Lineare Gemischte Modelle (2)

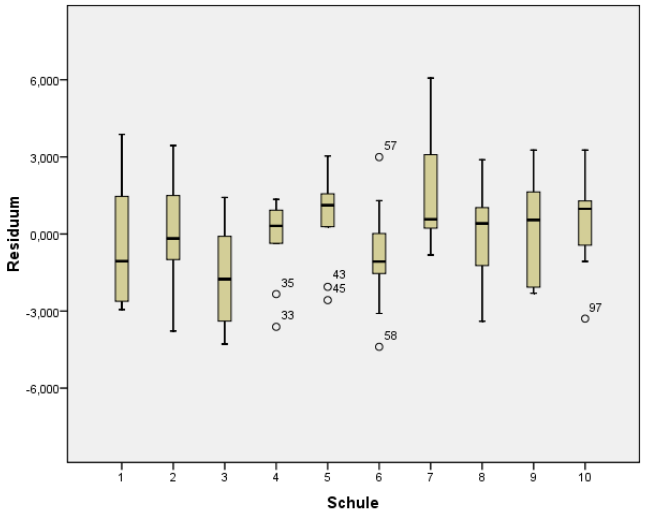
Wiederholung ...

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki} + ... + \epsilon_i$$

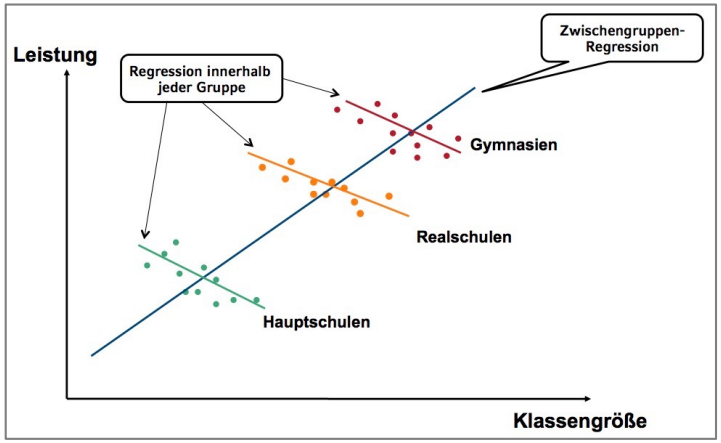
identically and independently distributed

$$\epsilon \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_\epsilon)$$

Problem bei gruppierten Daten



Nicht-unabhängige Residuen



Ökologischer Fehlschluss: Aggregationsebene berücksichtigen

Wiederholung:

- Wie könnte man nun mit herkömmlichen regressionsanalytischen Methoden gruppierte Daten analysieren?

  - ➢ Aggregation der Daten (d.h. nur auf Gruppenebene rechnen)
  - ➢ Disaggregation der Daten („Complete pooling“)
  - ➢ Gruppen separat analysieren („no pooling“)

Hauptproblem: Ökologischer Fehlschluss

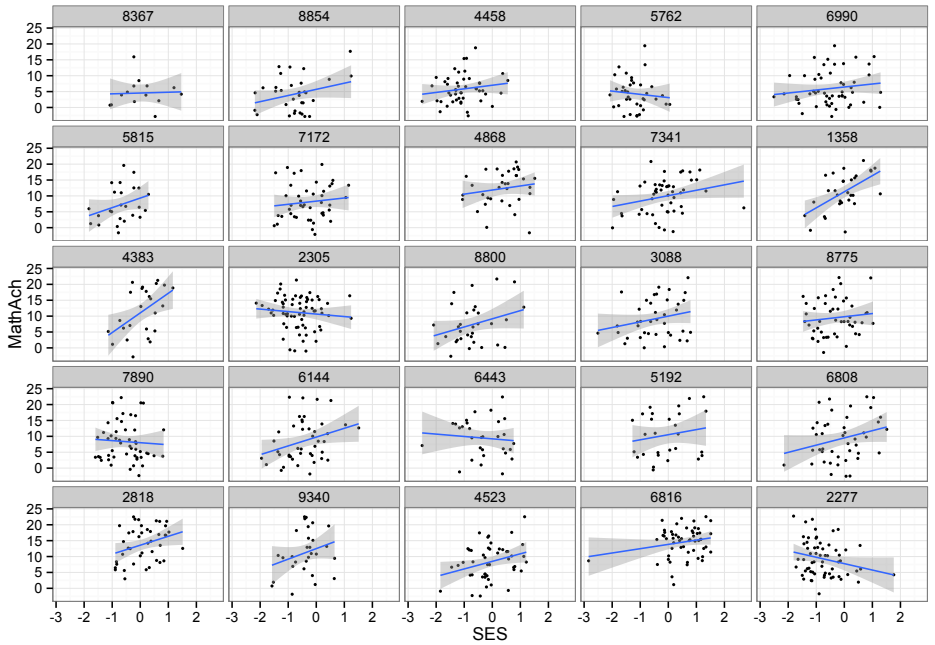
Hauptproblem: Erhöhte Typ-I-Fehlerrate bei L2-Variablen; ignoriert Variabilität der Gruppen

Hauptproblem: Overfitting in einzelnen Gruppen; wie fasst man die vielen Modellparameter optimal zusammen?

