AKiP Forschungscamp, 14.–15. September 21

Kurzworkshop Mehrebenenmodelle

Christopher Hautmann

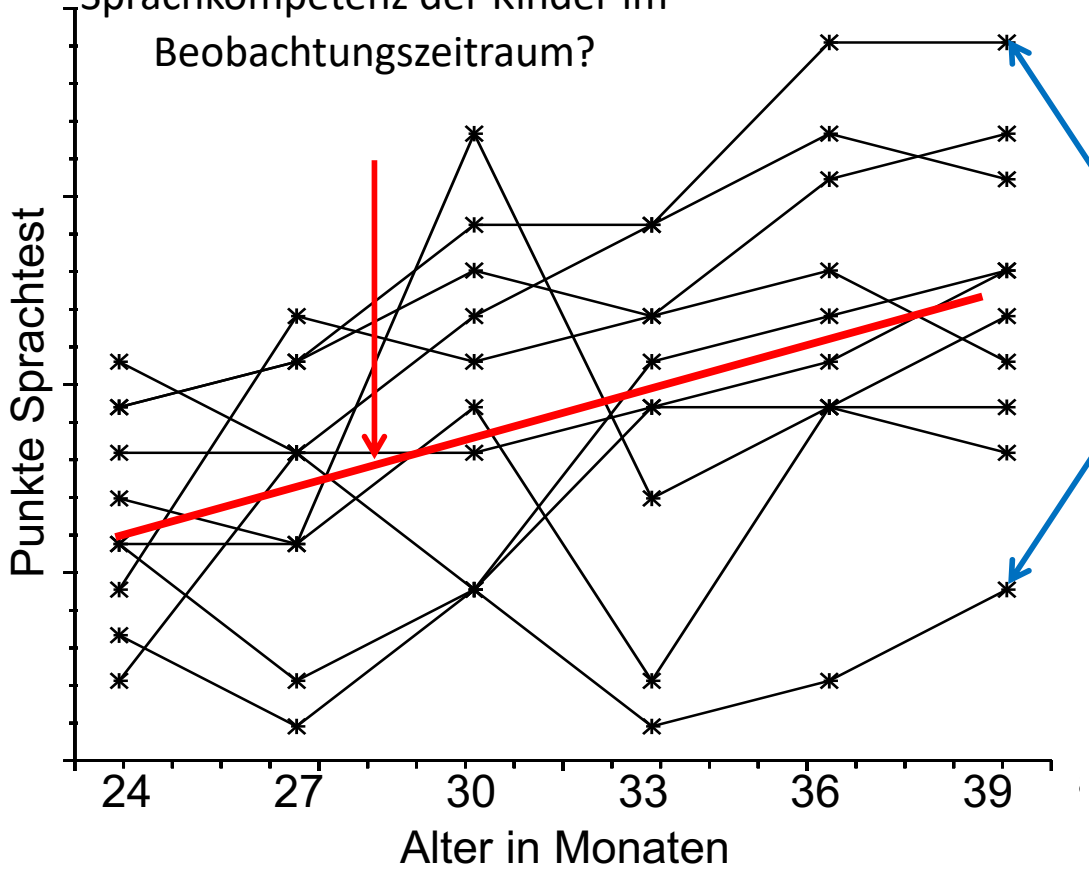
Ausbildungsinstitut für Kinder- und Jugendlichenpsychotherapie
am Klinikum der Universität zu Köln
www.akip.de

UNIKLINIK KÖLN

akipköln

Multilevel Modeling und Längsschnittdaten

Frage 1
Verbessert sich die Sprachkompetenz der Kinder im Beobachtungszeitraum?

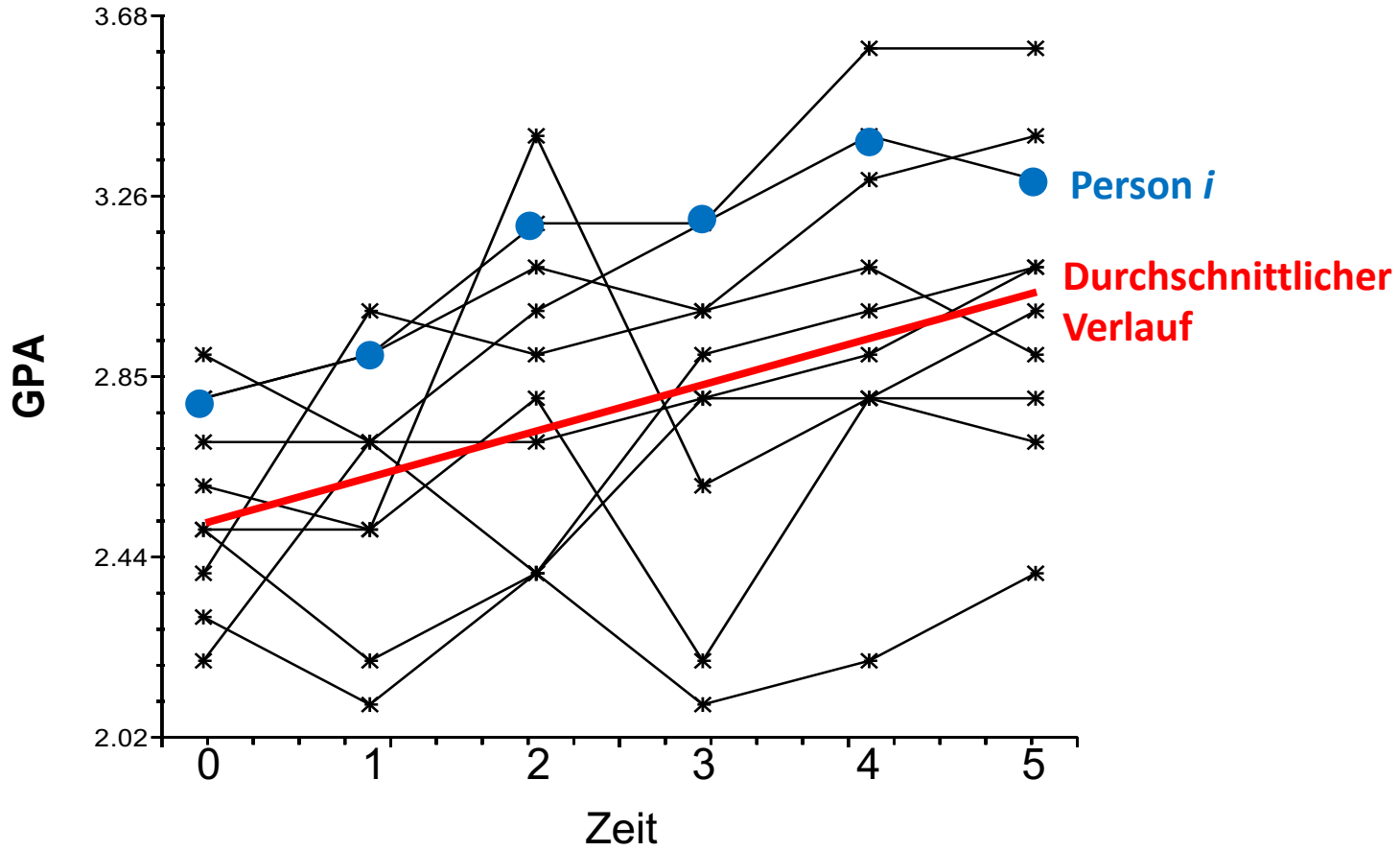


Frage 2
Gibt es interindividuelle Unterschiede in der Entwicklung der Sprachkompetenz?

