

Inhalt

- 1 Das lineare gemischte Modell
- 2 Likelihood-Schätzung für lineare gemischte Modelle
- 3 Likelihood-Inferenz im linearen gemischten Modell
- 4 Bayes-Schätzung für lineare gemischte Modelle
- 5 Additive gemischte Modelle
- 6 Das generalisierte lineare gemischte Modell
 - Wiederholung: GLM und LMM
 - $GLMM = GLM + LMM$
 - GLMM für Longitudinal-/Clusterdaten
 - GLMM in allgemeiner Form
 - GLMM: Marginale und konditionale Verteilung
 - Generalisierte Additive Gemischte Modelle (GAMMs)

Generalisiertes Lineares Modell (GLM): Strukturannahme

- **Strukturannahme:** Verknüpfung von Erwartungswert und linearem Prädiktor $\eta_i = \mathbf{x}_i' \boldsymbol{\beta}$ durch die Linkfunktion $g(\cdot)$:

$$E(y_i | \mathbf{x}_i) = g^{-1}(\eta_i) \Leftrightarrow g(E(y_i | \mathbf{x}_i)) = \eta_i$$

- Binäre Daten: z.B. Logitmodell: $E(y_i | \mathbf{x}_i) = P(y_i = 1 | \mathbf{x}_i) = \frac{\exp(\mathbf{x}_i' \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i' \boldsymbol{\beta})}$
- Zähldaten: z.B. log-lineares Modell: $E(y_i | \mathbf{x}_i) = \exp(\mathbf{x}_i' \boldsymbol{\beta})$

Generalisiertes Lineares Modell (GLM): Verteilungsannahme

- **Verteilungsannahme:** Gegeben η_i sind die y_i unabhängig verteilt und ihre Dichte $f(y_i)$ gehört zu einer Exponentialfamilie:

$$f(y_i|\theta_i) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{\phi_i} - c(y_i, \phi_i)\right)$$

$$\text{mit } E(y_i|\mathbf{x}_i) = b'(\theta_i) = g^{-1}(\mathbf{x}_i'\boldsymbol{\beta}); \text{Var}(y_i|\mathbf{x}_i) = \phi b''(\theta_i);$$

- θ_i : natürlicher/kanonischer Parameter
- ϕ_i : Skalenparameter ($\phi_i = 1$ für Binär-/Poisson-Daten)

Die Exponentialfamilie - Beispiele

Dichten / Wahrscheinlichkeitsfunktionen:

$$\text{Normal : } f(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - \mu)^2}{2\sigma^2}\right)$$

$$\text{Binomial : } f(y|\pi) = \binom{n}{y} \pi^y (1 - \pi)^{n-y}$$

$$\text{Poisson : } f(y|\lambda) = \frac{\lambda^y}{y!} \exp(-\lambda)$$

	μ	θ	ϕ	$b(\theta)$	$c(y, \phi)$
Normal	μ	μ	σ^2	$\frac{1}{2}\theta^2$	$\frac{y^2}{2\phi} - \frac{1}{2} \log(2\pi\phi)$
Binomial	π	$\log\left(\frac{\pi}{1-\pi}\right)$	1	$n \log(1 + \exp(\theta))$	$\log \binom{n}{y}$
Poisson	λ	$\log(\lambda)$	1	$\exp(\theta)$	$-y!$

GLM

- Strukturannahme:

$$E(y_i | \mathbf{x}_i) = g^{-1}(\mathbf{x}_i' \boldsymbol{\beta})$$

- Verteilungsannahme:

$$f(y_i | \theta_i) = \exp \left(\frac{y_i \theta_i - b(\theta_i)}{\phi_i} - c(y_i, \phi_i) \right)$$

⇒ Verknüpfung von Erwartungswertstruktur und Varianzstruktur

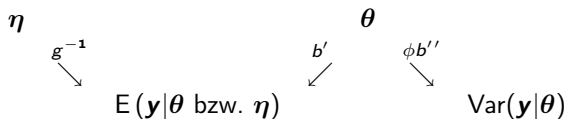
- $E(y_i | \mathbf{x}_i) = g^{-1}(\mathbf{x}_i' \boldsymbol{\beta}) = b'(\theta_i)$
- $\text{Var}(y_i | \mathbf{x}_i) = \phi_i b''(\theta_i)$