- Alle drei „work-arounds“ führen zu falschen Ergebnissen!

Hierarchische Lineare Modelle: Hinleitung

- Implikationen des herkömmlichen Regressionsansatzes:
  - β_0 und β_1 werden als **feste** Parameter der zu schätzenden *Populations*-Regressionsgeraden angenommen.
  - ➢ Es gibt jeweils nur **einen** zu schätzenden Parameter für β_0 und β_1.
  - ➢ Sobald eine gruppierte Datenstruktur vorliegt und β_0 und β_1 von Gruppe zu Gruppe variieren, kann dieses Modell nicht mehr gelten.
  - ➢ Flexibleres Regressionsmodell nötig!



Das Random Coefficient Model (RCM)

Struktur der Regressionsgleichung auf Level 1

- Standard-Regressionsgleichung (OLS-Regression):

$$y_i = \beta_0 + \beta_1 x_i + r_i$$

- Wenn β_0 und β_1 in jedem Cluster/Gruppe unterschiedlich sein darf:

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + r_{ij}$$

β_{0j} = Regressionskonstante (intercept) in Cluster j

β_{1j} = Steigungsparameter (slope) in Cluster j

r_{ij} = Fehlerterm (residual error) der Person i in Cluster j

i = Index für Individuum

j = Index für Level-2-Einheit (Gruppenindikator)

Struktur der Regressionsgleichung auf Level 1

- Jede Gruppe j (Level-2-Einheit) hat einen eigenen Intercept β_{0j}
- Jede Gruppe j hat einen eigenen Slope β_{1j}

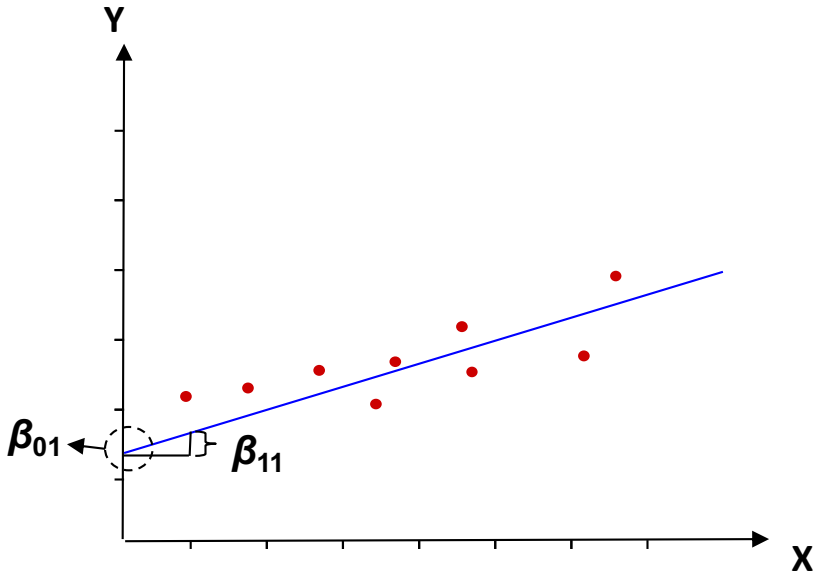
→ β_{0j} und β_{1j} können also über die Level-2-Einheiten hinweg variieren

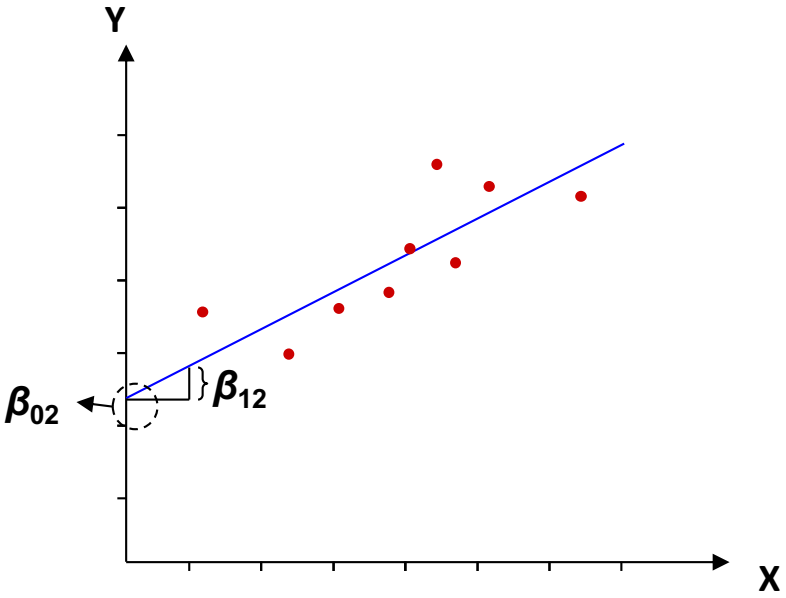
→ β_{0j} und β_{1j} weisen also jeweils Varianz auf