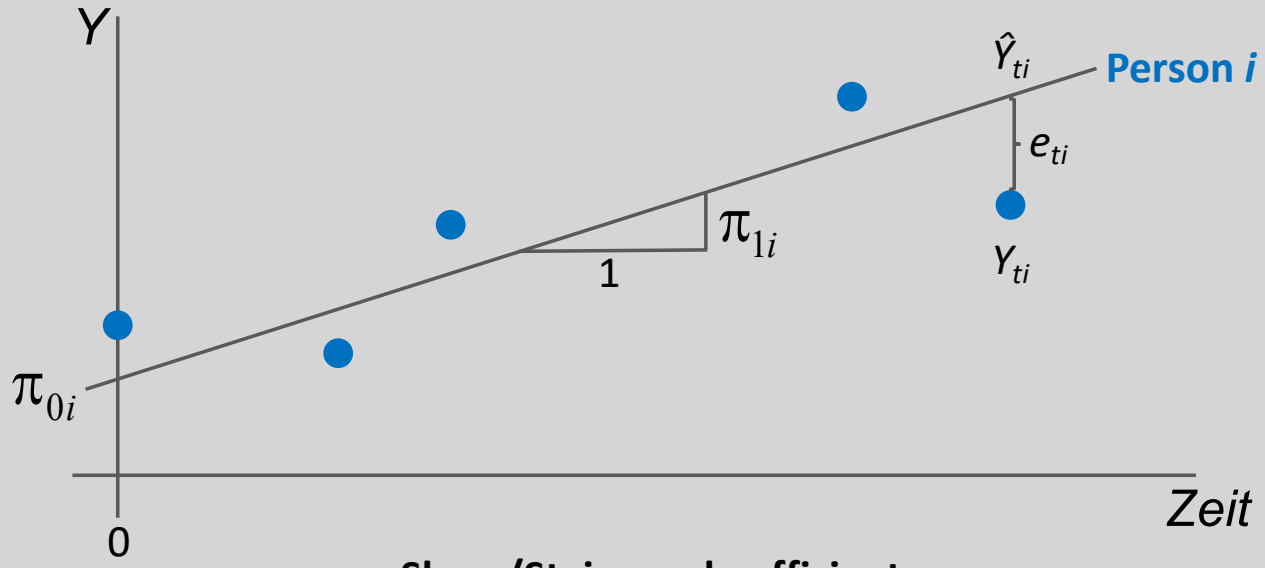
↓

Frage 3
Welche Faktoren können diese Unterschiede erklären (z. B. Geschlecht)?

Multilevel Modeling und Längsschnittdaten



Definition des Verlaufs von Person i



Slope/Steigungskoeffizient
Veränderung in $\hat{Y}$ von Person *i*, wenn die Prädiktorvariable *ZEIT* um den Wert 1 ansteigt

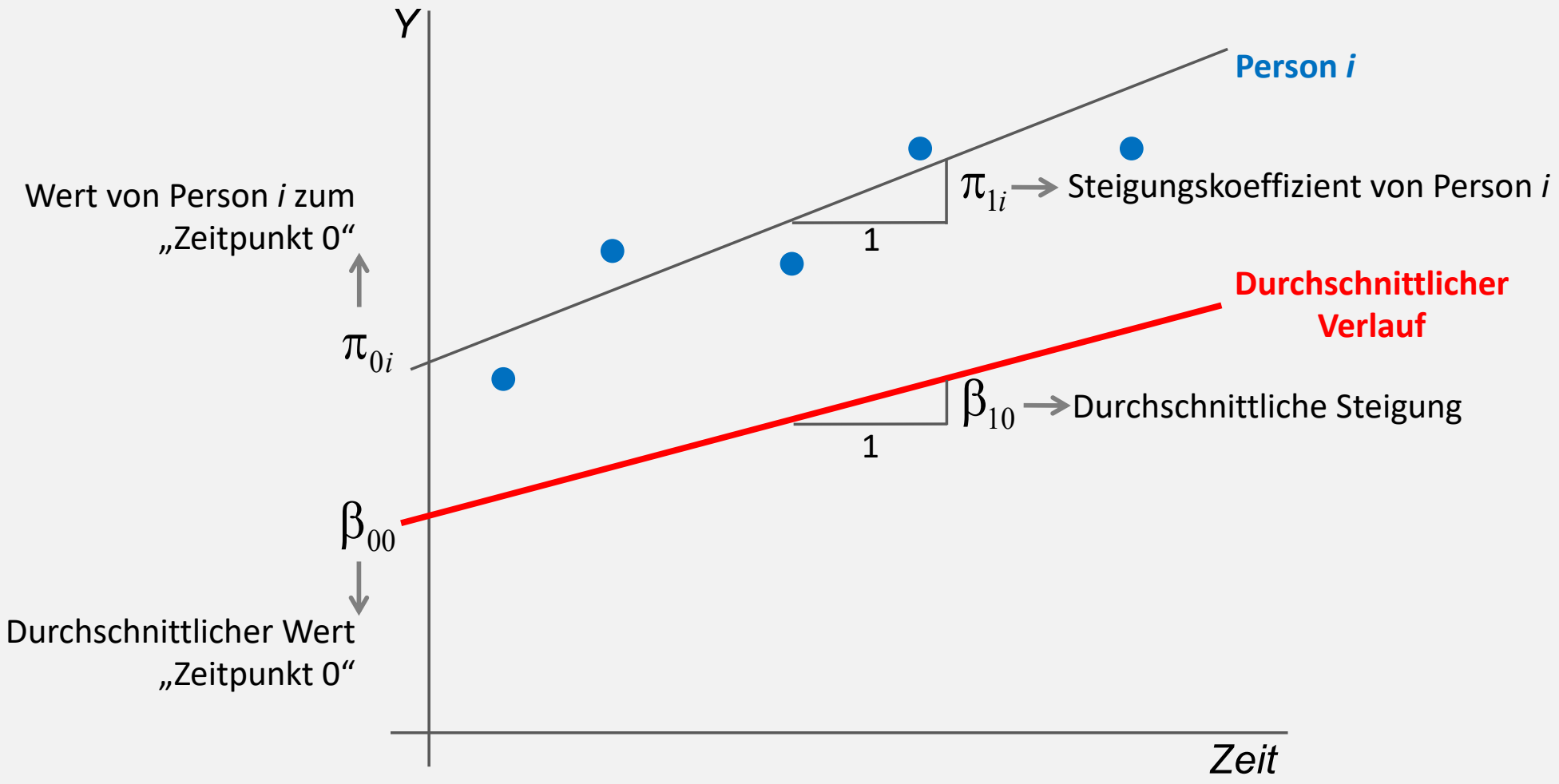
Beobachteter Wert Kriteriumsvariable *Y* von Person *i* zum Messzeitpunkt *t*

$$Y_{ti} = \pi_{0i} + \pi_{1i} ZEIT_{ti} + e_{ti}$$

Intercept/ Achsenabschnitt
Vorhergesagter Wert von Person *i* zum „Zeitpunkt 0“

Residuum
Differenz beobachteter Wert *Y* und vorhergesagter Wert $\hat{Y}$ von Person *i* zum Messzeitpunkt *t*