GLM

$$f(y_i|\theta_i) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{\phi_i} - c(y_i, \phi_i)\right)$$

- $E(y_i|\mathbf{x}_i) = g^{-1}(\mathbf{x}_i'\boldsymbol{\beta}) = b'(\theta_i)$
- $\text{Var}(y_i|\mathbf{x}_i) = \phi_i b''(\theta_i)$

Also:

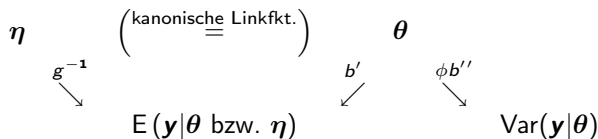


GLM mit kanonischem Link

$$f(y_i|\theta_i) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{\phi} - c(y_i, \phi)\right)$$

- $E(y_i|\mathbf{x}_i) = g^{-1}(\mathbf{x}'_i\boldsymbol{\beta}) = b'(\theta_i)$
- $\text{Var}(y_i|\mathbf{x}_i) = \phi b''(\theta_i)$
- kanonische Linkfunktion: $g^{-1}(\cdot) = b'(\cdot)$

Also:



- Beispiele: $g(\cdot) = \text{id}(\cdot)$ für Normalverteilung, $g(\cdot) = \text{logit}(\cdot)$ für Bernoulliverteilung, $g(\cdot) = \text{log}(\cdot)$ für Poissonverteilung

LM \Rightarrow LMM

- Was ist anders/neu im gemischten Modell?
- Wozu gemischte Modelle?
- Warum sind gemischte Modelle so beliebt?

Formal: Was ist anders/neu im LMM?

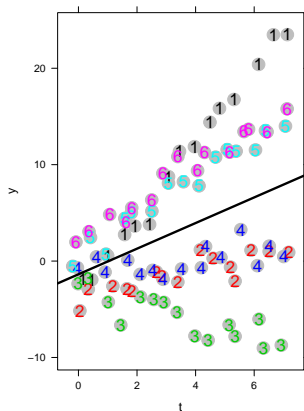
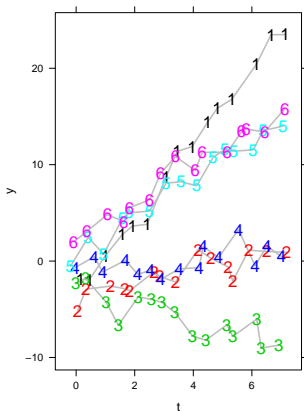
- Erweiterung des linearen Prädiktors um zufällige/penalisierte Effekte \mathbf{b} :
 $\eta = \mathbf{X}\beta + \mathbf{Zb}$
- Verteilungsannahme über \mathbf{b} : $\mathbf{b} \sim \mathcal{P}(\vartheta)$
üblich: $\mathbf{b} \sim N(\mathbf{0}, \mathbf{G}(\vartheta))$
- Inferenz über gemeinsame / penalisierte (Log-)Likelihood:
 $l_{pen}(\mathbf{y}|\beta, \mathbf{b}, \vartheta) = l(\mathbf{y}|\beta, \mathbf{b}, \vartheta) + l(\mathbf{b}|\vartheta)$
- konditionales Modell / marginales Modell:
 \mathbf{Z} und $\mathbf{G}(\vartheta)$ bestimmen marginale Kovarianzstruktur von \mathbf{y} .

$$\begin{aligned} \mathbf{y}|\mathbf{b} &\sim N(\mathbf{X}\beta + \mathbf{Zb}, \mathbf{R}(\vartheta)); \mathbf{b} \sim N(\mathbf{0}, \mathbf{G}(\vartheta)) \\ \Rightarrow \mathbf{y} &\sim N(\mathbf{X}\beta, \mathbf{ZG}(\vartheta)\mathbf{Z}' + \mathbf{R}(\vartheta)) \end{aligned}$$

Inhaltlich: Wozu LMMs?

