

Linear Mixed Model

André Meichtry

Statistik-Kolloquium für Interessierte, BFH Gesundheit

2024-11-06

1

2

Recap: General Linear Model

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

- Die Notation $\mathbf{x}_i^T \boldsymbol{\beta}$ steht für den linearen Prädiktor
- Der lineare Prädiktor ist eine Zahl, das Skalarprodukt

$$\mathbf{x}^T \boldsymbol{\beta} = (1, x_1, \dots, x_m) \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_m \end{pmatrix} = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m$$

- Unabhängige stochastische Fehler ϵ_i i.i.d., oft $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$

Matrixform

- Mit m Eingangsgrößen und 1 Intercept, also $p = m + 1$ Parametern

$$\begin{matrix} n \times 1 & n \times p & p \times 1 & n \times 1 \\ \mathbf{Y} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\epsilon} \end{matrix}$$

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_i \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1m} \\ 1 & x_{21} & \dots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{i1} & \dots & x_{im} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nm} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_m \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_i \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

- Man nennt die Matrix \mathbf{X} die Design-Matrix
- Wir haben also n Gleichungen mit p Unbekannten
- Diese Gleichungen werden bekanntlich gelöst mit Least-Squares

3

Kleinsten Quadrate Schätzer*

- Optimierungsfunktion

$$\hat{\beta} = \arg \min_{\beta} ||\mathbf{Y} - X\beta||^2.$$

- Ableiten der Optimierungsfunktion nach $\hat{\beta}$ gibt

$$(-2)X^T(\mathbf{Y} - X\hat{\beta}) = \mathbf{0}$$

- Das führt uns zu den **Normalgleichungen**

$$\overset{p \times p}{X^T X} \overset{p \times 1}{\hat{\beta}} = \overset{p \times 1}{X^T \mathbf{Y}}$$

- Das sind p lineare Gleichungen für p Unbekannte (Komponenten von $\hat{\beta}$).

4

5

Kleinsten Quadrate Schätzer*

- Der Kleinsten-Quadrate-Schätzer ist dann

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

- Aus den Residuen $r_i = Y_i - \mathbf{x}_i^T \hat{\beta}$ bekommt man schliesslich eine Schätzung für σ^2 ,

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^n r_i^2$$

6

Korrelierte Fehler

- Personen innerhalb einem Cluster
- Messwiederholungen innerhalb einer Person
- Die Beobachtungen **innerhalb** einer Person (Cluster) sind ähnlicher als die Beobachtungen **zwischen** Personen (Cluster)
- Die Fehler sind dann **korreliert**
- Statistische Einheit ist Person/Cluster, nicht Beobachtung

7

Bedingte Unabhängigkeit

$$\Pr(Y_1 \mid Y_2, \eta) = \Pr(Y_1 \mid \eta)$$

- Beispiel: Körpergröße (Y_1) und Wortschatz (Y_2) sind nicht unabhängig. Aber sie sind bedingt unabhängig, **wenn wir** den Wert des Alters (η) kennen.
- Wenn wir η nicht kennen, dann sind Y_1 und Y_2 marginal abhängig.

Linear Mixed Model*

- Das Modell für die Einheit i (Person, Cluster) mit n_i Beobachtungen ist

$$\mathbf{Y}_i = \underbrace{\mathbf{X}_i \boldsymbol{\beta}}_{\text{fixed}} + \underbrace{\mathbf{Z}_i \mathbf{U}_i}_{\text{random}} + \underbrace{\epsilon_i}_{\text{random}}, \quad i = 1, \dots, n$$

- \mathbf{Y}_i ist ein $n_i \times 1$ -Vektor der kontinuierlichen Zielgrößen für Einheit i (Individuum oder Cluster)
- \mathbf{X}_i ist eine $n_i \times p$ Design Matrix von erklärenden Variablen (Kovariaten), fixed effects.
- $\boldsymbol{\beta}$ ist ein $p \times 1$ Vektor von Populationsparametern, fixed effects Koeffizienten
- \mathbf{Z}_i ist eine $n_i \times k$ Design Matrix von Zufallseffekten
- \mathbf{U}_i ist ein $k \times 1$ Vektor von Zufallseffekten mit Durchschnitt 0 und Kovarianzmatrix \mathbf{D}
- ϵ_i ist ein $n_i \times 1$ Fehlerterm mit unabhängigen Komponenten, jede mit Durchschnitt 0 und within-subject Varianz σ^2