Individuelle Verläufe und die Population



Individuelle Verläufe und die Population

Level 1

$$Y_{ti} = \pi_{0i} + \pi_{1i} ZEIT_{ti} + e_{ti}$$

Level 2

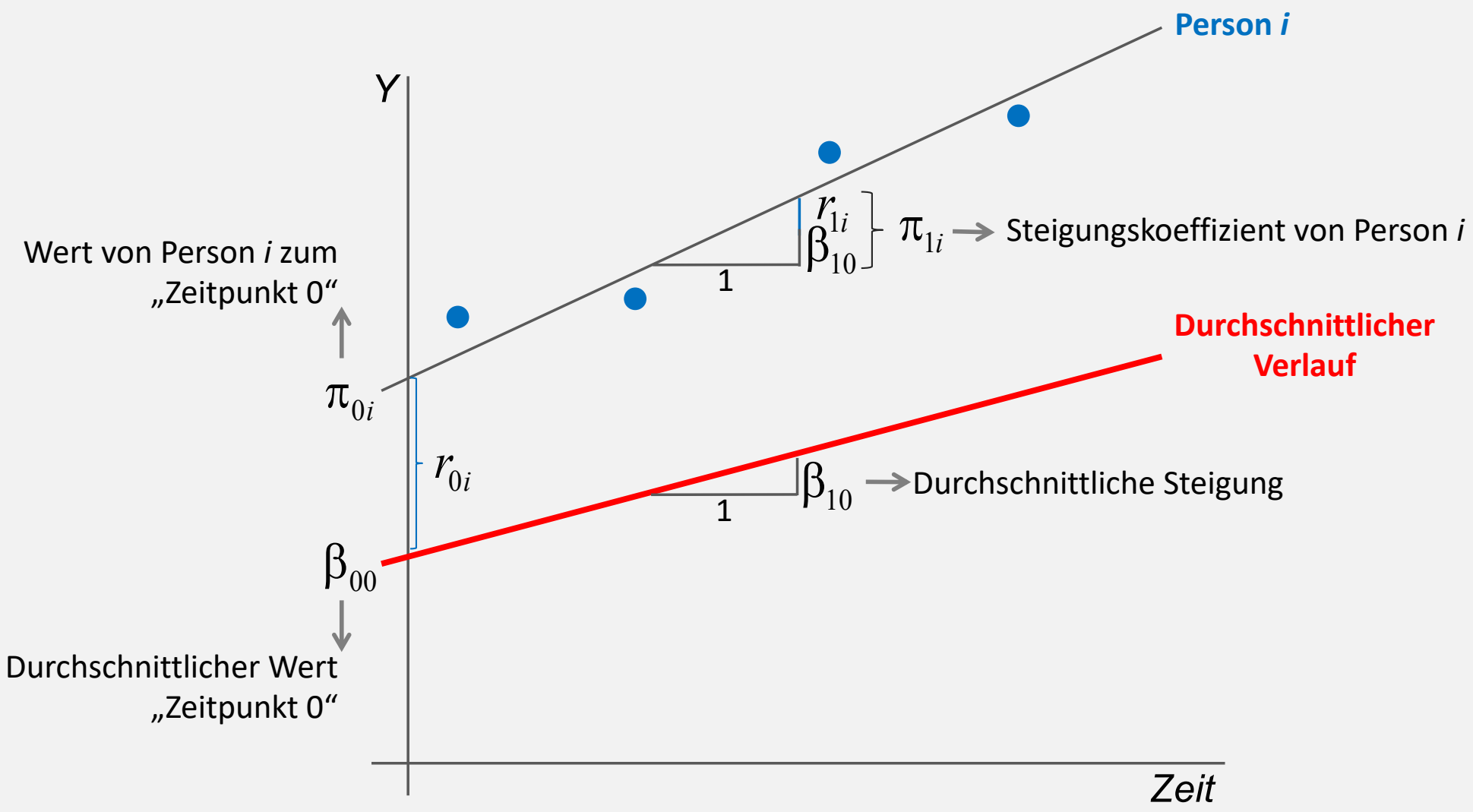
$$\pi_{0i} = \beta_{00} + r_{0i}$$

$$\pi_{1i} = \beta_{10} + r_{1i}$$

π_{0i}	Wert von Person *i* zum „Zeitpunkt 0“
β_{00}	Durchschnittlicher Wert aller Personen „Zeitpunkt 0“
r_{0i}	Individuelle Abweichung von Person *i* vom durchschnittlichen Wert „Zeitpunkt 0“

π_{1i}	Steigungskoffizient von Person *i*
β_{10}	Durchschnittliche Veränderung
r_{1i}	Individuelle Abweichung von Person *i* von der durchschnittlichen Veränderung

Individuelle Verläufe und die Population



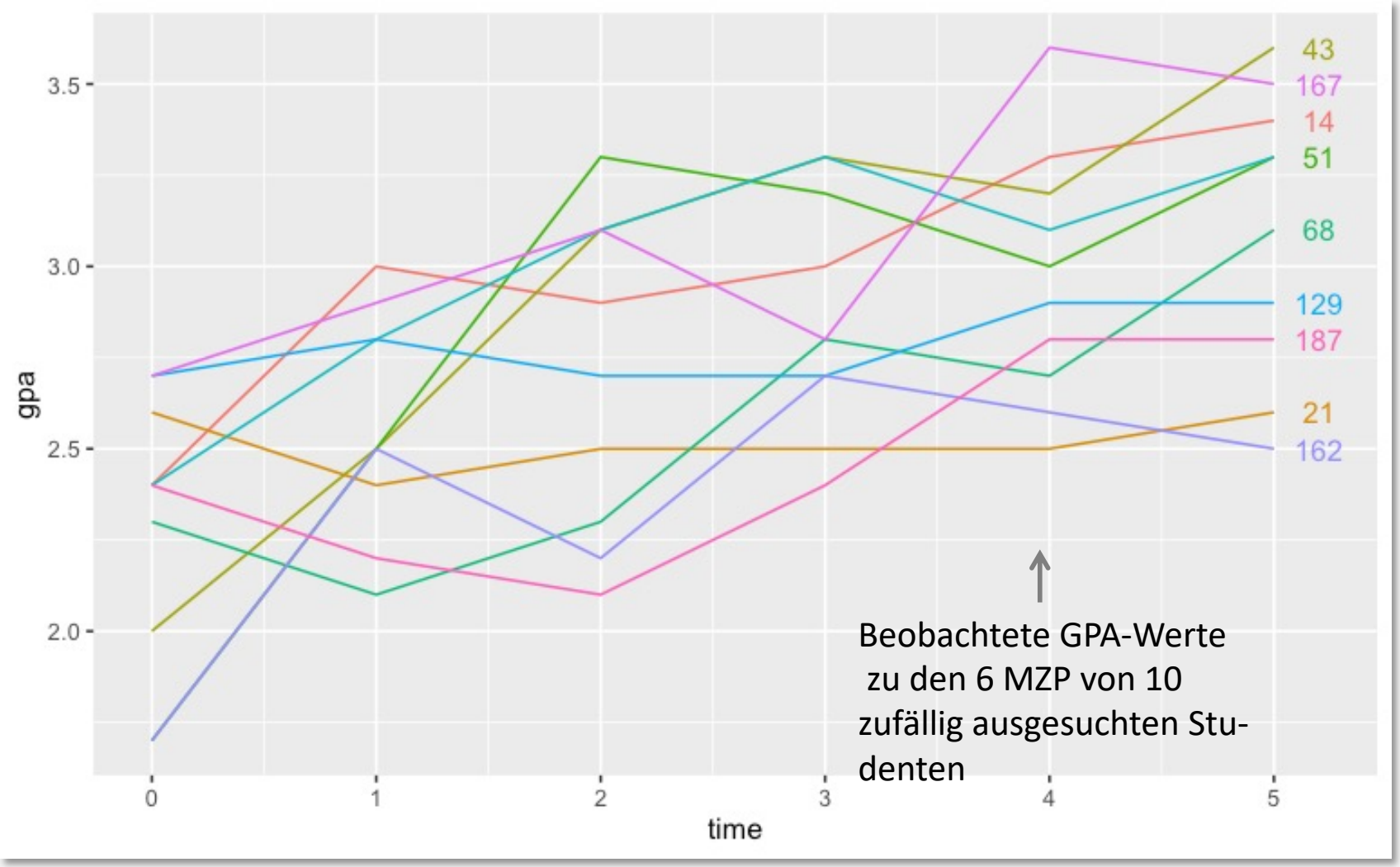
„Grade Point Average Data“ aus Hox (2010)

⇒ Quelle: Hox, J. J. (2010). *Multilevel Analysis: Techniques and applications* (2nd ed.). New York, NY: Routledge.

⇒ Datensatz: Grade Point Average (S. 81 ff. und S. 353)
Künstlich generierter Datensatz

⇒ Daten:
- 200 Studenten (Nutzung Studentenversion HLM möglich)
- AV: Grade Point Average (≈ Notendurchschnitt)
- 6 Messzeitpunkte (einer pro Semester)
- Kodierung Messzeitpunkte 0, …., 5 ⟶ Intercept kann als GPA-Wert zum 1. MZP interpretiert werden

„Grade Point Average Data“ aus Hox (2010)



„Grade Point Average Data“ aus Hox (2010)
SPSS-Daten – Wide Format

gpa2.sav [DataSet1] - IBM SPSS Statistics Dateneditor

Datei Bearbeiten Ansicht Daten Transformieren Analysieren Direktmarketing Grafik Extras Fenster Hilfe

1 :

	student	sex	gpa1	gpa2	gpa3	gpa4	gpa5	gpa6	var	var
1	1	1	2.3	2.1	3.0	3.0	3.0	3.3		
2	2	0	2.2	2.5	2.6	2.6	3.0	2.8		
3	3	1	2.4	2.9	3.0	2.8	3.3	3.4		
4	4	0	2.5	2.7	2.4	2.7	2.9	2.7		
5	5	0	2.8	2.8	2.8	3.0	2.9	3.1		
6	6	1	2.5	2.4	2.4	2.3	2.7	2.8		
7	7	0	2.4	2.4	2.8	2.6	3.0	3.0		
8	8	1	2.8	2.8	3.1	3.3	3.3	3.4		
9	9	0	2.8	2.7	2.7	3.1	3.1	3.5		
10	10	0	2.8	2.8	3.0	2.7	3.0	3.0		
11	11	1	2.6	2.9	3.2	3.6	3.6	3.8		
12	12	1	2.6	3.0	2.3	2.9	3.1	3.3		
13	13	0	2.8	3.1	3.5	3.6	3.9	3.9		
14	14	1	2.4	3.0	2.9	3.0	3.3	3.4		
15	15	0	2.2	2.7	2.6	2.6	2.8	2.8		
16	16	1	2.5	2.7	2.8	2.7	2.8	2.8		
17	17	0	2.6	2.2	2.6	2.5	2.6	2.8		
18	18	0	2.4	2.9	2.0	2.6	2.5	2.7		

„Grade Point Average Data“ aus Hox (2010)
SPSS-Daten – Long-Format

gpa2long.sav [DataSet2] - IBM SPSS Statistics Dateneditor

Datei Bearbeiten Ansicht Daten Transformieren Analysieren Direktmarketing Grafik Extras Fenster Hilfe