Konditional betrachtet:

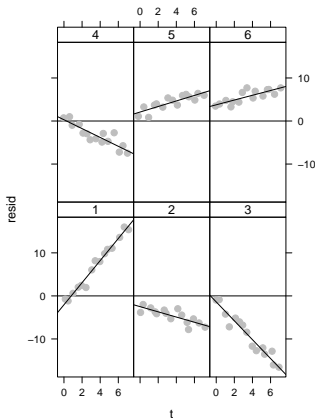
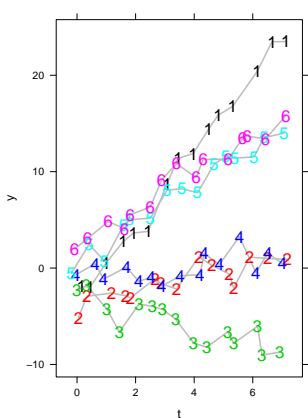
(zeitliche/räumliche/hierarchische) Struktur der Daten legt bestimmte Struktur der Abweichungen der Beobachtungen y von $X\beta$ nahe:



Inhaltlich: Wozu LMMs?

Konditional betrachtet:

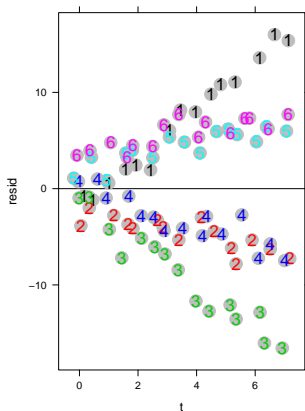
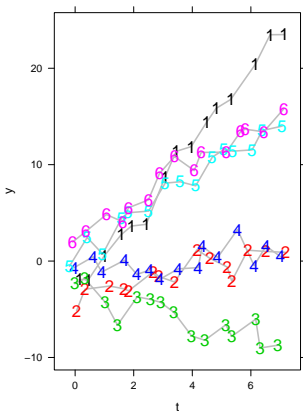
(zeitliche/räumliche/hierarchische) Struktur der Daten legt bestimmte Struktur der Abweichungen der Beobachtungen \mathbf{y} von $\mathbf{X}\beta$ nahe:



Inhaltlich: Wozu LMMs?

Marginal betrachtet:

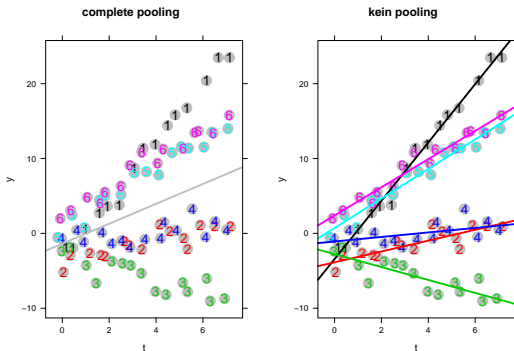
(zeitliche/räumliche/hierarchische) Struktur der Daten legt bestimmte Kovarianzstruktur der Beobachtungen nahe:



Warum sind LMMs so beliebt?

LMM bilden *Kompromiss* zwischen *komplettem pooling* und *keinem pooling*

- *komplettes pooling*: gewöhnliches LM ohne Einbezug der Gruppierungsstruktur
- *kein pooling*: separates LM für jede Ausprägung der Gruppierung



Warum sind LMMs so beliebt?

- LMM für hierarchische Daten als Kompromiss zwischen *komplettem pooling* und *keinem pooling*: Schätzung der Varianzparameter ϑ erfasst Heterogenität zwischen Gruppen auf Basis des gesamten Datensatzes \rightarrow Shrinkage der Koeffizienten zum Populationsmittel entsprechend stärker oder schwächer.
- penalisiertes Likelihoodkriterium stabilisiert Parameterschätzung (aber: Bias-Varianz-Tradeoff!)
- bayesianisch: priori-Annahmen über \mathbf{b} bilden Vorwissen über Struktur der Daten ab. Einführung dieser zusätzlichen Information erlaubt stabile Schätzung von Modellen mit sehr vielen Parametern.
- sehr vielseitig: Modelle für zeitliche, räumliche, hierarchische und gruppierte Datenstrukturen sowie nichtlineare Effekte lassen sich als (G)LMM auffassen.

