Inhalt

Generalisiertes Lineares Modell (GLM): Strukturannahme

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

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

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

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:

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

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

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

Die Exponentialfamilie - Beispiele

Dichten / Wahrscheinlichkeitsfunktionen:

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

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

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

	μ	θ	ϕ	$b(\theta)$
Normal	μ	μ	σ^2	$\frac{1}{2}\theta^2$
Binomial	π	$\log(\frac{\pi}{1-\pi})$	1	$n\log(1+\exp(\theta))$
Poisson	λ	$\log(\lambda)$	1	$\exp(\theta)$

GLM

- Strukturannahme:
$$E(y_i|\boldsymbol{x}_i) = g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta})$$
- Verteilungsannahme:
$$f(y_i|\theta_i) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{\phi_i} - c(y_i, \phi_i)\right)$$

$\Rightarrow$ Verknüpfung von Erwartungswertstruktur und Varianzstruktur
  - $\mathrm{E}(y_i|\boldsymbol{x}_i) = g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta}) = b'(\theta_i)$
  - $\mathrm{Var}(y_i|\boldsymbol{x}_i) = \phi_i b''(\theta_i)$

GLM

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

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

Also:

$$\begin{tikzcd}
\boldsymbol{\eta} \arrow[dr, "g^{-1}"] & & \boldsymbol{\theta} \arrow[dl, "b'"'] \arrow[dr, "\phi b''"] & \\
& \mathrm{E}\,(\boldsymbol{y}|\boldsymbol{\theta} \text{ bzw. } \boldsymbol{\eta}) & & \mathrm{Var}(\boldsymbol{y}|\boldsymbol{\theta})
\end{tikzcd}$$

GLM mit kanonischem Link

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

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

Also:

$$
\begin{tikzcd}
\boldsymbol{\eta} \arrow[dr, "g^{-1}"'] & \left(\overset{\text{kanonische Linkfkt.}}{=}\right) & \boldsymbol{\theta} \arrow[dl, "b'"'] \arrow[dr, "\phi b''"] & \\
& \mathrm{E}\,(\boldsymbol{y}|\boldsymbol{\theta} \text{ bzw. } \boldsymbol{\eta}) & & \mathrm{Var}(\boldsymbol{y}|\boldsymbol{\theta})
\end{tikzcd}
$$

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

LM $\Rightarrow$ LMM

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

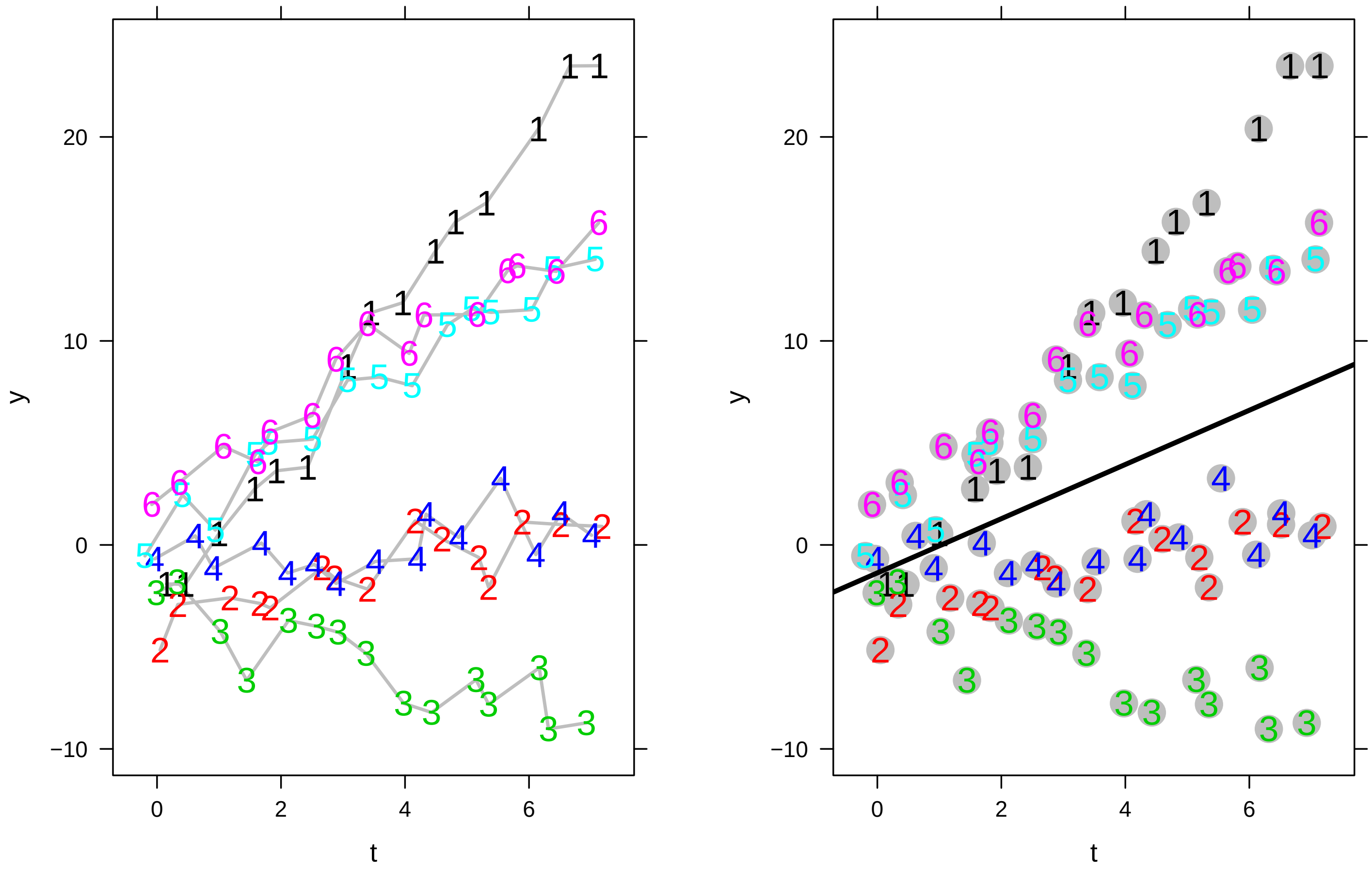
Formal: Was ist anders/neu im LMM?

