathbf{A} eine $(n \times (n - k))$ -Matrix ist, deren Spalten orthogonal zu den Spalten in \mathbf{X} sind
- Es folgt $\mathbf{U} \sim N(\mathbf{0}, \mathbf{A}^T \mathbf{V}(\alpha) \mathbf{A})$
- Man kann nun also α schätzen ohne Kenntnis über β zu haben

REML Schätzung

- Der Log-Likelihood-Kern des transformierten Datenvektors \mathbf{U} ist gegeben durch:

$$\tilde{l}(\alpha) = -\frac{1}{2} \left\{ |\mathbf{A}^T \mathbf{V}(\alpha) \mathbf{A}| + \mathbf{U}^T \mathbf{V}(\alpha)^{-1} \mathbf{U} \right\}$$

- Hieraus lässt sich wieder der Score ableiten, dessen Nullstelle, der REML-Schätzer $\hat{\alpha}_{REML}$, über das Fisher-Scoring Verfahren bestimmt werden kann
- Der entsprechende Schätzer für β ergibt sich dann aus

$$\hat{\beta}_{REML} = \hat{\beta}(\hat{\alpha}_{REML}) = (\mathbf{X}^T \mathbf{V}(\hat{\alpha}_{REML})^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}(\hat{\alpha}_{REML})^{-1} \mathbf{Y}$$

- Man kann zeigen, dass der REML-Schätzer $\hat{\alpha}_{REML}$ nicht von der Wahl der Matrix \mathbf{A} abhängt

Beispiel: Prostatadaten

- Wir betrachten erneut das bereits eingeführte Modell für die Prostatadaten
- Wir nehmen an, dass jedes der Profile gut durch eine quadratische Funktion (der Zeit vor der Diagnose) approximiert werden kann
- Für dieses Beispiel nehmen wir darüber hinaus an, dass $\Sigma_i = \sigma^2 I_{n_i}$ für alle i gilt
- Wir haben dann:

$$\begin{aligned} \ln(PSA_{ij} + 1) = & \beta_1 Age_i + \beta_2 C_i + \beta_3 B_i + \beta_4 L_i + \beta_5 M_i \\ & + (\beta_6 Age_i + \beta_7 C_i + \beta_8 B_i + \beta_9 L_i + \beta_{10} M_i) t_{ij} \\ & + (\beta_{11} Age_i + \beta_{12} C_i + \beta_{13} B_i + \beta_{14} L_i + \beta_{15} M_i) t_{ij}^2 \\ & + b_{1i} + b_{2i} t_{ij} + b_{3i} t_{ij}^2 + \epsilon_{ij} \end{aligned}$$

Beispiel: Prostatadaten

- Wir haben dann:

$$\begin{aligned} \ln(PSA_{ij} + 1) = & \beta_1 Age_i + \beta_2 C_i + \beta_3 B_i + \beta_4 L_i + \beta_5 M_i \\ & + (\beta_6 Age_i + \beta_7 C_i + \beta_8 B_i + \beta_9 L_i + \beta_{10} M_i) t_{ij} \\ & + (\beta_{11} Age_i + \beta_{12} C_i + \beta_{13} B_i + \beta_{14} L_i + \beta_{15} M_i) t_{ij}^2 \\ & + b_{1i} + b_{2i} t_{ij} + b_{3i} t_{ij}^2 + \epsilon_{ij} \end{aligned}$$

- In der SAS Prozedur Mixed lässt sich dies wie folgt programmieren:

```
proc mixed data=prostate method=reml;  
class id group;  
model lnpsa = group age time group*time age*time time2 group*time2 age*time2 / solution;  
random intercept time time2 / type=un subject=id;  
run;
```


Beispiel: Prostatadaten

- Die Option `type=un` im *random* Statement bedeutet, dass keine Struktur für die Kovarianzmatrix D der zufälligen Effekte angenommen wird
- Viele weitere Optionen sind hier möglich
- Im obigen Beispiel wurde $\Sigma_i = \sigma^2 I_{n_i}$ für alle i angenommen
- Über ein *repeated* Statement kann man die Form der Σ_i vorgeben
- Weiteres hierzu in der Literatur und der SAS-Hilfe