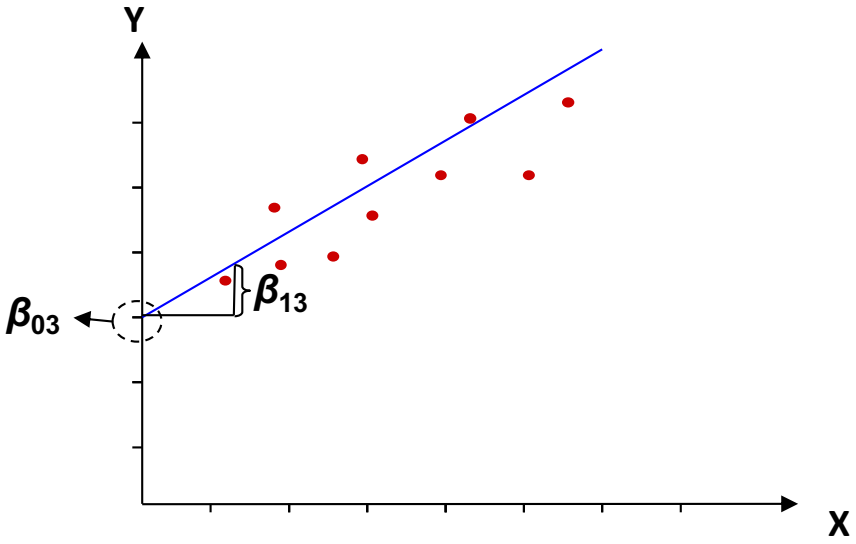
- Daher auch der Name „Random coefficient model“: Achsenabschnitt (β_0) und Steigungsparameter (β_1 etc.) können variieren → die Koeffizienten (d.h., die Modellparameter) sind „random“

- **Hinweis:** Es können auch Modelle geschätzt werden, die auf Gruppenebene nur bezüglich des Intercepts *oder* des Slopes variieren (siehe später)

Variation der Regressionskoeffizienten in den Gruppen – Beispiel anhand drei Gruppen







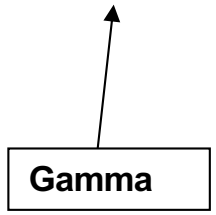
β_{0j} und β_{1j} variieren über die Gruppen hinweg

→ Regressionskoeffizienten weisen jeweils Varianz auf.

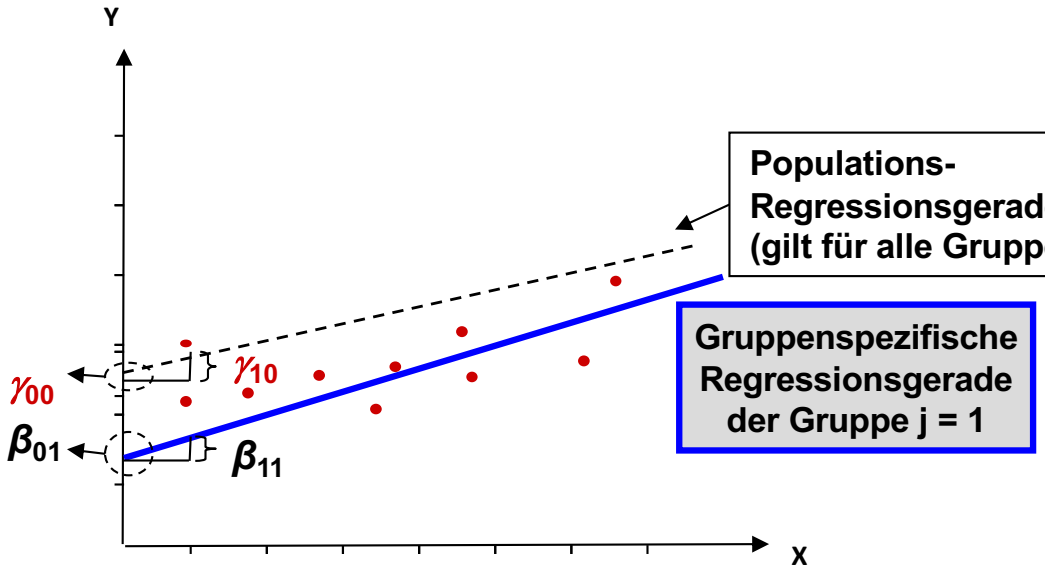
Einführung von γ (gamma)