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

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

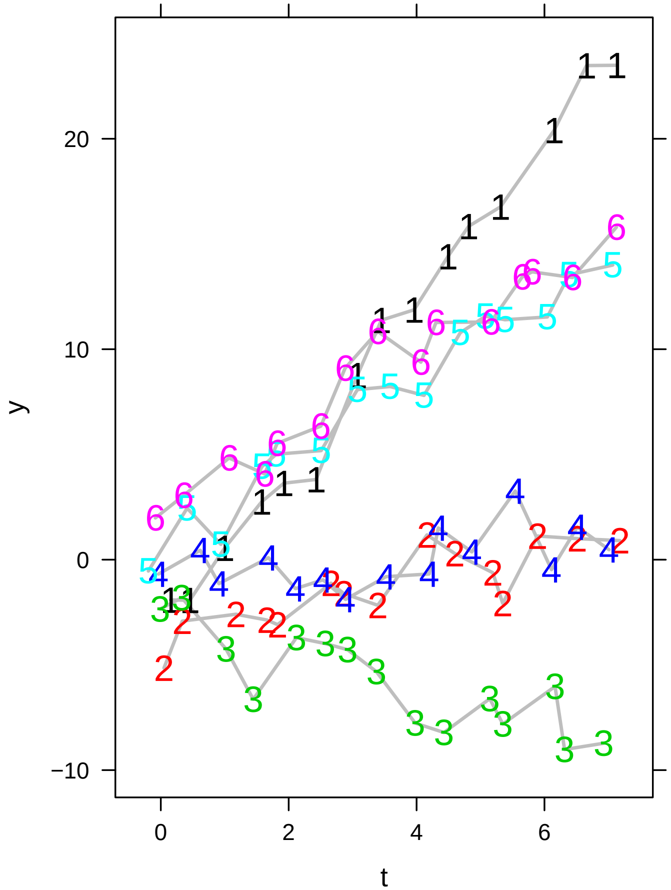
Inhaltlich: Wozu LMMs?