Linear Mixed Models

- Wir können die Korrelation der Beobachtungen innerhalb einer Einheit (Person, Cluster) durch das Hinzufügen von zufälligen oder latenten Variablen \mathbf{U}_i berücksichtigen.
- Diese zusätzliche(n) Zufallsvariable(n) induzieren eine Korrelation der wiederholten Beobachtungen.
- **Bedingte Unabhängigkeit:** Die latente Variable erklärt die Korrelation der wiederholten Beobachtungen. Gegeben die latente Variable (wenn wir diese “fixieren”), dann sind die Beobachtungen unabhängig.
- Der Vorteil, die \mathbf{U}_i als zufällig zu behandeln, ist:
 - Weniger Parameter. In linearen Mixed Modells (LMM) sind die zusätzlichen Parameter Varianzkomponenten, die die Varianz der \mathbf{U}_i spezifizieren nicht die \mathbf{U}_i selber.
 - Fixed-Effects-Parameter \mathbf{U}_i hätten keine Interpretation als Populationsparameter.

Random Intercept Model

- Wenn \mathbf{Z}_i eine Kolonne aus Einsen ($k = 1$), haben wir das **Random Intercept Model**

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{U}_i + \epsilon_i, \quad i = 1, \dots, n$$

- Eine Zeile (Beobachtung j von Einheit i) schreiben wir dann

$$Y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + U_i + \epsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, n_i$$

- Wir haben einfach zusätzlich ein random intercept für jede Einheit i , U_i .
- Das ist das bei weitem häufigste LMM. (Einfachster Spezialfall: Gepaarter t -Test).

Random Intercept Model

- Das **Random Intercept Model** ist

$$Y_{ij} = \mathbf{x}_{ij}^T \beta + U_i + \epsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, n_i.$$

- mit

- $U_i \sim \mathcal{N}(0, \nu^2)$ mit ν^2 als der between-subject Varianz
- $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ mit σ^2 als der within-subject Varianz
- $\text{Cor}(Y_{ij}, Y_{ij'}) = \frac{\nu^2}{\nu^2 + \sigma^2}$ als der Intraklassenkorrelation (ICC)

12

Simulation von Daten aus Modell*

```
1 simulate <- function() {
2   n <- 50 # number of subjects
3   K <- 4 # number of measurements per subject
4   # we construct a data frame with the design: everyone has a baseline measurement, and then measurements at two
5   d.SP <- data.frame(id = rep(seq_len(n), each = K), time = rep(1:K, n), group = rep(gl(2, n/2, labels = c("A", "B")), n))
6   # design matrices for the fixed and random effects
7   X <- model.matrix(~group * time, data = d.SP)
8   Z_base <- cbind(1, d.SP$time[1:K]) # intercept and time columns for one subject
9   Z <- kronecker(diag(n), Z_base) # Block-diagonal matrix for all subjects
10  beta <- c(10, 3, 1, 4) # fixed effects coefficients
11  ##
12  var_intercept <- 1.0 # variance of random intercepts
13  var_slope <- 1 # variance of random slopes
14  correlation <- 0 # correlation between intercept and slope
15  D <- matrix(c(var_intercept, correlation * sqrt(var_intercept * var_slope),
16               correlation * sqrt(var_intercept * var_slope), var_slope), nrow=2)
17  sigma <- 2 # sd error terms
18  # Simulate random effects for intercept and slope
19  set.seed(112)
20  B <- MASS::mvrnorm(n, rep(0, ncol(Z_base)), D)
21  b <- as.vector(t(B))
22  # linear predictor
23  eta <- X %*% beta + Z %*% b
24  # we simulate continuous longitudinal data
25  d.SP$y <- rnorm(n * K, mean = eta, sd = sigma)
26  res <- list(data=d.SP, params=list(random=apply(B, 2, var), fixed=beta, sigma2=sigma^2))
27 }
```

14

Split-Plot Design

- Häufiges Design/Analyse in Gesundheitswissenschaften
- Ein between-subject Faktor **group**, eine within-subject Variable **time**
- Design: Zwei Gruppen (between) über die Zeit (within)
- Man nennt dieses Design auch Split-Plot Design oder
- Repeated Measure ANOVA mit between-subject Faktor