21 :

	student	occas	gpa	var	var	var	var	var
1	1	0	2.3					
2	1	1	2.1					
3	1	2	3.0					
4	1	3	3.0					
5	1	4	3.0					
6	1	5	3.3					
7	2	0	2.2					
8	2	1	2.5					
9	2	2	2.6					
10	2	3	2.6					
11	2	4	3.0					
12	2	5	2.8					
13	3	0	2.4					
14	3	1	2.9					
15	3	2	3.0					
16	3	3	2.8					
17	3	4	3.3					
18	3	5	3.4					

Übung 1–6

Random Intercept, Random Slope Modell

Level 1

$$Y_{ti} = \pi_{0i} + \pi_{1i} ZEIT_{ti} + e_{ti}$$

Level 2

$$\pi_{0i} = \beta_{00} + r_{0i}$$

$$\pi_{1i} = \beta_{10} + r_{1i}$$

π_{0i}	Wert von Person *i* zum „Zeitpunkt 0“	π_{1i}	Steigungskoffizient von Person *i*
β_{00}	Durchschnittlicher Wert aller Personen „Zeitpunkt 0“	β_{10}	Durchschnittliche Veränderung
r_{0i}	Individuelle Abweichung von Person *i* vom durchschnittlichen Wert „Zeitpunkt 0“ -> Varianz von r_{0i} sei σ_{r0}^2	r_{1i}	Individuelle Abweichung von Person *i* von der durchschnittlichen Veränderung -> Varianz von r_{1i} sei σ_{r1}^2

Beantwortbare Fragen:

1. Wie lautet der Durchschnittswert zum „Zeitpunkt 0“ und ist er von Null verschieden?
2. Wie lautet die durchschnittliche Steigung und ist sie von Null verschieden?
3. Wie lautet die Streuung der Werte zum „Zeitpunkt 0“ und ist sie von Null verschieden?
4. Wie lautet die Streuung der Steigungskoeffizienten und ist sie von Null verschieden?

Spezifikation Random Intercept, Random Slope Modell mit dem Paket nlme

```
library("nlme")

gpa_lme <- lme(gpa ~ time, random = ~ 1 + time | student_id, data = data_gpa)
```

Level 1

$$Y_{ti} = \pi_{0i} + \pi_{1i} ZEIT_{ti} + e_{ti}$$

Durchschnittliche Veränderung

Level 2

$$\pi_{0i} = \beta_{00} + r_{0i}$$

$$\pi_{1i} = \beta_{10} + r_{1i}$$

π_{0i} Wert von Person *i* zum „Zeitpunkt 0“

β_{00} Durchschnittlicher Wert aller Personen „Zeitpunkt 0“

r_{0i} Individuelle Abweichung von Person *i* vom durchschnittlichen Wert „Zeitpunkt 0“
-> Varianz von r_{0i} sei σ_{r0}^2

π_{1i} Steigungskoffizient von Person *i*

β_{10}

r_{1i} Individuelle Abweichung von Person *i* von der durchschnittlichen Veränderung
-> Varianz von r_{1i} sei σ_{r1}^2

Beispiel „Grade Point Average Data“
Random Intercept, Random Slope Modell

```
> summary(gpa_lme)
Linear mixed-effects model fit by REML
  Data: data_gpa
       AIC      BIC    logLik
  272.9566 303.487 -130.4783

Random effects:
 Formula: ~1 + time | student_id
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev     Corr
(Intercept) 0.21259026 (Intr)
time        0.06711178 -0.098
Residual    0.20588283

Fixed effects:  gpa ~ time
                Value   Std.Error  DF   t-value p-value
(Intercept) 2.5992143 0.018357249 999 141.59062       0
time        0.1063143 0.005884789 999  18.06595       0
 Correlation:
     (Intr)
time -0.345
```

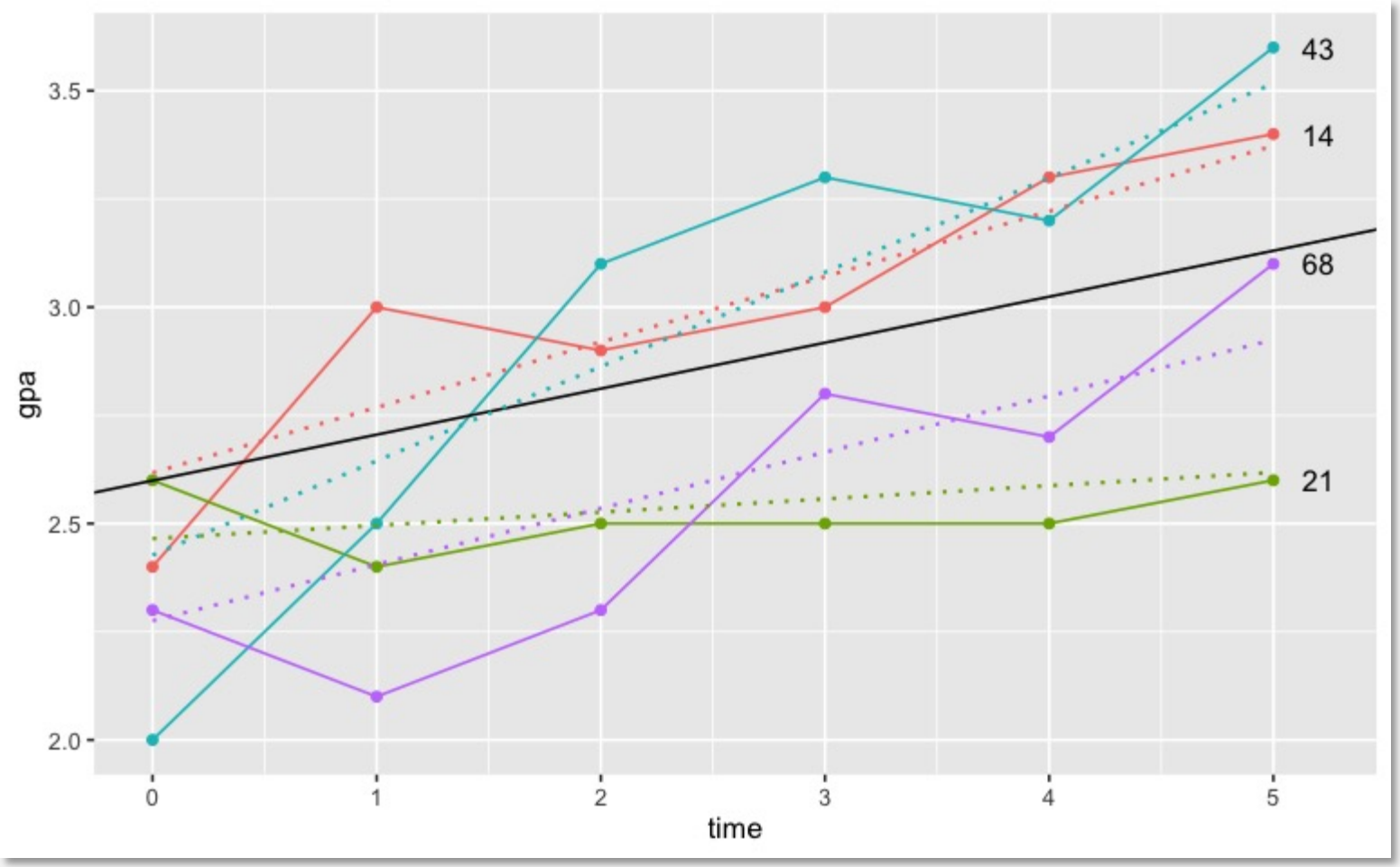
Beispiel „Grade Point Average Data“
Random Intercept, Random Slope Modell

```
> intervals(gpa_lme)
Approximate 95% confidence intervals

 Fixed effects:
                 lower      est.     upper
(Intercept) 2.56319109 2.5992143 2.6352375
time        0.09476632 0.1063143 0.1178623
attr(,"label")
[1] "Fixed effects:"

 Random Effects:
  Level: student_id
                           lower        est.      upper
sd((Intercept))        0.18367006  0.21259026 0.24606414
sd(time)               0.05770375  0.06711178 0.07805371
cor((Intercept),time) -0.28773958 -0.09825961 0.09862304
```

Beispiel „Grade Point Average Data“
Random Intercept, Random Slope Modell



Übung 7–9

Multilevel Modeling

- Random Intercept, Fixed Slope Modell
- Random Intercept, Random Slope Modell
- Intercepts- and Slopes-as-Outcomes Modell