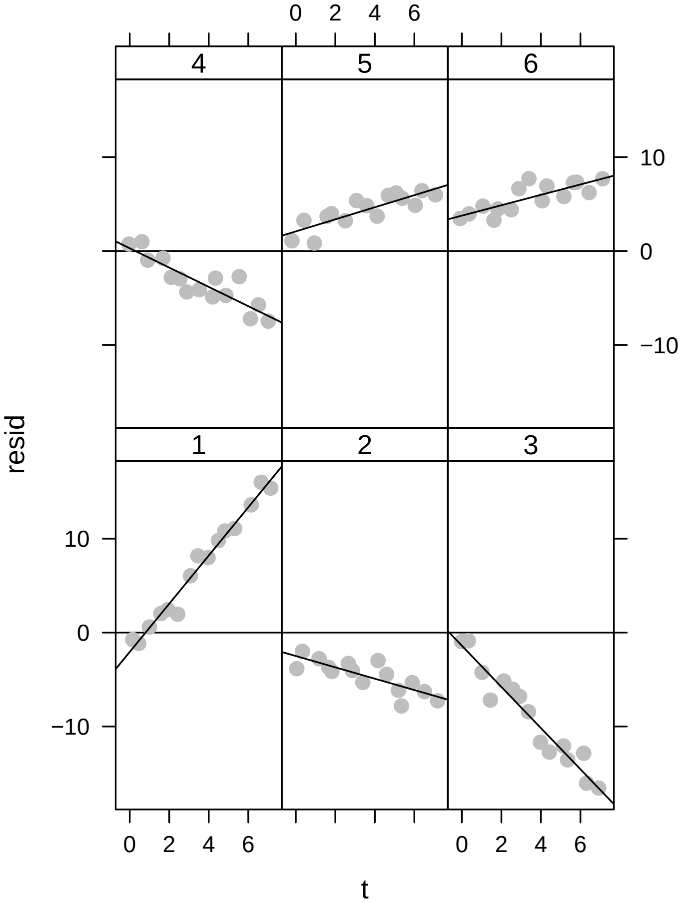
Konditional betrachtet:
(zeitliche/räumliche/hierarchische) Struktur der Daten legt bestimmte Struktur der Abweichungen der Beobachtungen $\boldsymbol{y}$ von $\boldsymbol{X}\beta$ nahe:



Inhaltlich: Wozu LMMs?