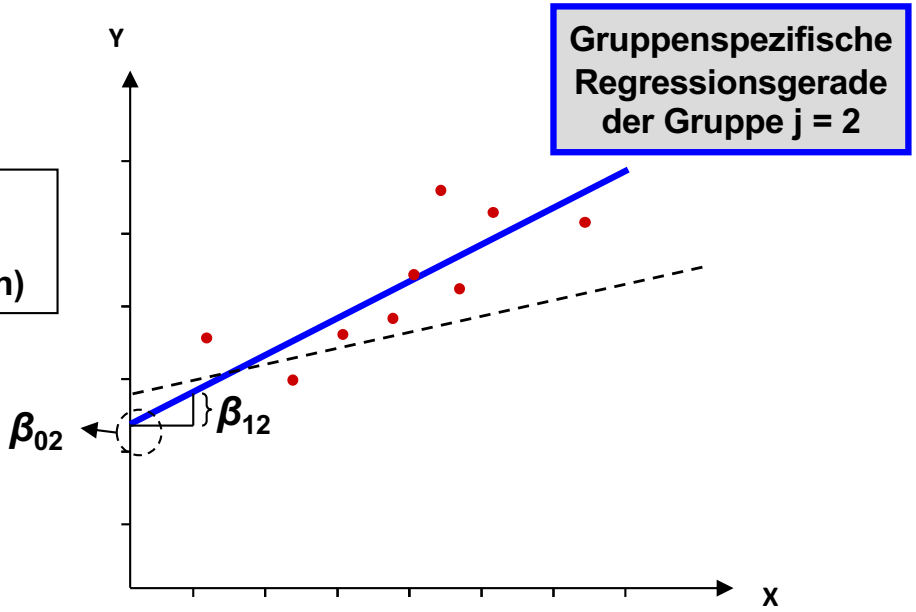
- Allerdings: Das Konzept **einer** Populations-Regressionsgleichung mit einem Populations-Regressionsintercept und einem Populations-Regressions-Slope wird beibehalten

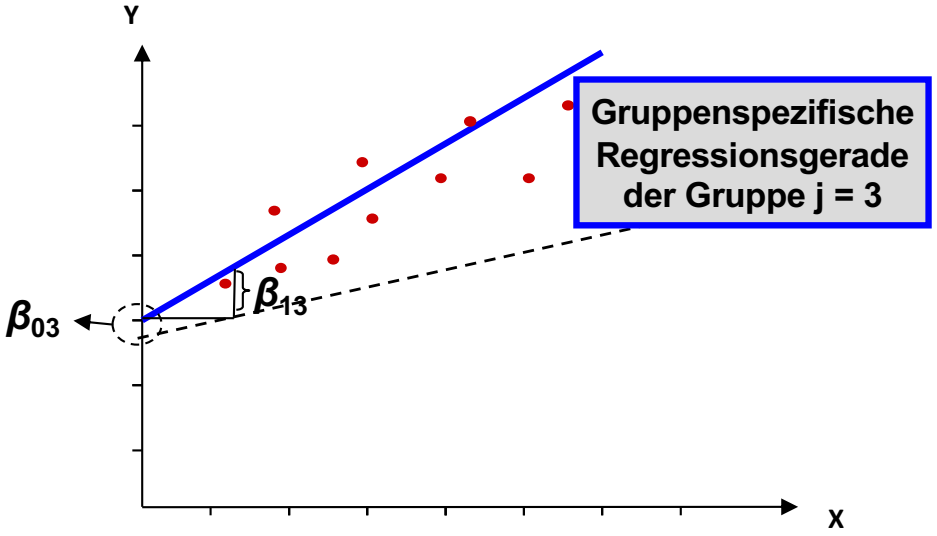
➢ Die Level-2-Gleichungen drücken daher aus, wie die Regressionskoeffizienten β_{0j} und β_{1j} **in Beziehung** zu den Populations-Regressionskoeffizienten γ_{00} (Intercept) und γ_{10} (Slope) stehen.



Variation der Regressionskoeffizienten in den Gruppen







β_{0j} und β_{1j} weichen jeweils um u_{0j} bzw. u_{1j} von den Populationsparametern γ_{00} und γ_{10} ab.

Gleichungen auf Level 2

- Die Regressionsparameter β_{0j} und β_{1j} werden also jetzt als Funktion von Level-2-Parametern modelliert:

- Level-1-Gleichung:

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + r_{ij}$$

- Level-2-Gleichungen:

u für Fehler auf Gruppenebene

$$\beta_{0j} = \gamma_{00} + u_{0j}, \quad u_{0j} \sim \mathcal{N}(0, \tau_{00})$$
$$\beta_{1j} = \gamma_{10} + u_{1j}, \quad u_{1j} \sim \mathcal{N}(0, \tau_{11})$$

Kein Index j mehr: gleich für alle L2-Einheiten

Gleichungen auf Level 2

- Level-2-Gleichungen:

$$\beta_{0j} = \gamma_{00} + u_{0j}, \quad u_{0j} \sim \mathcal{N}(0, \tau_{00})$$
$$\beta_{1j} = \gamma_{10} + u_{1j}, \quad u_{1j} \sim \mathcal{N}(0, \tau_{11})$$

- Unsystematische Variation von β_{0j} und β_{1j} wird auf Level 2 durch *gruppenspezifische* Zufallskomponenten u_{0j} und u_{1j} mit dem Erwartungswert null modelliert:
  - u_{0j}: gruppenspezifische Zufallskomponente des Intercepts β_{0j}
  - u_{1j}: gruppenspezifische Zufallskomponente des Slopes β_{1j}

- *Hinweis:* Später werden wir versuchen, die gruppenspezifischen Abweichungen vom Populationsintercept und –slope vorherzusagen. Ohne L2-Prädiktoren ist jedoch jede Abweichung zunächst mal unsystematisch.