LMM \Rightarrow GLMM: Flexibilisierung der Verteilungsannahme

- LMM: nur normalverteilter Response: $\mathbf{y}|\boldsymbol{\eta} \sim N(\boldsymbol{\eta}, \mathbf{R})$
- GLMM: Response aus beliebiger Exponentialfamilie:

$$E(\mathbf{y}|\boldsymbol{\eta}) = g^{-1}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\eta})$$
$$\mathbf{y}|\boldsymbol{\eta} \sim \text{Expo.fam.}$$

GLMM = GLM + LMM

- 1. Stufe:
 y_i bedingt auf den erweiterten linearen Prädiktor $\mathbf{X}\beta + \mathbf{Zb}$
unabhängig verteilt nach Exponentialfamilie (analog GLM)
- 2. Stufe:
Verteilungsannahme über $\mathbf{b} \Rightarrow$ regularisierte Schätzung (analog LMM)
Interpretation der Verteilungsannahme als Strafterm oder Eigenschaft der Grundgesamtheit (freq.) oder Priori-Annahme (bayes.)

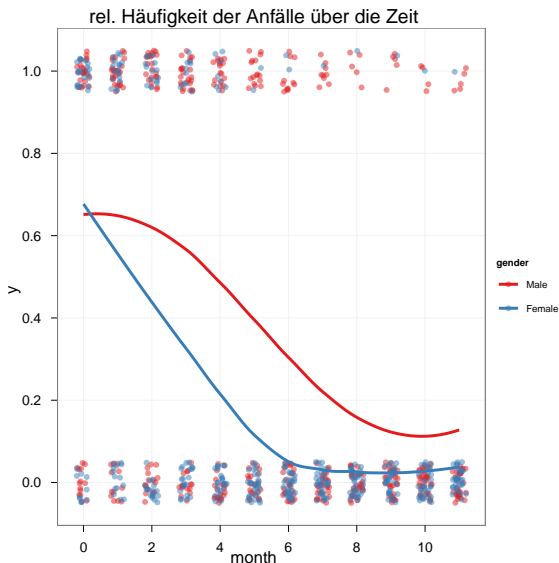
GLMM für Longitudinal-/Clusterdaten

- y_{ij} ; $i = 1, \dots, m$; $j = 1, \dots, n_i$ mit je n_i Messwiederholungen an m Beobachtungseinheiten (z.B. Binär-/Zählraten)
- Bedingte Verteilung von $y_{ij} | \mathbf{b}_i$ ist aus Exponentialfamilie (z.B. Bernoulli/Poisson)
- Bedingter Erwartungswert $\mu_{ij} = E(y_{ij} | \mathbf{b}) = g^{-1}(\mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{b}_i)$
- Die zufälligen Effekte \mathbf{b}_i sind normalverteilt,

$$\mathbf{b}_i \stackrel{iid}{\sim} N(\mathbf{0}, \mathbf{D}).$$

Beispiel: Madras-Daten

- 69 stationäre Schizophrenie-Patienten
- 1 Jahr lang nach Ersteinlieferung beobachtet
- 12 monatliche Beobachtungen pro Patient



Beispiel: Madras-Daten

- Response: $y_{it} \in 0, 1$: Patient/in i hat Schub im Monat t
- Kovariablen:
 - month: Zeit in Monaten nach Einlieferung (0-11)
 - gender: Geschlecht
 - id: Patient
- Logit-Modell mit zufälligem Intercept: → **Übung**

$$\text{logit}(P(y_{it} = 1|b_i)) = \beta_0 + b_i + \beta_M \text{month}_t + \beta_G \text{gender}_i$$