Random Intercept, Fixed Slope Modell

Level 1

$$Y_{ti} = \pi_{0i} + \pi_1 ZEIT_{ti} + e_{ti}$$

Level 2

$$\pi_{0i} = \beta_{00} + r_{0i}$$

Alle Personen haben den gleichen Steigungskoeffizienten

$$\pi_1 = \beta_{10}$$

π_{0i} Wert von Person *i* zum „Zeitpunkt 0“

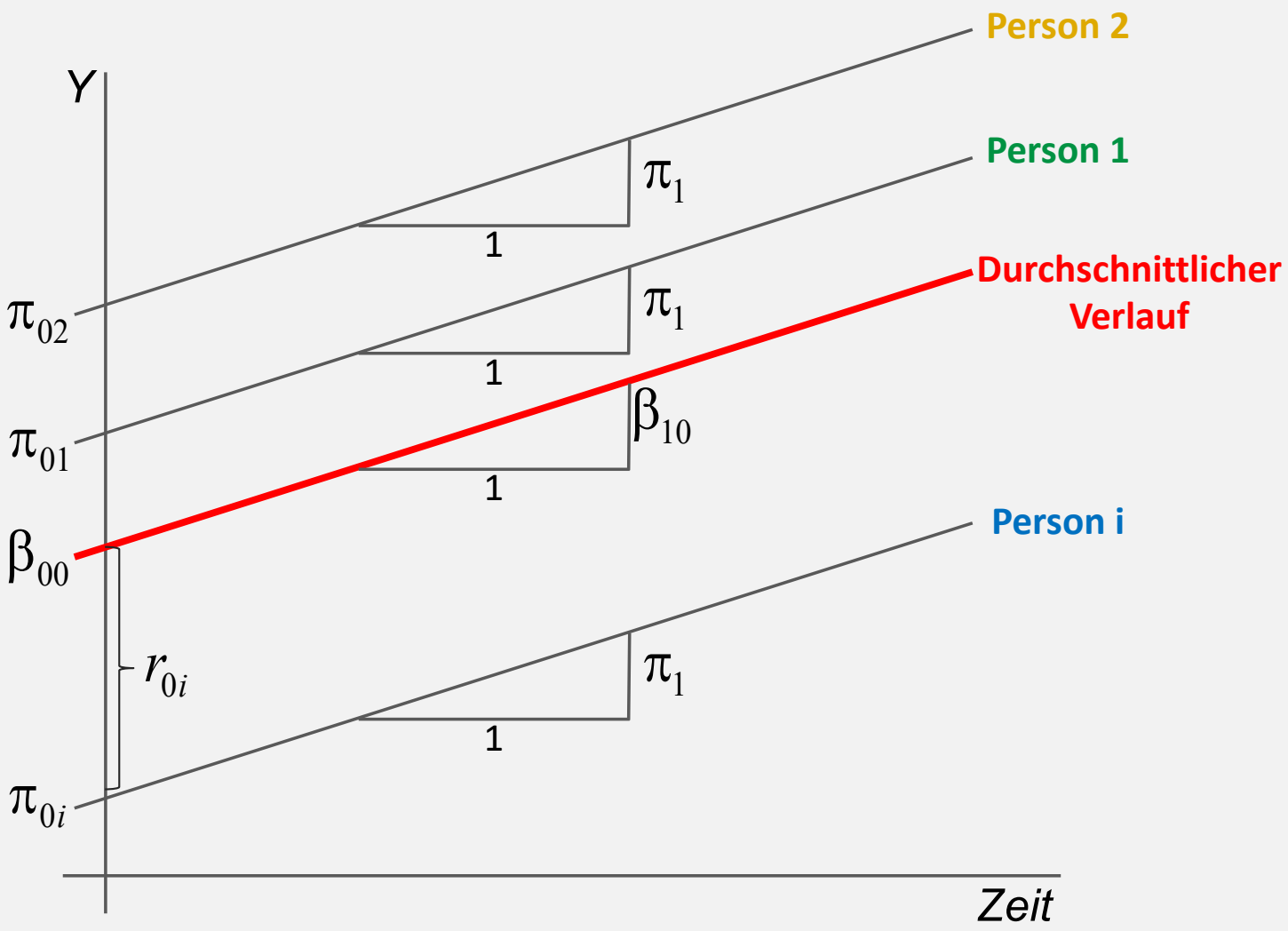
β_{00} Durchschnittlicher Wert aller Personen „Zeitpunkt 0“

r_{0i} Individuelle Abweichung von Person *i* vom durchschnittlichen Wert „Zeitpunkt 0“

π_1 Individueller Steigungskoeffizient

β_{10} Durchschnittliche Veränderung

Random Intercept, Fixed Slope Modell



Beispiel „Grade Point Average Data“ Random Intercept, Fixed Slope Modell

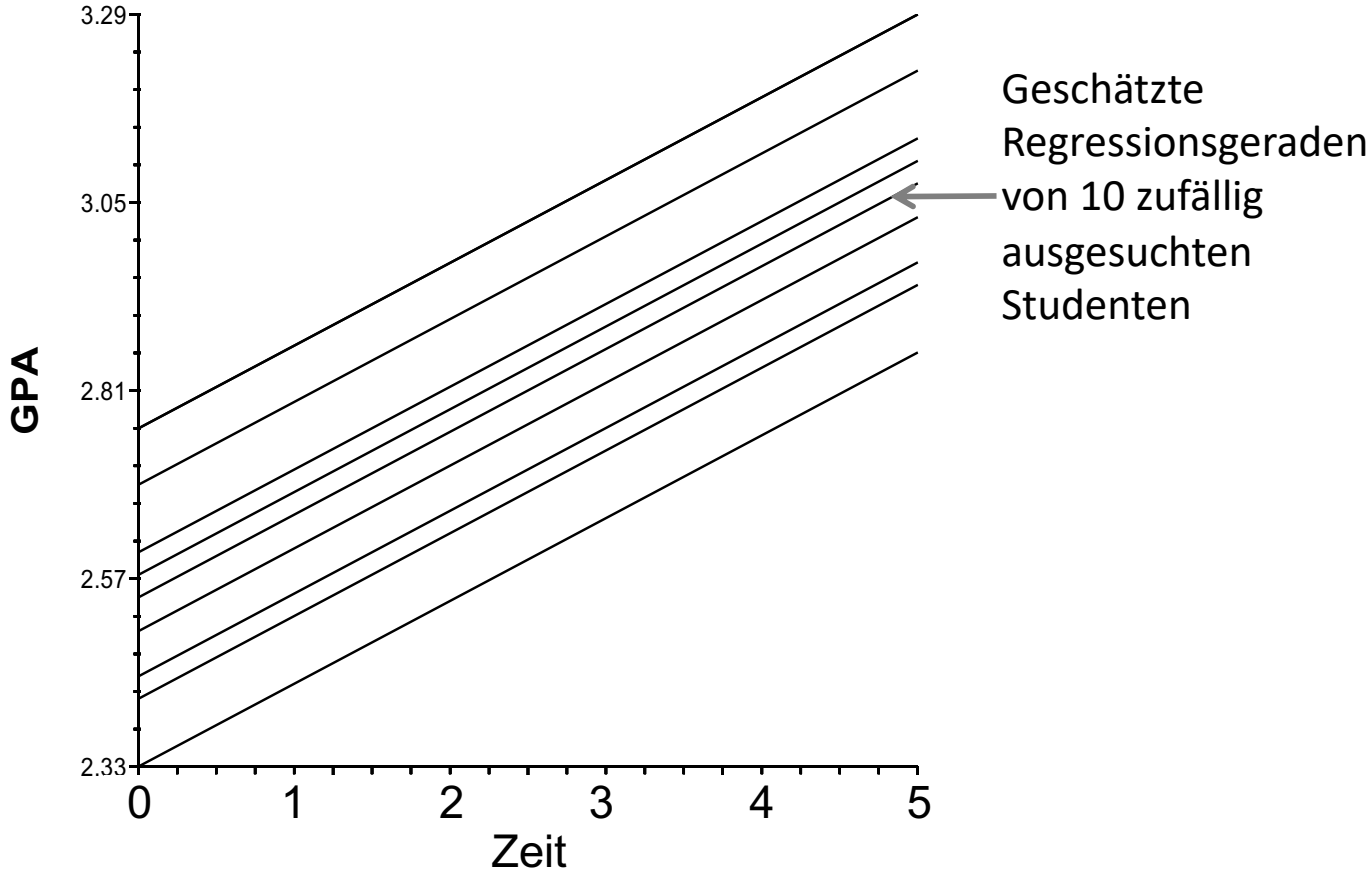
```
gpa_fixed_slope <- lme(gpa ~ time, random = ~ 1 | student_id, data = data_gpa)
```

```
> summary(gpa_fixed_slope)
Linear mixed-effects model fit by REML
  Data: data_gpa
       AIC      BIC    logLik
  416.8929 437.2465 -204.4464

Random effects:
 Formula: ~1 | student_id
        (Intercept)  Residual
StdDev:   0.2524261 0.2410156

Fixed effects:  gpa ~ time
                Value   Std.Error  DF   t-value p-value
(Intercept) 2.5992143 0.021696341 999 119.79966       0
time        0.1063143 0.004073908 999  26.09639       0
 Correlation:
     (Intr)
time -0.469
```