Gleichung auf Level 2: Bedeutung der Level-2-Zufallskomponenten u_{kj}

- Level-2-Gleichungen, Interpretation der **Intercepts**:
- $\beta_{0j} = \gamma_{00} + u_{0j}$
  - β_{0j} ist also Funktion eines **festen (fixed)** Populationseffektes γ_{00} plus einer zufälligen (random) Abweichung $u_{0j} = \beta_{0j} - \gamma_{00}$ vom Populations-**Intercept**.
  - $u_{0j} \sim \mathcal{N}(0, \tau_{00})$

Vorsicht: **τ_{00} ist eine Varianz und keine Standardabweichung!**

  - Die Varianz τ_{00} ist also ein Maß dafür, wie stark die Gruppenintercepts β_{0j} vom Gesamtintercept γ_{00} abweichen.
  - γ_{00} ist dementsprechend der Erwartungswert der β_{0j}
  - Ist $\tau_{00} = 0$, dann sind alle $\beta_{0j} = \gamma_{00}$ (so wie im herkömmlichen nicht-hierarchischen Regressionsmodell)

Gleichung auf Level 2: Bedeutung der Level-2-Zufallskomponenten u_{kj}

- Level-2-Gleichungen, Interpretation der **Slopes**:

- $\beta_{1j} = \gamma_{10} + u_{1j}$

  - β_{1j} ist also Funktion eines **festen (fixed)** Populationseffektes γ_{10} plus einer zufälligen (random) Abweichung $u_{1j} = \beta_{1j} - \gamma_{10}$ vom Populations-**Slope**.

  - $u_{1j} \sim \mathcal{N}(0, \tau_{11})$

Vorsicht: τ_{11} ist eine Varianz und keine Standardabweichung!

  - Die Varianz τ_{11} ist also ein Maß dafür, wie stark die Gruppenslopes β_{1j} vom Gesamtslope γ_{10} abweichen.
  - γ_{10} ist dementsprechend der Erwartungswert der β_{1j}
  - Ist $\tau_{11} = 0$, dann sind alle $\beta_{1j} = \gamma_{10}$ (so wie im herkömmlichen nicht-hierarchischen Regressionsmodell)

Begriffsklärung: feste und zufällige Effekte (fixed vs. random)

- Feste Effekte / *fixed effects*:

  Slopes/Intercepts über alle Gruppen hinweg betrachtet → „Populationsintercept“/“-slope“.

  - ➢ Es sind durchschnittliche (gruppenunspezifischen) Effekte
  - ➢ In der „klassischen“ (multiple) Regression sind alles feste Effekte

- Zufällige Effekte / *random effects*:

  Regressions-Achsenabschnitte (Intercepts) und -Steigungen (Slopes) können irgendwie (= zufällig) zwischen Gruppen hinweg variieren.

  - ➢ Eventuell kann man ihre Varianz mit anderen Variablen erklären

- Gemischte Effekte / *Mixed effect models:*
  Regressionsmodelle, die sowohl feste als auch zufällige Effekte enthalten.

Einsetzen in eine kombinierte Gleichung

- Level-1-Gleichung: $y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + r_{ij}$

- Level-2-Gleichungen:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$

- Einsetzen von Level 2 in Level 1:

$$y_{ij} = \gamma_{00} + u_{0j} + (\gamma_{10} + u_{1j}) x_{ij} + r_{ij}$$

- Umstellen und Ausmultiplizieren:

$$y_{ij} = \underbrace{\gamma_{00} + \gamma_{10} x_{ij}}_{\text{fixed part}} + \underbrace{u_{0j} + u_{1j} x_{ij} + r_{ij}}_{\text{random part}}$$

Varianzkomponenten

Durch Berücksichtigung der Gruppierungsstruktur gibt es drei **Quellen** von Zufallsvariationen:

1. Level-1-Zufallsfehler r_{ij} in den y-Werten:
   Abweichungen der Probandenwerte von der Regression.
2. Level-2-Abweichungen u_{0j} der Random-Intercepts um den Populations-Intercept.
3. Level-2-Abweichungen u_{1j} der Random-Slopes um den Populations-Slope

➢ Diese Koeffizienten bezeichnen (in Kombination) die Abweichung **einer** spezifischen Person (L1) oder **einer** spezifischen Gruppe (L2).

→ Jede dieser Abweichungen wird durch ihre **Varianz** beschrieben.

→ **Es ergeben sich also mehrere voneinander trennbare Fehlervarianzkomponenten**

Varianzkomponenten: Notation

Varianzkomponenten:

σ^2: Zufallsfehlervarianz auf Level 1, d.h. Varianz der r_{ij} (Residuen)

τ_{00}: Varianz der Achsenabschnitte (Intercepts), d.h. Varianz der u_{0j}

τ_{11}: Varianz der Steigungskoeffizienten (Slopes), d.h. Varianz der u_{1j}

τ_{01}: *Kovarianz* zwischen den Steigungskoeffizienten (Slopes) und Achsenabschnitten (Intercepts), d.h. Kovarianz zwischen u_{0j} und u_{1j}

- **τ_{01}:** z.B. positive Kovarianz: Gruppen, die ein höheres Intercept haben, haben einen höheren Slope.
- Für jeden weiteren Prädiktor auf Level 1 gibt es (z.B. für den zweiten Prädiktor) ein β_{2j} mit den dazugehörigen Varianzkomponenten:
  - τ_{22} als Varianz der u_{2j},
  - τ_{21} als Kovarianz der u_{2j} und u_{1j},
  - τ_{20} als Kovarianz der u_{2j} und u_{0j}.