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

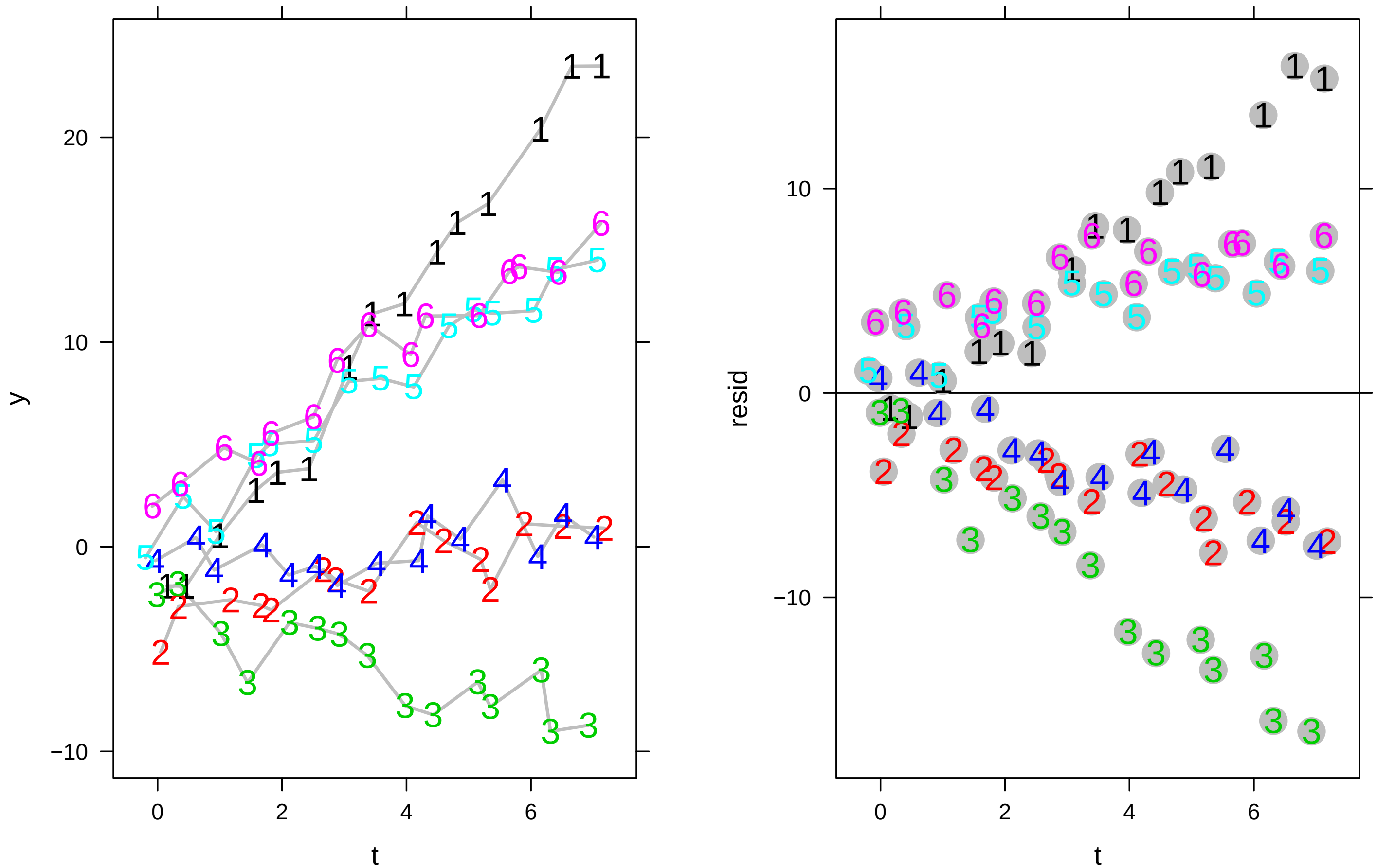




Inhaltlich: Wozu LMMs?

Marginal betrachtet:

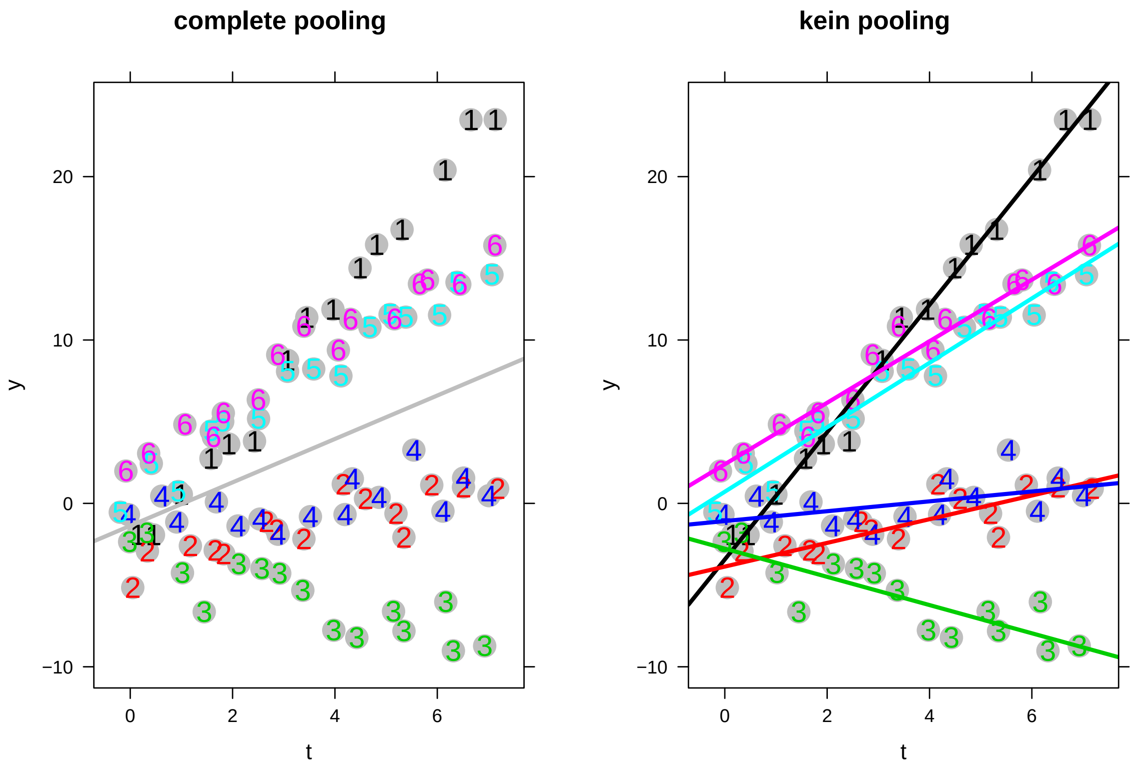
(zeitliche/räumliche/hierarchische) Struktur der Daten legt bestimmte Kovarianzstruktur der Beobachtungen nahe:



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 (oder ein fester Effekt) 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 $\boldsymbol{\vartheta}$ 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 $\boldsymbol{b}$ 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: $\boldsymbol{y}|\boldsymbol{\eta} \sim N(\boldsymbol{\eta}, \boldsymbol{R})$
- GLMM: Response aus beliebiger Exponentialfamilie:

$$\begin{aligned} \mathrm{E}(\boldsymbol{y}|\boldsymbol{\eta}) &= g^{-1}(\boldsymbol{X}\beta + \boldsymbol{Z}\boldsymbol{b}) \\ \boldsymbol{y}|\boldsymbol{\eta} &\sim \text{Expo.fam.} \end{aligned}$$

GLMM = GLM + LMM

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

GLMM für Longitudinal-/Clusterdaten

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

$$\boldsymbol{b_i} \overset{iid}{\sim} N(\boldsymbol{0}, \boldsymbol{D}).$$

Allg. GLMM

Beispiel: Madras-Daten

- 69 stationäre Schizophrenie-Patienten
- 1 Jahr lang nach Ersteinlieferung beobachtet
- 12 monatliche Beobachtungen pro Patient ob Schub (ja/nein)

Beispiel: Madras-Daten

- Response: $\mathtt{y}_{it} \in 0, 1$: 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: $\rightarrow$ **Übung**

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

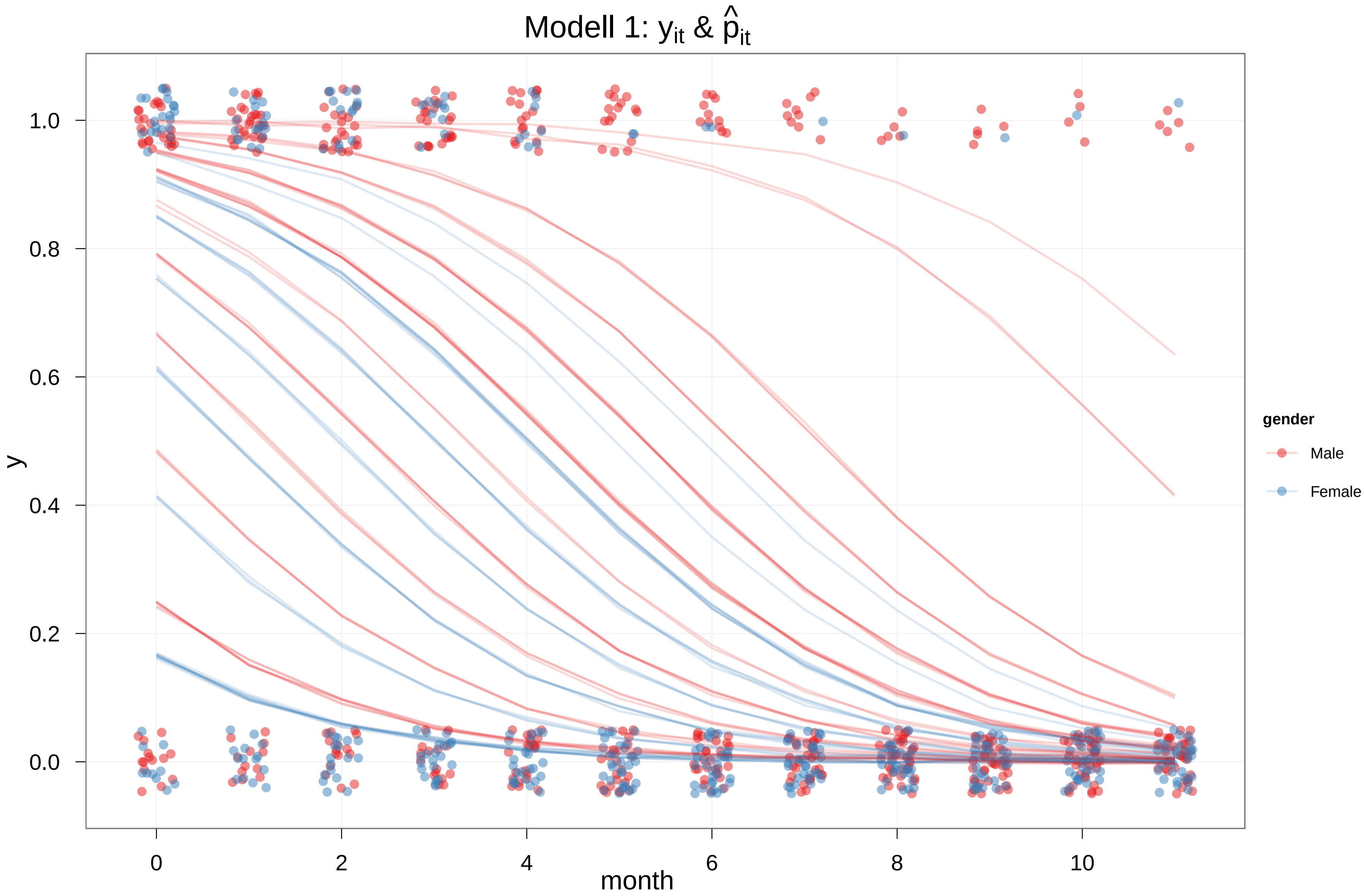
Beispiel: Madras-Daten - Ergebnisse

```
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
```