Beispiel: Prostatadaten – Ergebnisse für fixe Effekte

Effect	Parameter	MLE (s.e.)	REMLE (s.e.)
Age	β_1	0.026 (0.013)	0.027 (0.014)
Intercepts:			
Control	β_2	-1.077 (0.919)	-1.098 (0.976)
BPH	β_3	-0.493 (1.026)	-0.523 (1.090)
lok. cancer	β_4	0.314 (0.997)	0.296 (1.059)
Met. cancer	β_5	1.574 (1.022)	1.549 (1.086)
Age*time	β_6	-0.010 (0.020)	-0.011 (0.021)
Time Effects:			
Control	β_7	0.511 (1.359)	0.568 (1.437)
BPH	β_8	0.313 (1.511)	0.396 (1.638)
lok. cancer	β_9	-1.072 (1.469)	-1.036 (1.593)
Met. cancer	β_{10}	-1.657 (1.499)	-1.605 (1.626)
Age*time ²	β_{11}	0.002 (0.008)	0.002 (0.009)
Time ² Effects:			
Control	β_{12}	-0.106 (0.549)	-0.130 (0.610)
BPH	β_{13}	-0.119 (0.604)	-0.158 (0.672)
lok. cancer	β_{14}	0.350 (0.590)	0.342 (0.656)
Met. cancer	β_{15}	0.411 (0.598)	0.395 (0.666)

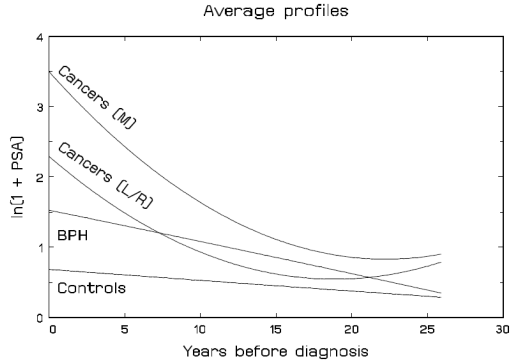
Schätzer für die fixen Effekte im Prostatadatenbeispiel

Beispiel: Prostatadaten – Ergebnisse für zufällige Effekte

Effect	Parameter	MLE (s.e.)	REMLE (s.e.)
Covariance of \mathbf{b}_i :			
$\text{var}(b_{1i})$	d_{11}	0.398 (0.083)	0.452 (0.098)
$\text{var}(b_{2i})$	d_{22}	0.768 (0.187)	0.915 (0.230)
$\text{var}(b_{3i})$	d_{33}	0.103 (0.032)	0.131 (0.041)
$\text{cov}(b_{1i}, b_{2i})$	$d_{12} = d_{21}$	-0.443 (0.113)	-0.518 (0.136)
$\text{cov}(b_{2i}, b_{3i})$	$d_{23} = d_{32}$	-0.273 (0.076)	-0.336 (0.095)
$\text{cov}(b_{1i}, b_{3i})$	$d_{13} = d_{31}$	0.133 (0.043)	0.163 (0.053)
Residual variance:			
$\text{var}(\epsilon_{ij})$	σ^2	0.028 (0.002)	0.028 (0.002)

Ergebnisse für die zufälligen Effekte im Prostatadatenbeispiel

Beispiel: Prostatadaten – Mittlere Profile



Mittlere angepasste Profile des transformierten PSA-Wertes für Männer mit medianem Alter bei Diagnose – aus Verbeke & Molenberghs (2000)

Schätzproblem: Negative Varianzkomponenten

Negative Varianzkomponenten

- Wir betrachten erneut das Beispiel der Ratten. Hier war das angepasste Modell:

$$Y_{ij} = (\beta_0 + b_{1i}) + (\beta_1 L_i + \beta_2 H_i + \beta_3 C_i)t_{ij} + \epsilon_{ij}$$

- Passt man dieses Modell an die Daten an, erhält man

Effect	Parameter	REMLE (s.e.)
Intercept	β_0	68.606 (0.325)
Time effects:		
Low Dose	β_1	7.503 (0.228)
High Dose	β_2	6.887 (0.231)
Control	β_3	7.319 (0.285)
Covariance of \mathbf{b}_i :		
$\text{var}(b_{1i})$	d_{11}	0.369 (1.123)
$\text{var}(b_{2i})$	d_{22}	0.000 (—)
$\text{cov}(b_{1i}, b_{2i})$	$d_{12} = d_{21}$	0.090 (0.381)
Residual variance:		
$\text{var}(\epsilon_{ij})$	σ^2	1.445 (0.145)
REML log-likelihood		-466.173