Beispiel: Madras-Daten - Ergebnisse

Random effects:

Groups Name	Variance	Std.Dev.
id (Intercept)	4.91	2.22

Number of obs: 828, groups: id, 69

Fixed effects:

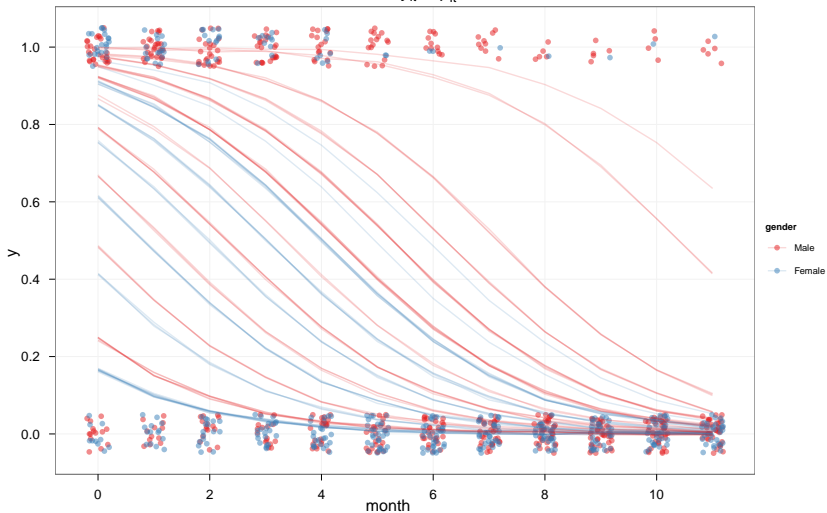
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.933	0.437	4.42	9.8e-06 ***
month	-0.573	0.045	-12.72	< 2e-16 ***
genderFemale	-1.574	0.595	-2.65	0.0081 **

```
R> quantile(exp(ranef(m1)$id$(Intercept)'), p=c(.2,.4,.6,.8))
```

20%	40%	60%	80%
0.138	0.634	2.167	5.828

⇒ relativ starke individuelle Heterogenität

Beispiel: Madras-Daten - Ergebnisse

Modell 1: y_{it} & \hat{p}_{it} 

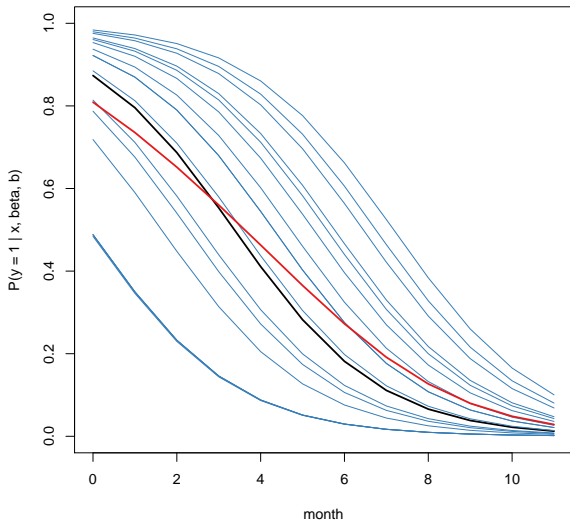
Beispiel: Madras-Daten - Interpretation

Wie sind die Parameter in diesem Logit-Modell mit Random Intercept zu interpretieren?

	beta	exp(beta)
month	-0.573	0.564

Graphisch:

- $\frac{\exp(\beta_0 + \beta_1 \text{month}_t)}{1 + \exp(\beta_0 + \beta_1 \text{month}_t)}$
- $P(y = 1 | x, \beta, b)$ für Quantile von b
- $E_b[P(y = 1 | x, \beta, b)] = P(y = 1 | x, \beta)$



Interpretation Random-Intercept-Logitmodell

Konditional: Odds-Ratio

$$\frac{\frac{P(y_{it}=1|x_{it}+1, b_i)}{P(y_{it}=0|x_{it}+1, b_i)}}{\frac{P(y_{it}=1|x_{it}, b_i)}{P(y_{it}=0|x_{it}, b_i)}} = \frac{\exp((x_{it} + 1)\beta + b_i)}{\exp(x_{it}\beta + b_i)} = \exp(\beta)$$