Beispiel „Grade Point Average Data“ Random Intercept, Fixed Slope Modell



Geschätzte Regressionsgeraden von 10 zufällig ausgesuchten Studenten

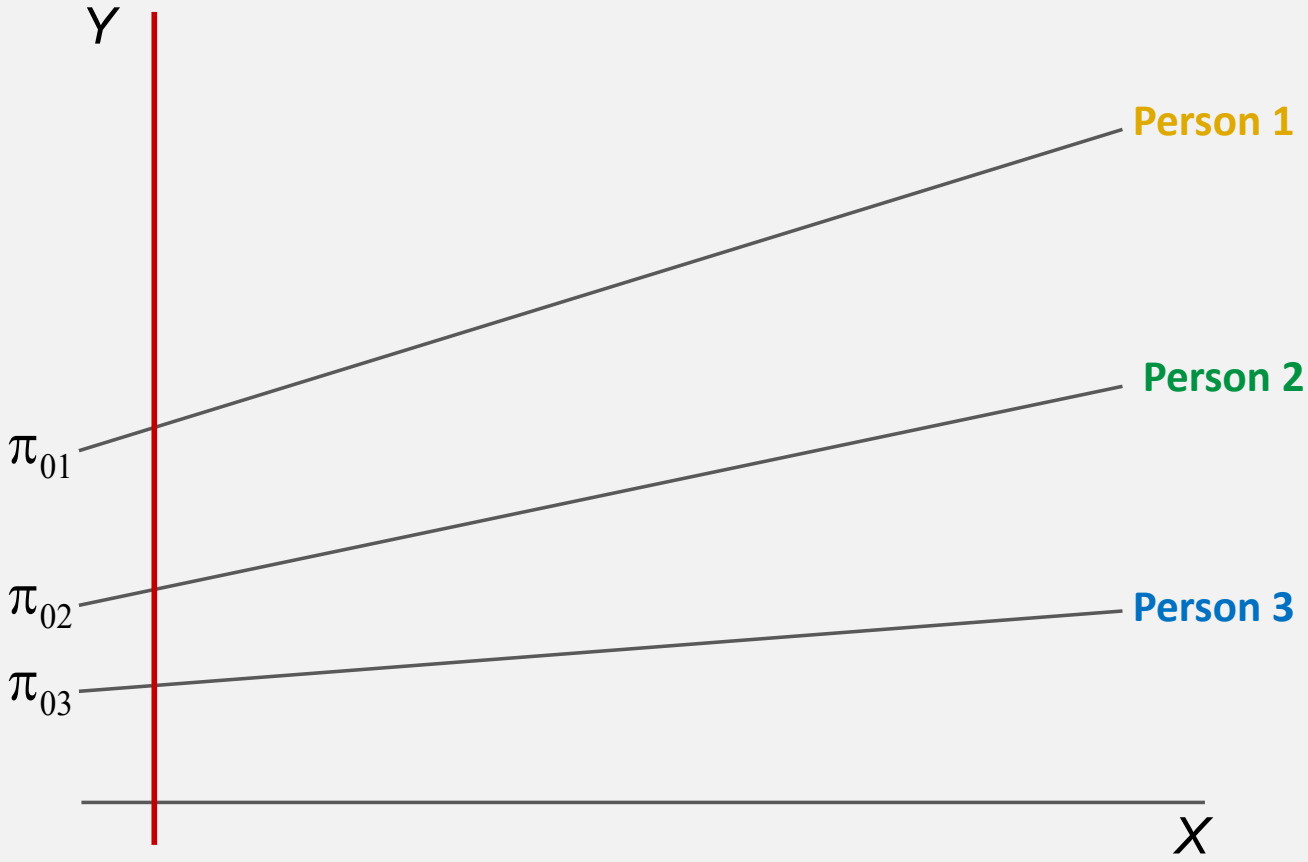
Übung 10–12

Kodierung Zeit und Bedeutung Intercept

$$\hat{Y}_{ti} = \pi_{0i} + \pi_{1i} ZEIT_{ti}$$

$$\hat{Y}_{ti} = \pi_{0i} + \pi_{1i} 0$$

$$\hat{Y}_{ti} = \pi_{0i}$$



Messzeitpunkt	1	2	3	4	5
Zeitkodierung A	0	1	2	3	4

Kodierung Zeit und Bedeutung des Intercept

	Messzeitpunkt				
	1	2	3	4	5
Zeitkodierung A	**0**	1	2	3	4

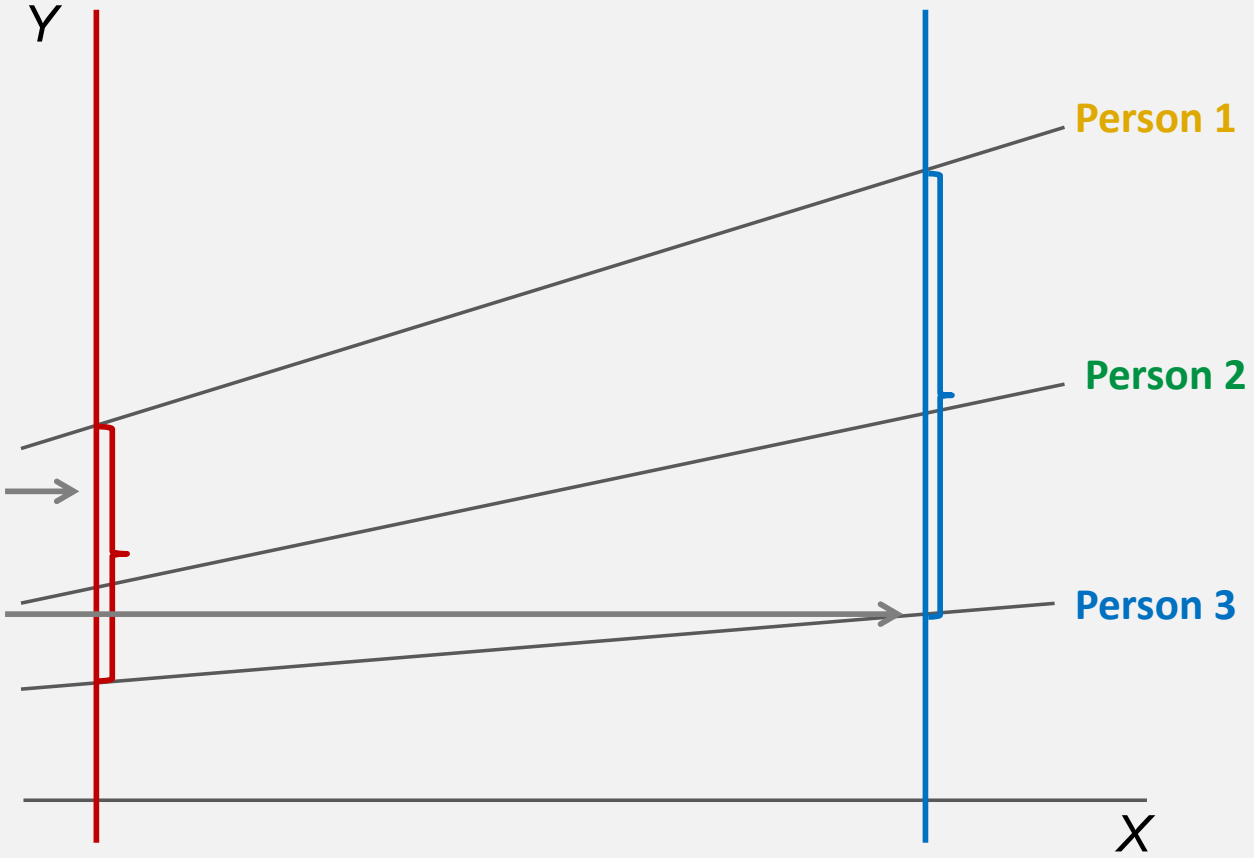
π_{0i} entspricht dem vorhergesagten Wert von Person *i* zum Beginn der Studie

	Messzeitpunkt				
	1	2	3	4	5
Zeitkodierung B	-4	-3	-2	-1	**0**

π_{0i} entspricht dem vorhergesagten Wert von Person *i* am Ende der Studie

Kodierung Zeit und Modellparameter

Im Random Intercept, Random Slope Modell ist die Varianz im Intercept gegenüber verschiedenen Zentrierungsformen nicht invariant



Messzeitpunkt	1	2	3	4	5
Zeitkodierung A	0	1	2	3	4
Zeitkodierung B	-4	-3	-2	-1	0

Kodierung Zeit und Modellparameter
Bsp. „GPA Data“ - Random Intercept, Random Slope Modell