Negative Varianzkomponenten

- Für $d_{22} = \text{var}(b_{2i})$ wird ein Schätzwert von 0 und kein Standardfehler ausgegeben
- Die SAS Prozedur Proc Mixed maximiert die REML log-likelihood unter den Bedingungen, dass die Diagonalelemente von D , sowie σ^2 positiv sind
- Dies ist dadurch gerechtfertigt, dass diese Parameter im hierarchischen Modell als Varianzen der Zufallsterme interpretiert werden können
- Im marginalen Modell haben die Parameter nichtmehr zwangsläufig die Interpretation von Varianzen
- Die Vermutung liegt nahe, dass die REML log-likelihood weiter maximiert werden kann, wenn man negative Varianzkomponenten zulässt

Negative Varianzkomponenten

- Lässt man auch negative Varianzkomponenten zu (Option *nobound* in Proc Mixed) ergibt sich folgendes Resultat:

Effect	Parameter	Restrictions	
		$d_{ij} \geq 0, \sigma^2 \geq 0$	$d_{ij} \in \mathbb{R}, \sigma^2 \in \mathbb{R}$
Effect	Parameter	REMLE (s.e.)	REMLE (s.e.)
Intercept	β_0	68.606 (0.325)	68.618 (0.313)
Time effects:			
Low Dose	β_1	7.503 (0.228)	7.475 (0.198)
High Dose	β_2	6.887 (0.231)	6.890 (0.189)
Control	β_3	7.319 (0.285)	7.284 (0.254)
Covariance of \mathbf{b}_j :			
$\text{var}(b_{1i})$	d_{11}	0.369 (1.123)	2.921 (1.019)
$\text{var}(b_{2i})$	d_{22}	0.000 (—)	-0.287 (0.169)
$\text{cov}(b_{1i}, b_{2i})$	$d_{12} = d_{21}$	0.090 (0.381)	0.462 (0.357)
Residual variance:			
$\text{var}(\epsilon_{ij})$	σ^2	1.445 (0.145)	1.522 (0.65)
REML log-likelihood		-466.173	-465.193

Negative Varianzkomponenten

- Die REML log-likelihood konnte weiter maximiert werden
- Erhalten einen negativen Schätzer für d_{22}
- Mit diesem Wert für d_{22} ist eine hierarchische Interpretation nicht länger möglich
- Wir erhalten jedoch ein gültiges marginales Modell, solange die marginalen Kovarianzen $V_i = Z_i D Z_i^T + \sigma^2 I_{n_i}$ positiv (semi-) definit sind
- Dies ist der Fall, wenn sich die Kovariablen Z_i in einem bestimmten Bereich realisieren
- Dies zeigt erneut: hierarchische und marginale Modelle sind nicht äquivalent!

Interpretation der negativen Varianzkomponenten

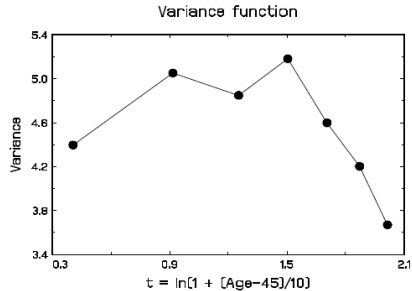
- Im Beispiel mit der negativen Varianzkomponente ist die geschätzte Varianzfunktion

$$\begin{aligned}\text{Var}(\mathbf{Y}_i(t)) &= \begin{pmatrix} 1 & t \end{pmatrix} \hat{D} \begin{pmatrix} 1 \\ t \end{pmatrix} + \hat{\sigma}^2 \\ &= \hat{d}_{22}t^2 + 2\hat{d}_{12}t + \hat{d}_{11} + \hat{\sigma}^2 \\ &= -0.287t^2 + 0.924t + 4.443\end{aligned}$$

- Die geschätzte Varianzfunktion ist also eine nach unten geöffnete Parabel
- Zur informellen Prüfung betrachten wir die geschätzte Varianzfunktion eines herkömmlichen linearen Regressionsmodells mit der gleichen Erwartungswertstruktur wie das hier angepasste marginale Modell
- Ignorieren damit die Korrelationsstruktur der Daten

Interpretation der negativen Varianzkomponenten

- Die so erhaltene Varianz lässt sich wie folgt darstellen:



Sample Varianz für herkömmliche KQ-Schätzung – aus Verbeke & Molenberghs (2000)

- Die vorhergehenden Betrachtungen werden dadurch weiter gestützt

Inferenz im marginalen Modell

Generelles zur Inferenz in gemischten Modellen

- In der Regel unterscheidet man zwischen der Inferenz für das marginale Modell und das hierarchische Modell
- Im marginalen Modell können wir Inferenz für die fixen Effekte und die Varianzkomponenten machen
- Im hierarchischen Modell gibt es Methoden zur Inferenz für die zufälligen Effekte
- In dieser VL beschränken wir uns auf die Inferenz für fixe Effekte im marginalen Modell

Inferenz für fixe Effekte

- Wie bereits diskutiert, wird der Vektor der fixen Effekte β geschätzt durch:

$$\hat{\beta}(\alpha) = (\mathbf{X}^T \mathbf{V}(\alpha)^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}(\alpha)^{-1} \mathbf{Y}$$

- Der unbekannte Vektor der Varianzkomponenten α wird durch seinen ML- oder REML-Schätzer ersetzt
- Man kann zeigen, dass unter dem marginalen Modell und bei gegebenem α :

$$\hat{\beta}(\alpha) \stackrel{a}{\sim} N(\beta, (\mathbf{X}^T \mathbf{V}(\alpha)^{-1} \mathbf{X})^{-1})$$

- In der Praxis wird die Kovarianzmatrix geschätzt, indem α wieder durch den ML- oder REML-Schätzer ersetzt wird

Wald Test für fixe Effekte

- Sind nun wieder interessiert lineare Hypothesen der Form

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

zu Testen

- Aus der Normalverteilungseigenschaft des ML-Schätzers $\hat{\beta}(\alpha)$ folgt, dass

$$W = (\mathbf{C}\hat{\beta} - \mathbf{d})^T [\mathbf{C}(\mathbf{X}^T \mathbf{V}(\alpha)^{-1} \mathbf{X})^{-1} \mathbf{C}^T]^{-1} (\mathbf{C}\hat{\beta} - \mathbf{d}) \stackrel{a}{\sim} \chi_r^2$$

mit $r = \text{Rang}(\mathbf{C})$

- Wir können H_0 verwerfen, falls $W \geq Q_r^{\chi^2}(1 - \alpha)$
- In der Praxis ersetzt man α wieder durch seinen ML- oder REML-Schätzer

Approximativer F-Test für fixe Effekte

- Die geschätzten Standardfehler der Wald Teststatistik berücksichtigen nicht, dass in $\hat{\beta}(\hat{\alpha})$ das α geschätzt wird
- Es kommt in der Regel zu einer Unterschätzung der tatsächlichen Variabilität von $\hat{\beta}(\hat{\alpha})$
- In der Praxis versucht man dies dadurch zu kompensieren, dass man approximative F-Statistiken zum Testen der linearen Hypothesen benutzt

Approximativer F-Test für fixe Effekte

- Die entsprechende F-Statistik ist:

$$F = \frac{(\mathbf{C}\hat{\beta} - \mathbf{d})^T [\mathbf{C}(\mathbf{X}^T \mathbf{V}(\alpha)^{-1} \mathbf{X})^{-1} \mathbf{C}^T]^{-1} (\mathbf{C}\hat{\beta} - \mathbf{d})}{Rang(\mathbf{C})}$$

- Die Zählerfreiheitsgrade sind $Rang(\mathbf{C})$, die Nennerfreiheitsgrade müssen geschätzt werden
- Es existieren verschiedene Verfahren die Nennerfreiheitsgrade zu schätzen (auch in SAS implementiert)
- Dies ist nicht Gegenstand dieser Vorlesung

Likelihood-Ratio-Test für fixe Effekte

- Ein alternativer Test für die linearen Hypothesen ist der Likelihood-Ratio-Test
- Die Teststatistik ist gegeben durch

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

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

- Unter der Nullhypothese ist $lq \stackrel{a}{\sim} \chi_r^2$ mit $r = \text{Rang}(\mathbf{C})$. Wir können H_0 also verwerfen, falls $W \geq Q_r^{\chi^2}(1 - \alpha)$
- **Achtung:** Das Resultat für den Likelihood-Quotienten-Test ist nur gültig, wenn die Modelle mit ML und nicht mit REML angepasst werden!