$\Rightarrow$ relativ starke individuelle Heterogenität

Beispiel: Madras-Daten - Ergebnisse



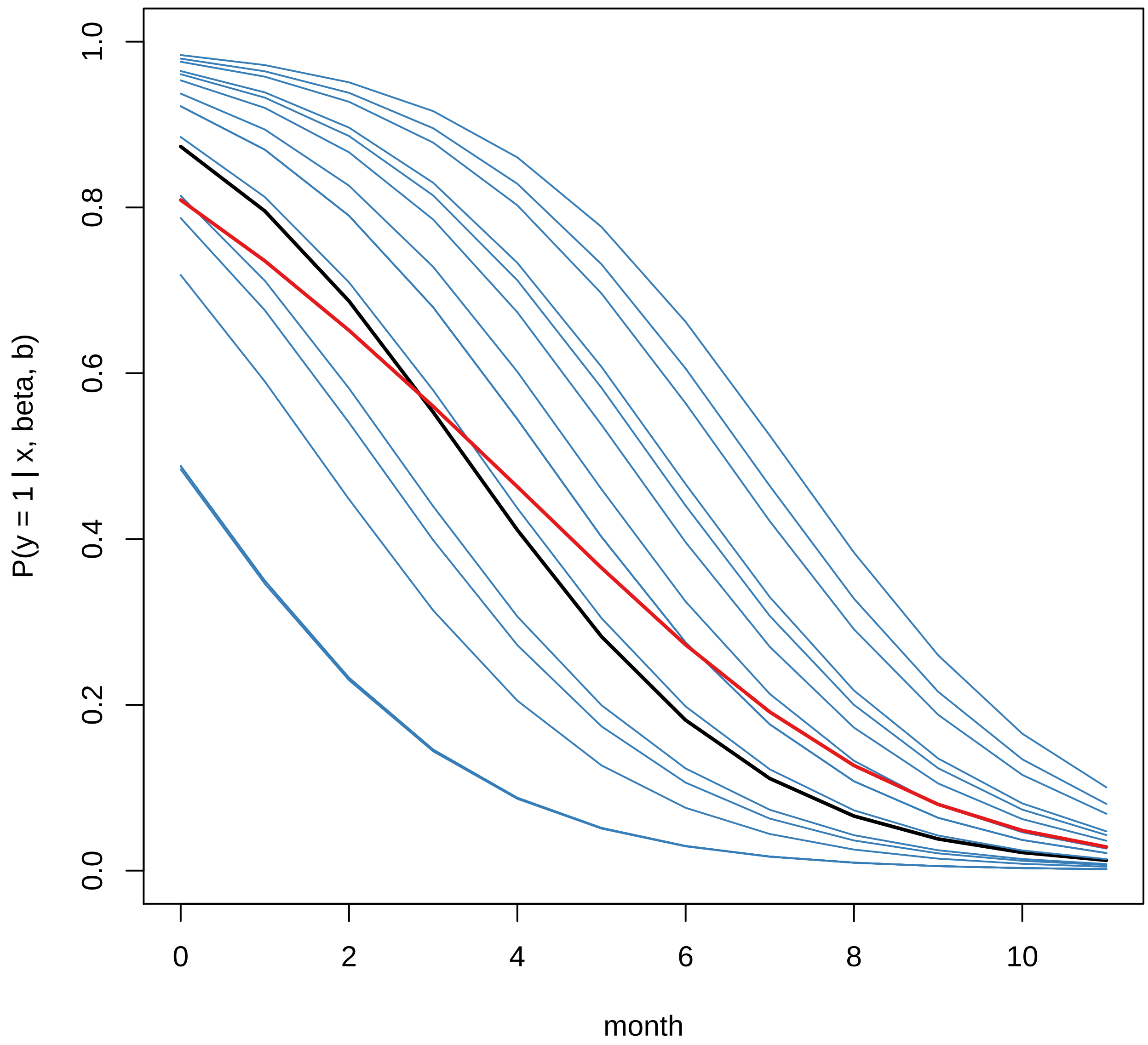
Beispiel: Madras-Daten - Interpretation

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

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

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

$\Rightarrow$ Interpretation von β nur bedingt auf zufällige Effekte b_i zulässig
$\Rightarrow$ β ist individuenspezifischer Parameter, nicht Populationsparameter