13

Simulation von Daten aus Modell*

True values

```
1 res<-simulate()
2 d.SP<-res$data
3 res$params
```

```
$random
[1] 0.7758 1.0514
```

```
$fixed
[1] 10 3 1 4
```

```
$sigma2
[1] 4
```

Design

- Fully crossed group(2) × time(4) between-within Design

```
1 with(d.SP, xtabs(~group+time))
```

```
time
group 1 2 3 4
A 25 25 25 25
B 25 25 25 25
```

15

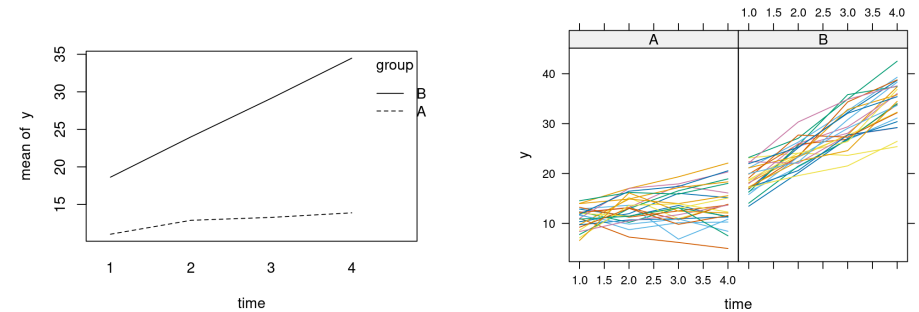
Daten

```
1 psych::headTail(d.SP,10,10)
```

```
   id time group  y
1    1     1    A 12.84
2    1     2    A 11.45
3    1     3    A 13.59
4    1     4    A 11.25
5    2     1    A  14
6    2     2    A 17.02
7    2     3    A 19.34
8    2     4    A 22.09
9    3     1    A 12.33
10   3     2    A 10.21
... .. <NA> ...
191 48     3    B 21.53
192 48     4    B 26.45
193 49     1    B 22.11
194 49     2    B 30.33
195 49     3    B 34.86
196 49     4    B 37.57
197 50     1    B 21.98
198 50     2    B 25.25
199 50     3    B 32.07
200 50     4    B 35.4
```

Beschreiben

```
1 with(d.SP,interaction.plot(x.factor=time,trace.factor=group,response=y))
2 lattice::xyplot(data = d.SP, y ~ time | group, groups = id, type = "l")
```



16

17

Beschreiben

```
1 psych::describeBy(d.SP$y, list(d.SP$group,d.SP$time), mat = TRUE)
```

	item	group1	group2	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
X11	1	A	1	1	25	11.03	2.178	11.59	11.12	1.923	6.577	14.54	7.964	-0.39497	-0.8672	0.4355
X12	2	B	1	1	25	18.63	2.718	18.14	18.66	2.665	13.444	23.23	9.782	0.09828	-0.9920	0.5436
X13	3	A	2	1	25	12.88	2.646	13.00	12.95	2.784	7.254	17.02	9.763	-0.10523	-0.9265	0.5293
X14	4	B	2	1	25	23.99	2.538	23.60	23.91	2.789	19.574	30.33	10.752	0.33035	-0.2770	0.5075
X15	5	A	3	1	25	13.27	3.391	13.11	13.40	3.554	6.180	19.34	13.158	-0.14966	-0.7198	0.6782
X16	6	B	3	1	25	29.14	3.733	27.74	29.17	2.395	21.531	35.81	14.279	0.13509	-0.8811	0.7466
X17	7	A	4	1	25	13.88	4.230	13.68	13.90	3.585	4.961	22.09	17.124	0.07687	-0.6399	0.8459
X18	8	B	4	1	25	34.48	4.183	35.40	34.68	4.315	25.392	42.51	17.118	-0.41266	-0.5375	0.8366

```
1 ##
2 aggregate(y~time+group,data=d.SP,summary)
```

	time	group	y.Min.	y.1st Qu.	y.Median	y.Mean	y.3rd Qu.	y.Max.
1	1	A	6.577	9.725	11.594	11.031	12.326	14.541
2	2	A	7.254	11.250	13.003	12.881	14.881	17.017
3	3	A	6.180	10.931	13.105	13.271	16.035	19.338
4	4	A	4.961	11.291	13.675	13.877	16.093	22.086
5	1	B	13.444	16.887	18.144	18.628	21.123	23.226
6	2	B	19.574	22.285	23.596	23.994	25.668	30.326
7	3	B	21.531	26.861	27.744	29.140	32.216	35.810
8	4	B	25.392	32.189	35.398	34.480	37.455	42.509