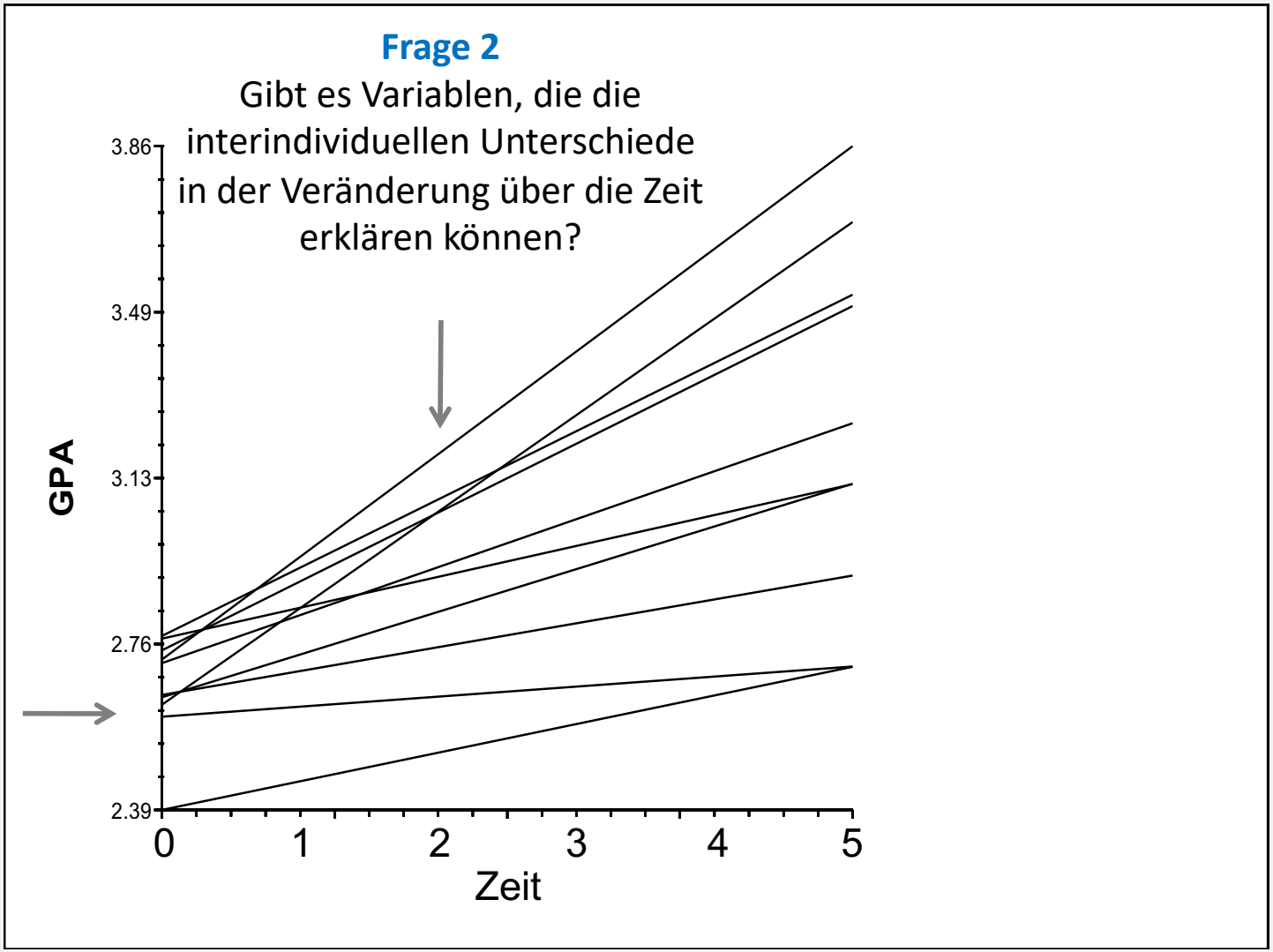
	Modell 1 Erster MZP hat Wert 0	Modell 2 Letzter MZP hat Wert 0
β_{00} ø Intercept	2.599214	3.130786
β_{10} ø Slope	0.106314	0.106314
τ_{00} Varianz r_{0i}	0.04519	0.14378
τ_{11} Varianz r_{1i}	0.00450	0.00450
σ^2 Varianz e_{ti}	0.04239	0.04239

Multilevel Modeling

- Random Intercept, Fixed Slope Modell
- Random Intercept, Random Slope Modell
- Intercepts- and Slopes-as-Outcomes Modell

Beispiel „Grade Point Average Data“ Intercepts- and Slopes-as-Outcomes Modell

Frage 1
Gibt es Variablen, die die interindividuellen Unterschiede zum „Zeitpunkt 0“ erklären können?



Intercepts- and Slopes-as-Outcomes Modell

Level 1

$$Y_{ti} = \pi_{0i} + \pi_{1i} ZEIT_{ti} + e_{ti}$$

Level 2

$$\pi_{0i} = \beta_{00} + \beta_{01} \overbrace{X_{1i}}^{\text{Level 2 Prädiktorvariable}} + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11} \overbrace{X_{1i}}^{\text{Level 2 Prädiktorvariable}} + r_{1i}$$

π_{0i} Wert von Person *i* zum „Zeitpunkt 0“

β_{00} Durchschnittlicher Zustand zum „Zeitpunkt 0“ von Personen, die in Bezug auf die Prädiktor-variable X_1 die Ausprägung Null haben

β_{01} Wenn β_{01} signifikant von 0 verschieden ist, eignet sich die Variable X_1 zur Vorhersage der Werte von „Zeitpunkt 0“

π_{1i} Steigungskoeffizient von Person *i*

β_{10} Durchschnittlicher Steigungskoeffizient von Personen, die in Bezug auf die Prädiktorvariable X_1 die Ausprägung Null haben

β_{11} Wenn β_{11} signifikant von 0 verschieden ist, eignet sich die Variable X_1 zur Erklärung inter-individueller Unterschiede in den Steigungs-koeffizienten

Intercepts- and Slopes-as-Outcomes Modell

$$Y_{ti} = \pi_{0i} + \pi_{1i} ZEIT_{ti} + e_{ti}$$

$$= \boxed{\beta_{00} + \beta_{01} X_{1i} + r_{0i}} + \boxed{(\beta_{10} + \beta_{11} X_{1i} + r_{1i})} ZEIT_{ti} + e_{ti}$$

$$= \beta_{00} + \beta_{01} X_{1i} + \beta_{10} ZEIT_{ti} + \beta_{11} \boxed{X_{1i} ZEIT_{ti}} + r_{0i} + r_{1i} ZEIT_{ti} + e_{ti}$$

↓

Interaktionsterm

Wenn die Interaktion einen bedeutsamen Effekt hat ($\beta_{11} \neq 0$),
ist die Veränderung über die Zeit abhängig von X_1