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

GLMM in allgemeiner Form

- Verteilungsannahme I: Gegeben die zufälligen Effekte $\boldsymbol{b}$ sind die y_i bedingt unabhängig und die bedingte Dichte $f(y_i|\boldsymbol{b})$ gehört zur Exponentialfamilie
- Strukturannahme: Der **bedingte** Erwartungswert $\mu_i = \mathrm{E}(y_i|\boldsymbol{b}_i)$ ist mit dem linearen Prädiktor
$$\eta_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{z}_i'\boldsymbol{b}$$
durch $\mu_i = g^{-1}(\eta_i)$ verknüpft mit bekannter Linkfunktion $g(\cdot)$.
- Verteilungsannahme II: Die zufälligen Effekte $\boldsymbol{b}$ sind normalverteilt,
$$\boldsymbol{b} \sim N(\boldsymbol{0}, \boldsymbol{G})$$
(auch andere Verteilungen möglich).

Hierarchisches GLMM Spezialfall mit $\boldsymbol{X} = (\boldsymbol{X}_1'|\ldots|\boldsymbol{X}_N')'$, $\boldsymbol{Z} = \mathrm{diag}(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_N)$ und $\boldsymbol{G} = \mathrm{diag}(\boldsymbol{D}, \ldots, \boldsymbol{D})$ analog zum hierarchischen LMM. Hierarch. GLMM

GLMM: Marginale und konditionale Verteilung

Erinnerung:

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

Für $g^{-1}(x) \neq x$ gilt:

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

GLMM mit $g^{-1}(x) \neq x$
$\Rightarrow$ marginaler Erwartungswert $E_b(E(y_i|\eta_i)) \neq g^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta})$
Parameter $\boldsymbol{\beta}$ haben nur konditionale Interpretation.

GLMM: Marginale und konditionale Verteilung

- Konditionale Kovarianzstruktur: Gegeben die zufälligen Effekte $\boldsymbol{b}$ sind die y_i bedingt unabhängig
$$\mathrm{Cov}(y_i, y_{i'} | \boldsymbol{b}) = 0, \quad i \neq i'.$$
(Es gibt keine Residuenkovarianzmatrix $\boldsymbol{R}$ wie im LMM.)

- Marginale Kovarianzstruktur: Es folgt für die marginale Kovarianz
$$\begin{aligned} \mathrm{Cov}(y_i, y_{i'}) &= \mathrm{Cov}(\mathrm{E}(y_i|\boldsymbol{b}), \mathrm{E}(y_{i'}|\boldsymbol{b})) + \mathrm{E}(\mathrm{Cov}(y_i, y_{i'}|\boldsymbol{b})) \\ &= \mathrm{Cov}(\mu_i, \mu_{i'}) + 0 \\ &= \mathrm{Cov}(g^{-1}(\boldsymbol{x_i'}\beta + \boldsymbol{z_i'b}), g^{-1}(\boldsymbol{x_{i'}'}\beta + \boldsymbol{z_{i'}'b})), i \neq i'. \end{aligned}$$
Im Allgemeinen hängt die marginale Korrelation von den Kovariablen $\boldsymbol{x_i}$ ab.