Marginal: Odds-Ratio

$$\frac{\frac{P(y_{it}=1|x_{it}+1)}{P(y_{it}=0|x_{it}+1)}}{\frac{P(y_{it}=1|x_{it})}{P(y_{it}=0|x_{it})}} = \frac{\frac{E_{b_i}[P(y_{it}=1|x_{it}+1, b_i)]}{E_{b_i}[P(y_{it}=0|x_{it}+1, b_i)]}}{\frac{E_{b_i}[P(y_{it}=1|x_{it}, b_i)]}{E_{b_i}[P(y_{it}=0|x_{it}, b_i)]}} \neq \exp(\beta)$$

$$\text{mit } E_{b_i}[P(y_{it} = 1|x_{it}, b_i)] = \int \frac{\exp(x_{it}\beta + b_i)}{1 + \exp(x_{it}\beta + b_i)} p(b_i) db_i \quad \text{etc.}$$

⇒ Interpretation von β nur **bedingt auf zufällige Effekte b_i** zulässig

⇒ β ist **individuen-spezifischer** Parameter, **nicht Populationsparameter**

GLMM in allgemeiner Form

- **Verteilungsannahme I:** Gegeben die zufälligen Effekte \mathbf{b} sind die y_i bedingt unabhängig und die bedingte Dichte $f(y_i|\mathbf{b})$ gehört zur Exponentialfamilie
- **Strukturannahme:** Der **bedingte** Erwartungswert $\mu_i = E(y_i|\mathbf{b}_i)$ ist mit dem linearen Prädiktor

$$\eta_i = \mathbf{x}'_i\boldsymbol{\beta} + \mathbf{z}'_i\mathbf{b}$$

durch $\mu_i = g^{-1}(\eta_i)$ verknüpft mit bekannter Linkfunktion $g(\cdot)$.

- **Verteilungsannahme II:** Die zufälligen Effekte \mathbf{b} sind normalverteilt,

$$\mathbf{b} \sim N(\mathbf{0}, \mathbf{G})$$

(auch andere Verteilungen möglich).

Hierarchisches GLMM Spezialfall mit $\mathbf{X} = (\mathbf{X}'_1 | \dots | \mathbf{X}'_m)'$, $\mathbf{Z} = \text{diag}(\mathbf{Z}_1, \dots, \mathbf{Z}_m)$ und $\mathbf{G} = \text{diag}(\mathbf{D}, \dots, \mathbf{D})$ analog zum hierarchischen LMM. [▶ Hierarch. GLMM](#)

GLMM: Marginale und konditionale Verteilung

Erinnerung:

$$E(y_i|\eta_i) = g^{-1}(\mathbf{x}'_i\boldsymbol{\beta} + \mathbf{z}'_i\mathbf{b})$$

Für $g^{-1}(x) \neq x$ gilt:

$$\begin{aligned} E_b(E(y_i|\eta_i)) &= \int E(y_i|\eta_i)p(\mathbf{b}|\boldsymbol{\vartheta})d\mathbf{b} \\ &= \int g^{-1}(\mathbf{x}'_i\boldsymbol{\beta} + \mathbf{z}'_i\mathbf{b})p(\mathbf{b}|\boldsymbol{\vartheta})d\mathbf{b} \\ &\neq g^{-1}(\mathbf{x}'_i\boldsymbol{\beta}) \end{aligned}$$

GLMM mit $g^{-1}(x) \neq x$

\Rightarrow marginaler $E_b(E(y_i|\eta_i)) \neq g^{-1}(\mathbf{x}'_i\boldsymbol{\beta})$

Parameter $\boldsymbol{\beta}$ haben nur **konditionale** Interpretation.