Das Random-Intercept-Modell

…als Spezialfall des Random Coefficient Models: von all den möglichen Koeffizienten des Modells darf nur das Intercept zufällig zwischen den Gruppen variieren (nicht die Slopes).

Random-Intercept-Modell ohne Prädiktor (= random effects ANOVA)

- Das Intercept darf zwischen den L2-Einheiten variieren
- Es gibt **keine** Prädiktorvariable

- Level-1-Gleichung: $y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + r_{ij}$

- Level-2-Gleichungen: $\beta_{0j} = \gamma_{00} + u_{0j}$

  $\beta_{1j} = \gamma_{10} + u_{1j}$

- Kombinierte Gleichung: $y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$

- Ein Datenpunkt wird vorhergesagt durch das Gesamtintercept, die gruppenspezifische Abweichung vom Gesamtintercept, und die Abweichung relativ zum Gruppenmittelwert.

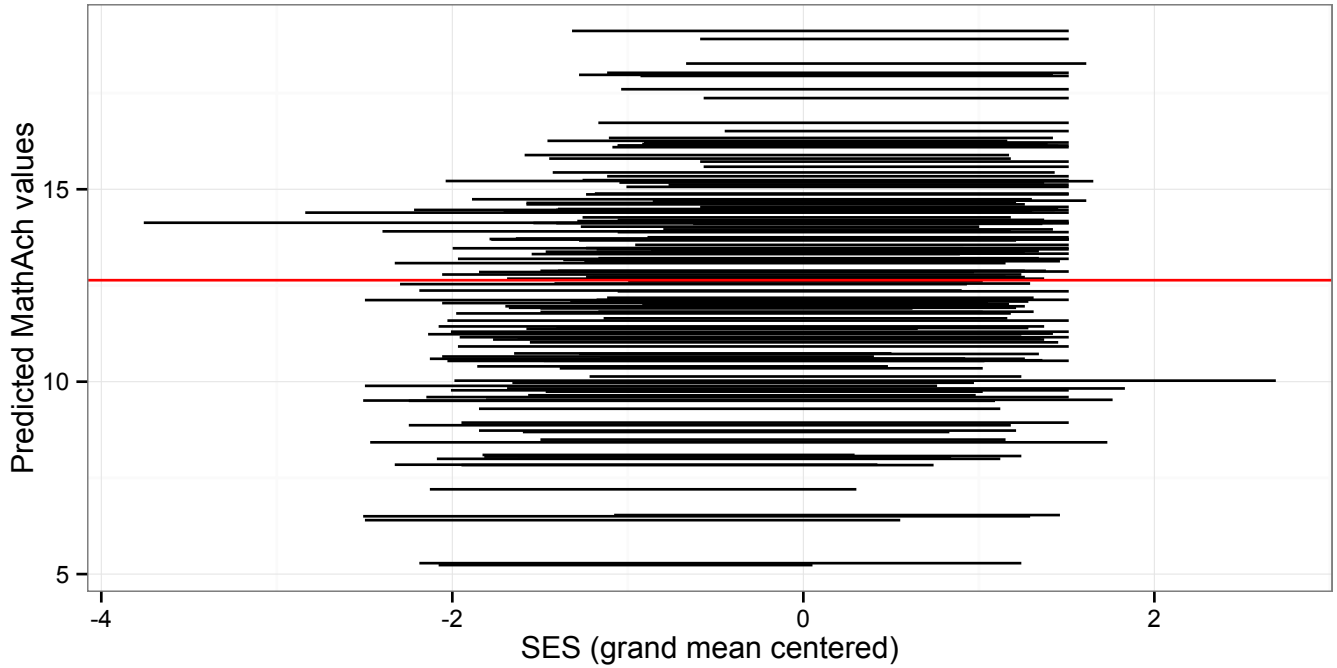
Random-Intercept-Modell ohne Prädiktor (= random effects ANOVA)


```
library(lme4)
lmer(MathAch ~ 1 + (1|School), data=HSB)
```

Kriterium | **Gesamt-intercept** | **Random intercept ...** | **für jede L2-Einheit**

```
Random effects:
 Groups   Name        Variance Std.Dev.
 School   (Intercept)  8.614   2.935
 Residual             39.148   6.257
Number of obs: 7185, groups:  School, 160

Fixed effects:
            Estimate Std. Error       df t value
(Intercept)  12.6370     0.2444 156.6500   51.71
```



Hinweis: Wir verwenden für lineare gemischte Modelle das *lme4* Paket in R (für die Details siehe Seminar im SoSe)

Random-Intercept-Modell ohne Prädiktor (= random effects ANOVA)

Varianzkomponenten:

σ^2: Zufallsfehlervarianz auf Level 1, d.h. Varianz der r_{ij}

τ_{00}: Varianz der Achsenabschnitte (Intercepts), d.h. Varianz der u_{0j}

Fixed effects:

γ_{00}: Erwartungswert für eine beliebige Person

```
Random effects:
 Groups   Name        Variance Std.Dev.
 School   (Intercept)  8.614   2.935
 Residual             39.148   6.257
Number of obs: 7185, groups:  School, 160

Fixed effects:
            Estimate Std. Error       df t value
(Intercept)  12.6370     0.2444 156.6500   51.71
```

geschätztes τ_{00}

geschätztes σ^2

geschätztes γ_{00}

$$ICC = \frac{\hat{\tau}_{00}}{\hat{\tau}_{00} + \hat{\sigma}^2} = \frac{8.614}{8.614 + 39.148} = 18\%$$

→ 18% der Varianz der Mathematik-Leistung geht auf Unterschiede zwischen den Schulen zurück, 82% auf Unterschiede innerhalb der Schulen.