GLMM + AMM = GAMM

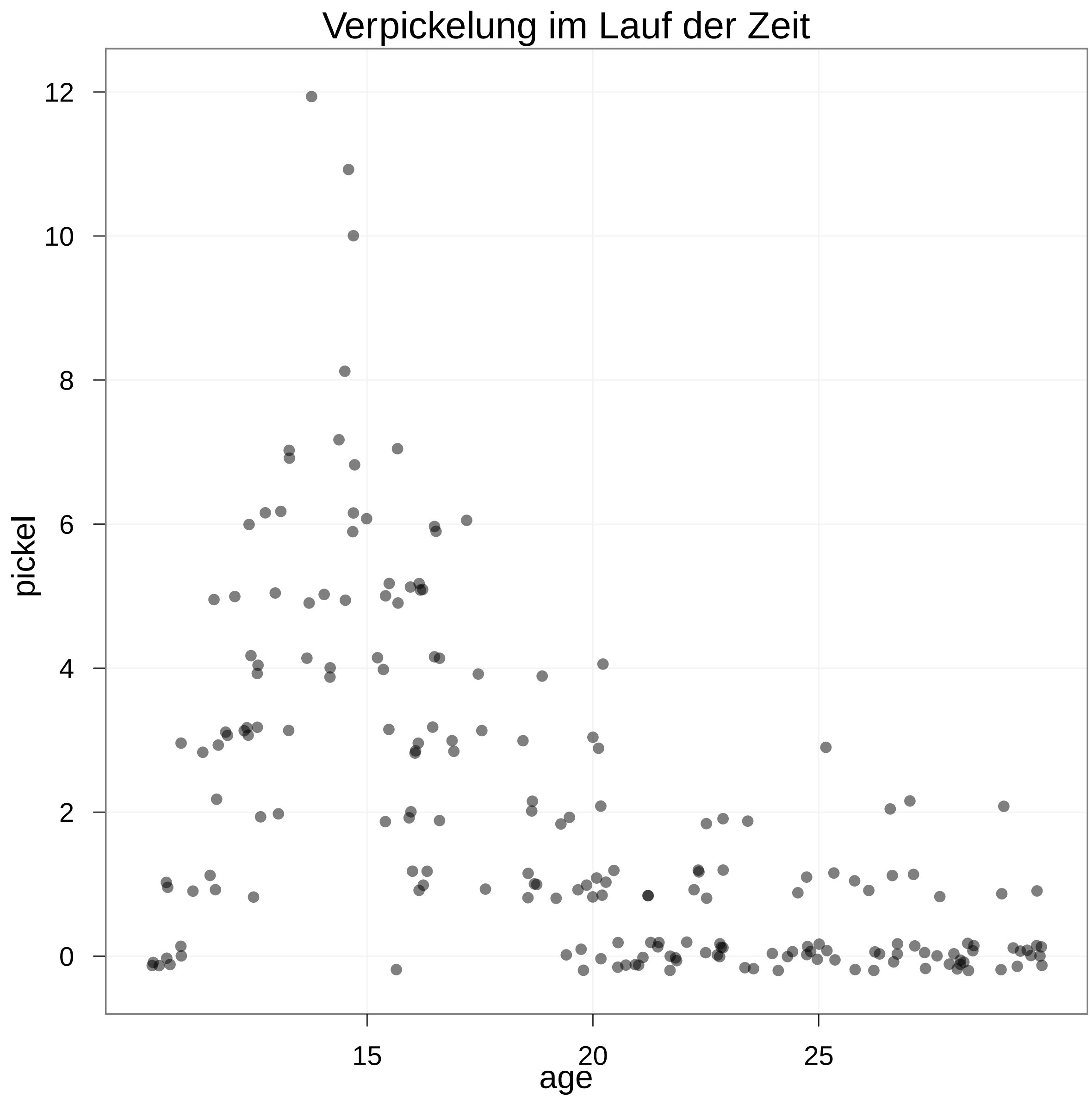
- Additive gemischte Modelle (AMMs) und
- Generalisierte lineare gemischte Modelle (GLMMs)

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

Hierzu werden die Designmatrizen $\boldsymbol{X}$ und $\boldsymbol{Z}$ und die Kovarianz $\boldsymbol{G}$ der zufälligen Effekte $\boldsymbol{b}$ 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 $\texttt{pickel}_i$: Person i hat $\texttt{pickel}_i$ Pickel im Gesicht
- Kovariable $\texttt{alter}_i$: Alter in Jahren $(10-30)$
- Modell für **pickel** $= (\texttt{pickel}_1, \ldots, \texttt{pickel}_n)$ mit Splines aus der TP(2)-Basis für **alter** $= (\texttt{alter}_1, \ldots, \texttt{alter}_n)$:

$$
\begin{aligned}
\mathbf{pickel}|\boldsymbol{b} &\sim \mathit{Poisson}\left(\exp\left(f(\mathbf{alter})\right)\right)\\
f(\mathbf{alter}) &= \boldsymbol{X}\beta + \boldsymbol{Zb}\\
\boldsymbol{X} &= (\mathbf{1}, \mathbf{alter}, \mathbf{alter}^2)\\
\boldsymbol{Z} &= \left((\texttt{alter}_i - \kappa_k)_+^2\right)_{i=1,\ldots,n;k=1,\ldots,m}\\
\boldsymbol{b} &\sim N(\mathbf{0}, \tau^2 \boldsymbol{I}_m)
\end{aligned}
$$

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

Beispiel Verpickelung: Ergebnis