LMM anpassen

- Anpassen mit Funktionen `lmer()` aus Packet `lme4`, immer auch Packet `lmerTest` laden für Tests.
- Das neue Element in der Syntax ist ein Term `(1|subject)` oder `(time|subject)` in der Modellformel.
- Dieser Term steht dann für ein Random Intercept- oder ein Random Slope Modell.

18

19

Modelle anpassen

- Fixed effects: `group`, `time` und `group:time` Interaktion
- Verschiedene Random-Effects- Strukturen möglich:
- Random intercept model

```
1 library(lme4)
2 library(lmerTest)
3 m1 <- lmer(y ~ group * time + (1 | id), data = d.SP)
```

- Random slope model (hier ohne Intercept-Slope Korrelation)
 - `time|id`: korrelierte random effects
 - `time||id`: nicht korrelierte random effects

```
1 m2 <- lmer(y ~ group * time + (time || id), data = d.SP)
```

Random intercept

```
1 summary(m1, cor = FALSE)
```

Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: y ~ group * time + (1 | id)
Data: d.SP

REML criterion at convergence: 988.7

Scaled residuals:

	Min	1Q	Median	3Q	Max
	-2.9853	-0.4988	-0.0115	0.6250	2.6959

Random effects:

Groups	Name	Variance	Std.Dev.
id	(Intercept)	5.06	2.25
Residual		5.69	2.39

Number of obs: 200, groups: id, 50

Fixed effects:

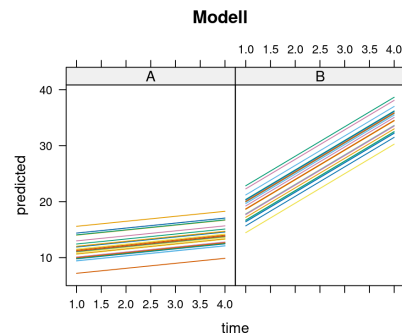
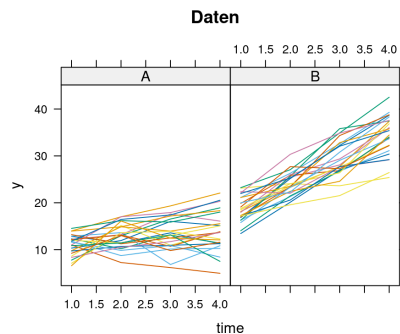
	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	10.533	0.737	151.817	14.28	< 2e-16
groupB	2.852	1.043	151.817	2.73	0.007
time	0.893	0.213	148.000	4.19	4.9e-05
groupB:time	4.378	0.302	148.000	14.51	< 2e-16

20

21

Random intercept

```
1 predicted <- predict(m1)
2 lattice::xyplot(data = d.SP, y ~ time | group, groups = id, type = "l", main="Daten")
3 lattice::xyplot(data = d.SP, predicted ~ time | group, groups = id, type = "l", main="Modell")
```



Random slope for time

```
1 summary(m2, cor = FALSE)
```

Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: y ~ group * time + (time || id)
Data: d.SP

REML criterion at convergence: 944.5

Scaled residuals:

	Min	1Q	Median	3Q	Max
	-2.2822	-0.5329	-0.0014	0.5793	2.7807

Random effects:

Groups	Name	Variance	Std.Dev.
id	(Intercept)	0.775	0.880
id.1	time	0.874	0.935
Residual		3.773	1.942

Number of obs: 200, groups: id, 50

Fixed effects:

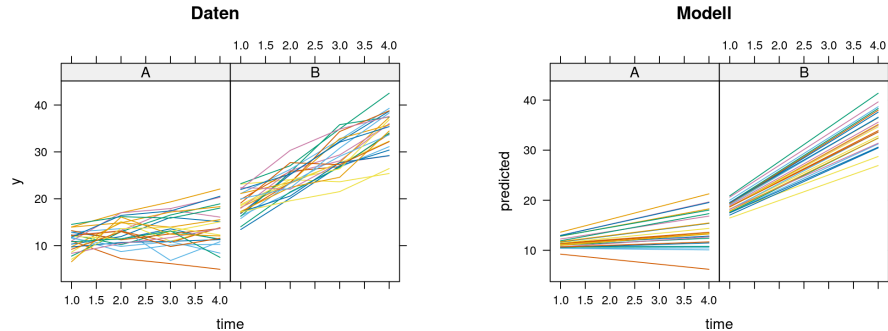
	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	10.533	0.507	93.599	20.76	< 2e-16
groupB	2.852	0.717	93.599	3.98	0.00014
time	0.893	0.255	93.606	3.50	0.00072

22

23

Random slope for time

```
1 predicted <- predict(m2)
2 lattice::xyplot(data = d.SP, y ~ time | group, groups = id, type = "l", main="Daten")
3 lattice::xyplot(data = d.SP, predicted ~ time | group, groups = id, type = "l", main="Modell")
```



Testing random effects

```
1 anova(m1, m2)
```

Data: d.SP
Models:
m1: y ~ group * time + (1 | id)
m2: y ~ group * time + (time || id)
npair AIC BIC logLik deviance Chisq Df Pr(>Chisq)
m1 6 999 1019 -494 987
m2 7 957 980 -471 943 44.3 1 2.8e-11

```
1 lmerTest::ranova(m2)
```

ANOVA-like table for random-effects: Single term deletions

Model:
y ~ group + time + (1 | id) + (0 + time | id) + group:time
npair logLik AIC LRT Df Pr(>Chisq)
<none> 7 -472 959
(1 | id) 6 -473 957 0.7 1 0.39
time in (0 + time | id) 6 -494 1001 44.1 1 3.1e-11

24

25

Testings fixed effects

```
1 drop1(m2)
```

Single term deletions using Satterthwaite's method:

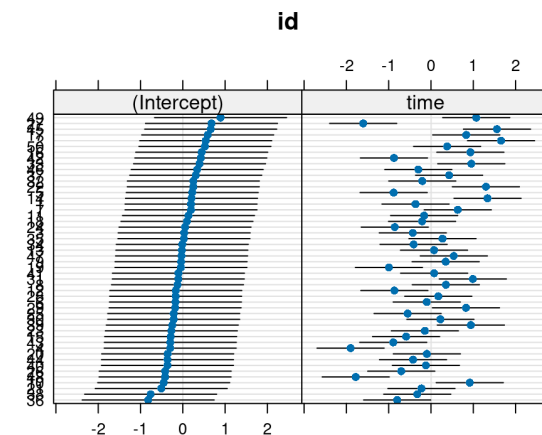
Model:
y ~ group * time + ((1 | id) + (0 + time | id))
Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
group:time 555 555 1 93.6 147 <2e-16

- `drop1()` testet nur die Interaktion
- Haupteffekte sollten in der Gegenwart von Interaktion nicht getestet werden sollten

Plotting random effects

```
1 dotplot.ranef.mer(ranef(m2))
```

\$id

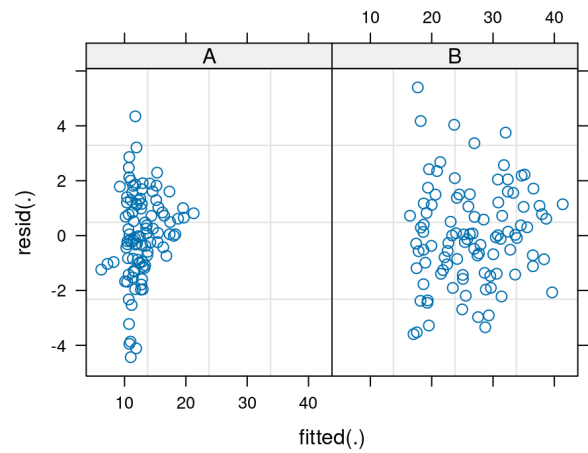


26

27

Modellannahmen

```
1 plot(m2,resid(.)~fitted(.)|group)
```



Modellannahmen