Random-Intercept-Modell ohne Prädiktor: Einen Messwert zerlegen


```
Random effects:
 Groups   Name        Variance Std.Dev.
 School   (Intercept)  8.614   2.935
 Residual             39.148   6.257
Number of obs: 7185, groups:  School, 160

Fixed effects:
            Estimate Std. Error       df t value
(Intercept)  12.6370     0.2444 156.6500   51.71
```

$$y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$$

$$5.876 = 12.637 - 2.664 - 4.097$$

geschätztes γ_{00}

y

geschätzte r_{ij} (Residuum auf L1)

geschätzte u_{0j} (Residuum auf L2)

	School	Minority	Sex	SES	MathAch	MEANSES	resid
1	1224	No	Female	-1.528	5.876	-0.428	-4.097039
2	1224	No	Female	-0.588	19.708	-0.428	9.734961
3	1224	No	Male	-0.528	20.349	-0.428	10.375961
4	1224	No	Male	-0.668	8.781	-0.428	-1.192039
5	1224	No	Male	-0.158	17.898	-0.428	7.924961
6	1224	No	Male	0.022	4.583	-0.428	-5.390039

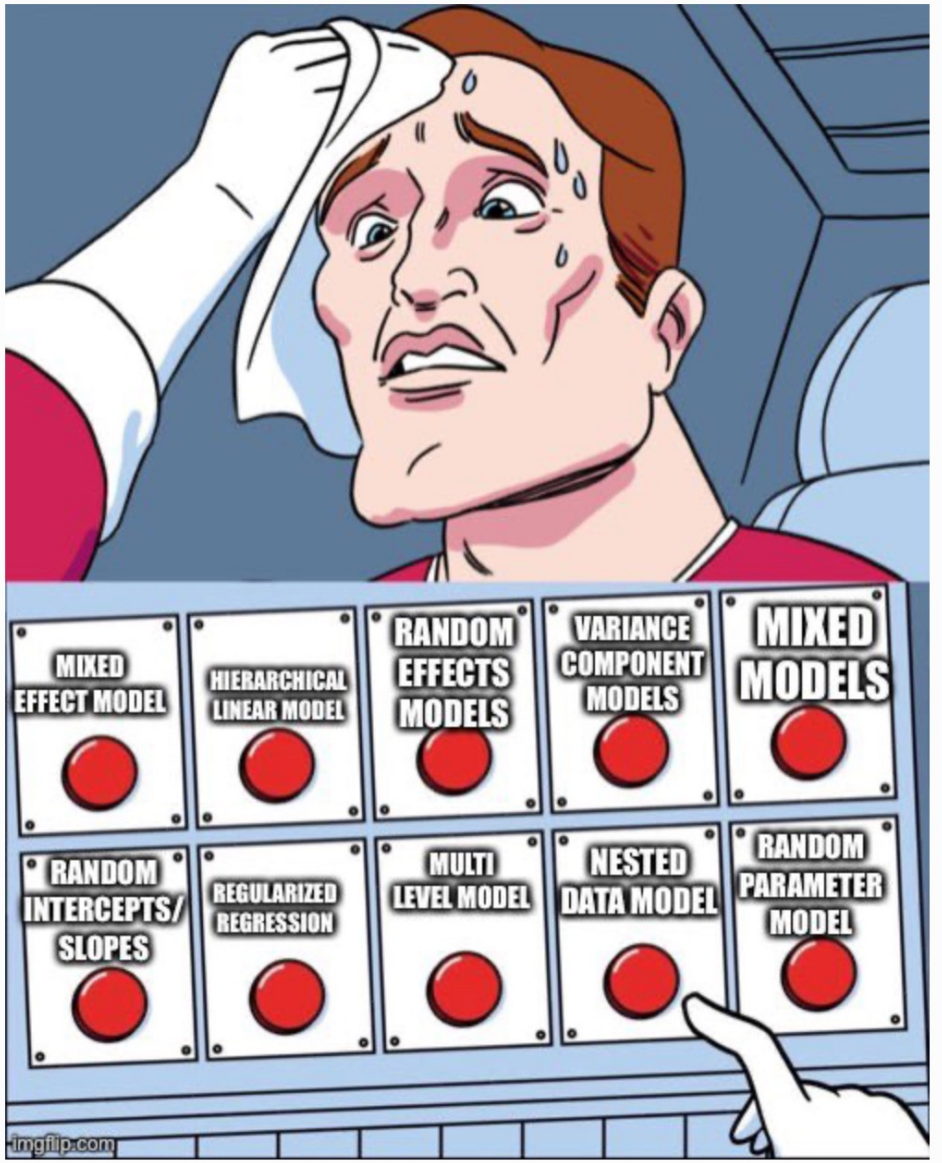
	school	(Intercept)
1	1224	-2.66393477
2	1288	0.73940980
3	1296	-4.56846563
4	1308	2.94851711
5	1317	0.49394606
6	1358	-1.24251161