GLMM: Marginale und konditionale Verteilung

- **Konditionale Kovarianzstruktur:** Gegeben die zufälligen Effekte \mathbf{b} sind die y_i bedingt unabhängig

$$\text{Cov}(y_i, y_{i'} | \mathbf{b}) = 0, \quad i \neq i'.$$

(Es gibt keine Residuenkovarianzmatrix \mathbf{R} wie im LMM.)

- **Marginale Kovarianzstruktur:** Es folgt für die marginale Kovarianz

$$\begin{aligned}\text{Cov}(y_i, y_{i'}) &= \text{Cov}(E(y_i | \mathbf{b}), E(y_{i'} | \mathbf{b})) + E(\text{Cov}(y_i, y_{i'} | \mathbf{b})) \\ &= \text{Cov}(\mu_i, \mu_{i'}) + 0 \\ &= \text{Cov}(g^{-1}(\mathbf{x}'_i \boldsymbol{\beta} + \mathbf{z}'_i \mathbf{b}), g^{-1}(\mathbf{x}'_{i'} \boldsymbol{\beta} + \mathbf{z}'_{i'} \mathbf{b})), i \neq i' .\end{aligned}$$

Im Allgemeinen hängt die marginale Korrelation von den Kovariablen \mathbf{x}_i ab.

GLMM + AMM = GAMM

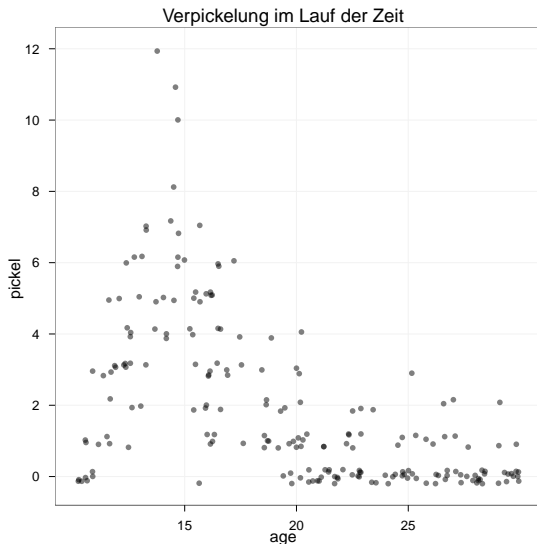
- Additive Gemischte Modelle (AMMs) und
- Generalisierte lineare gemischte Modelle (GLMMs)

lassen sich zu generalisierten additiven gemischten Modellen (GAMMs) verbinden.

Hierzu werden die Designmatrizen \mathbf{X} und \mathbf{Z} und die Kovarianz \mathbf{G} der zufälligen Effekte \mathbf{b} im GLMM analog zu den additiven gemischten Modellen gewählt, siehe Kapitel 5.

Beispiel Verpickelung

- keine echten Daten
- Pickel auf den Gesichtern von 200 Teens und Twens wurden gezählt (1 Beob./Person)
- Verlauf der Verpickelung über die Zeit



Beispiel Verpickelung: Modell

- Response pickel_i : Person i hat pickel_i Pickel im Gesicht
- Kovariable alter_i : Alter in Jahren (10 – 30)
- Modell für $\mathbf{pickel} = (\text{pickel}_1, \dots, \text{pickel}_n)$ mit Splines aus der TP(2)-Basis für $\mathbf{alter} = (\text{alter}_1, \dots, \text{alter}_n)$:

$$\mathbf{pickel} | \mathbf{b} \sim \text{Poisson}(\exp(f(\mathbf{alter})))$$

$$f(\mathbf{alter}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b}$$

$$\mathbf{X} = (\mathbf{1}, \mathbf{alter}, \mathbf{alter}^2)$$

$$\mathbf{Z} = ((\text{alter}_i - \kappa_k)_+^2)_{i=1, \dots, n; k=1, \dots, m}$$

$$\mathbf{b} \sim N(\mathbf{0}, \tau^2 \mathbf{I}_m)$$

- Wir verwenden für den Fit auf der nächsten Folie quadratische B-Splines mit Differenzen 2. Ordnung. → **Übung**

Beispiel Verpickelung: Ergebnis