Freiheitsgrade und *p*-Werte in LMM

- Hypothesentests in frequentistischen LMMs kompliziert: *df* abhängig von ICC
  - Extremfall ICC = 1: Es gibt keine Varianz innerhalb der Gruppen; (d.h., alle Messwerte innerhalb einer Gruppe haben den selben Wert → effektiv gibt es nur J (=Zahl der Gruppen) Freiheitsgrade (minus die Zahl der geschätzten Parameter) für einige Parameter
  - Extremfall ICC = 0: Die Gruppenzugehörigkeit klärt keine Varianz in den Messwerten auf (d.h., jede Person kann frei variieren
    → es gibt i (= Zahl der Personen) Freiheitsgrade (minus die Zahl der geschätzten Parameter)
- Bei 0 < ICC < 1: Effektive Zahl der Freiheitsgrade irgendwo zwischen Zahl der Gruppen (L2) und Zahl der Personen (L1)
- *Satterthwaite* oder *Kenward-Roger*-Korrektur der Freiheitsgrade liefert effektive *df* zwischen diesen beiden Extremen (kann auch eine Kommazahl sein)
- Achtung: LMM-Software berichtet manchmal keine *df* (*lme4*), oder eine der beiden Korrekturen – ohne *df* kein *p*-Wert!
- Das R-Paket *lmerTest* erweitert *lme4* und liefert als default *df* nach der Satterthwaite-Korrektur und *p*-Werte.
- Je nach angewandter *df*-Korrektur können sich *df* und *p*-Werte geringfügig zwischen den Softwarelösungen unterscheiden.

Exkurs: Viele Bezeichnungen für dieselben bzw. ähnliche Klassen von Modellen …

- Betonung der gruppierten Struktur:
  - Multilevel Model (MLM)
  - Linear Multilevel Model (LMLM)
  - Hierarchical Linear Model (HLM)
    - HLM ist auch Name einer Software
    - bezieht sich nur auf streng hierarchische Modelle; schließt genau genommen cross-classified mixed models aus.
- Betonung, dass feste *und* zufällige Koeffizienten modelliert werden:
  - Linear mixed models (LMM) / lineare gemischte Modelle (LGM):
    Mixed = fixed & random effects
  - Mixed effects models (MEM)
  - Linear mixed effects models (LMEM)
- Betonung der zufälligen Koeffizienten (dass zusätzlich auch feste Effekte drin sind, wird ohnehin angenommen):
  - Random effects models
  - Random coefficient models
- **Breiteste Anwendung: Linear Mixed Models / lineare gemischte Modelle (LMM / LGM)**



(meme courtesy of @ChelseaParlett)

Exkurs: Begriffsklärung – feste und zufällige Effekte

(Exkurs) Begriffsklärung: feste und zufällige Effekte (fixed vs. random)

[...] we briefly review **what is meant by fixed and random effects**. It turns out that different—in fact, incompatible—definitions are used in different contexts. Here we outline **five definitions** that we have seen:

1. Fixed effects are constant across individuals, and random effects vary. For example, in a growth study, a model with random intercepts α_i and fixed slope β corresponds to parallel lines for different individuals i, or the model $y_{it} = \alpha_i + \beta_t$. (Kreft and de Leeuw, 1998).
2. Effects are fixed if they are interesting in themselves or random if there is interest in the underlying population. (Searle, Casella and McCulloch, 1992)
3. “When a sample exhausts the population, the corresponding variable is fixed; when the sample is a small (i.e., negligible) part of the population the corresponding variable is random” (Green and Tukey, 1960).
4. “If an effect is assumed to be a realized value of a random variable, it is called a random effect” (LaMotte, 1983).
5. Fixed effects are estimated using least squares (or, more generally, maximum likelihood) and random effects are estimated with shrinkage. This definition is standard in the multilevel modeling literature (e.g., Snijders & Bosker, 1999) and in econometrics.

Gelman (2005, p. 20)

(Exkurs) Begriffsklärung: feste und zufällige Effekte (fixed vs. random)

[...] we briefly review **what is meant by fixed and random effects**. It turns out that different—in fact, incompatible—definitions are used in different contexts. Here we outline **five definitions** that we have seen:

1. Fixed effects are constant across individuals, and random effects vary. For example, in a growth study, a model with random intercepts α_i and fixed slope β corresponds to parallel lines for different individuals i, or the model $y_{it} = \alpha_i + \beta_t$. (Kreft and de Leeuw, 1998).

We prefer to sidestep the overloaded terms “fixed” and “random” with a cleaner distinction by simply renaming the terms in definition 1 above. We define effects (or coefficients) in a multilevel model as **constant** if they are identical for all groups in a population and **varying** if they are allowed to differ from group to group.

Gelman (2005, p.21)