↓

X_1 als Moderatorvariable für die Veränderung über die Zeit

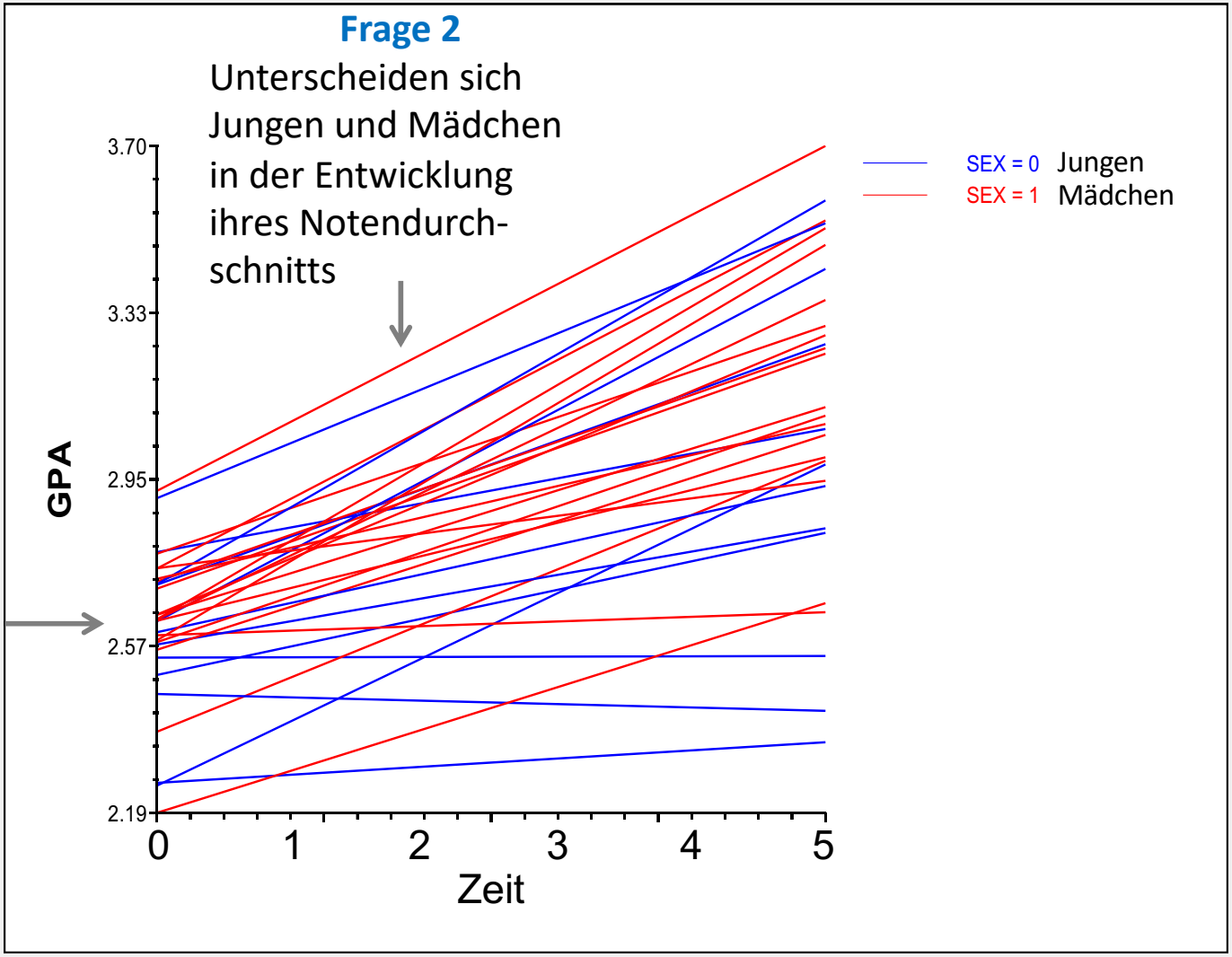
```
lme(gpa ~ time*sex, random = ~ 1 + time | student_id, data = gpa_data_long)
```

Übung 13

Beispiel „Grade Point Average Data“ Intercepts- and Slopes-as-Outcomes Modell

Frage 1

Unterscheidet sich der Notendurchschnitt von Jungen und Mädchen im ersten Semester



Beispiel „Grade Point Average Data“
Intercepts- and Slopes-as-Outcomes Modell

Level 1

$$GPA_{ti} = \pi_{0i} + \pi_{1i} ZEIT_{ti} + e_{ti}$$

Level 2

$\Rightarrow$ Level 2 Prädiktor SEX

$$\left.\begin{array}{ll} \text{Jungen} & \rightarrow 0 \\ \text{Mädchen} & \rightarrow 1 \end{array}\right\} \text{Dummy-Codierung}$$

$$\pi_{0i} = \beta_{00} + \beta_{01} SEX_i + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11} SEX_i + r_{1i}$$

Beispiel „Grade Point Average Data“
Intercepts- and Slopes-as-Outcomes Modell

⇒ Intercept Jungen

$$\pi_{0i} = \beta_{00} + \beta_{01} SEX_i + r_{0i}$$

Individueller Intercept von Person *i*

$$\pi_{0i} = \beta_{00} + \beta_{01} 0 + r_{0i}$$

$$\pi_{0i} = \beta_{00} + r_{0i}$$

Durchschnittlicher
Wert Jungen „Zeitpunkt 0“

Beispiel „Grade Point Average Data“ Intercepts- and Slopes-as-Outcomes Modell

⇒ Intercept Mädchen

$$\pi_{0i} = \beta_{00} + \beta_{01} SEX_i + r_{0i}$$

Individueller Intercept von Person *i*

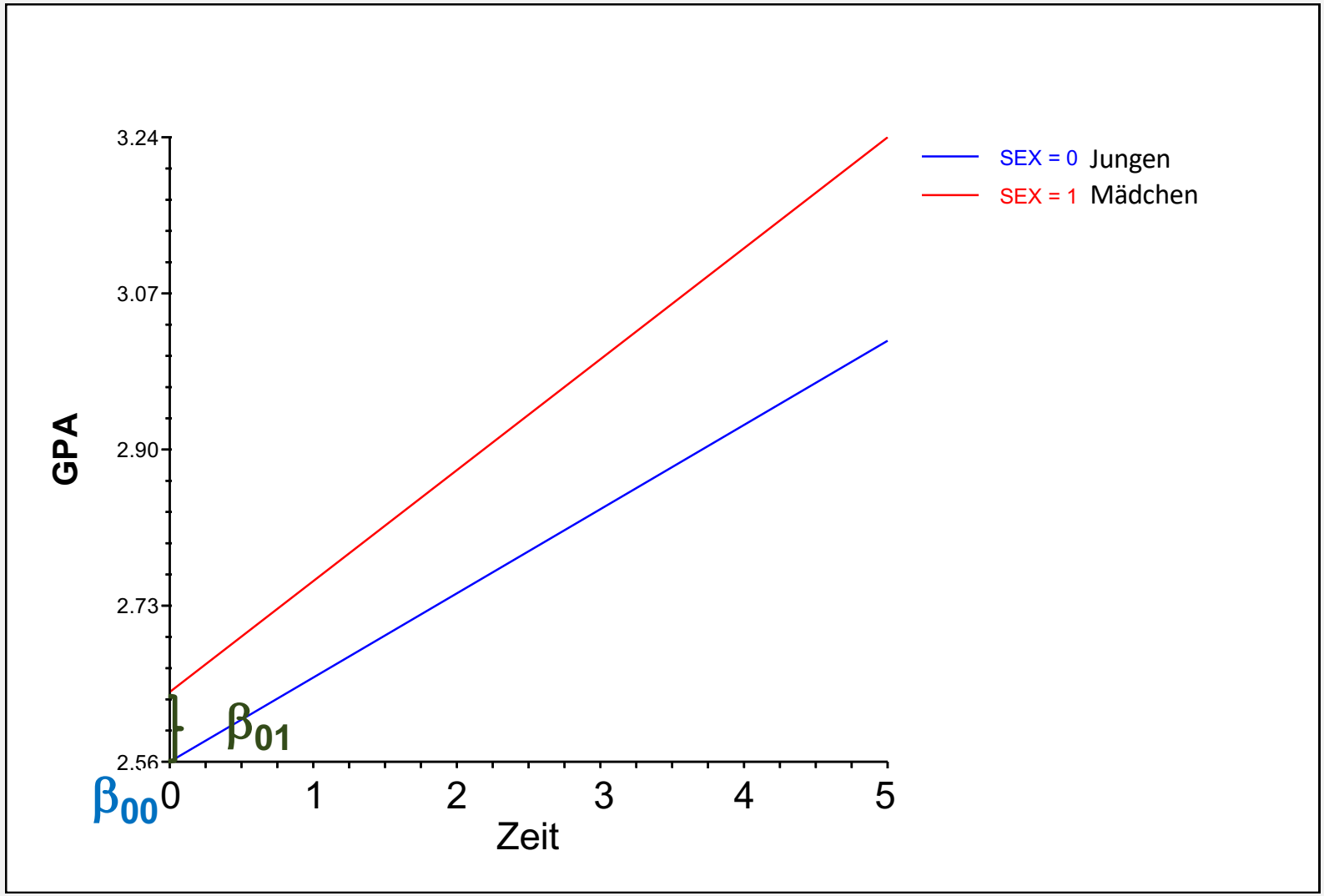
Durchschnittlicher Wert Mädchen „Zeitpunkt 0“

$$\pi_{0i} = \overbrace{\beta_{00} + \beta_{01}}\, 1 + r_{0i}$$

Durchschnittlicher Wert Jungen „Zeitpunkt 0“

Differenz Jungen und Mädchen „Zeitpunkt 0“

Beispiel „Grade Point Average Data“
Intercepts- and Slopes-as-Outcomes Modell



Beispiel „Grade Point Average Data“ Intercepts- and Slopes-as-Outcomes Modell

⇒ Slope Jungen

$$\pi_{1i} = \beta_{10} + \beta_{11} SEX_i + r_{1i}$$

Individueller Slope von Person *i*

$$\pi_{1i} = \beta_{10} + \beta_{11} 0 + r_{1i}$$

$$\pi_{1i} = \beta_{10} + r_{1i}$$

Durchschnittliche Veränderung über die Zeit bei Jungen

Beispiel „Grade Point Average Data“ Intercepts- and Slopes-as-Outcomes Modell

⇒ Slope Mädchen

$$\pi_{1i} = \beta_{10} + \beta_{11} SEX_i + r_{1i}$$

Individueller Slope von Person *i*

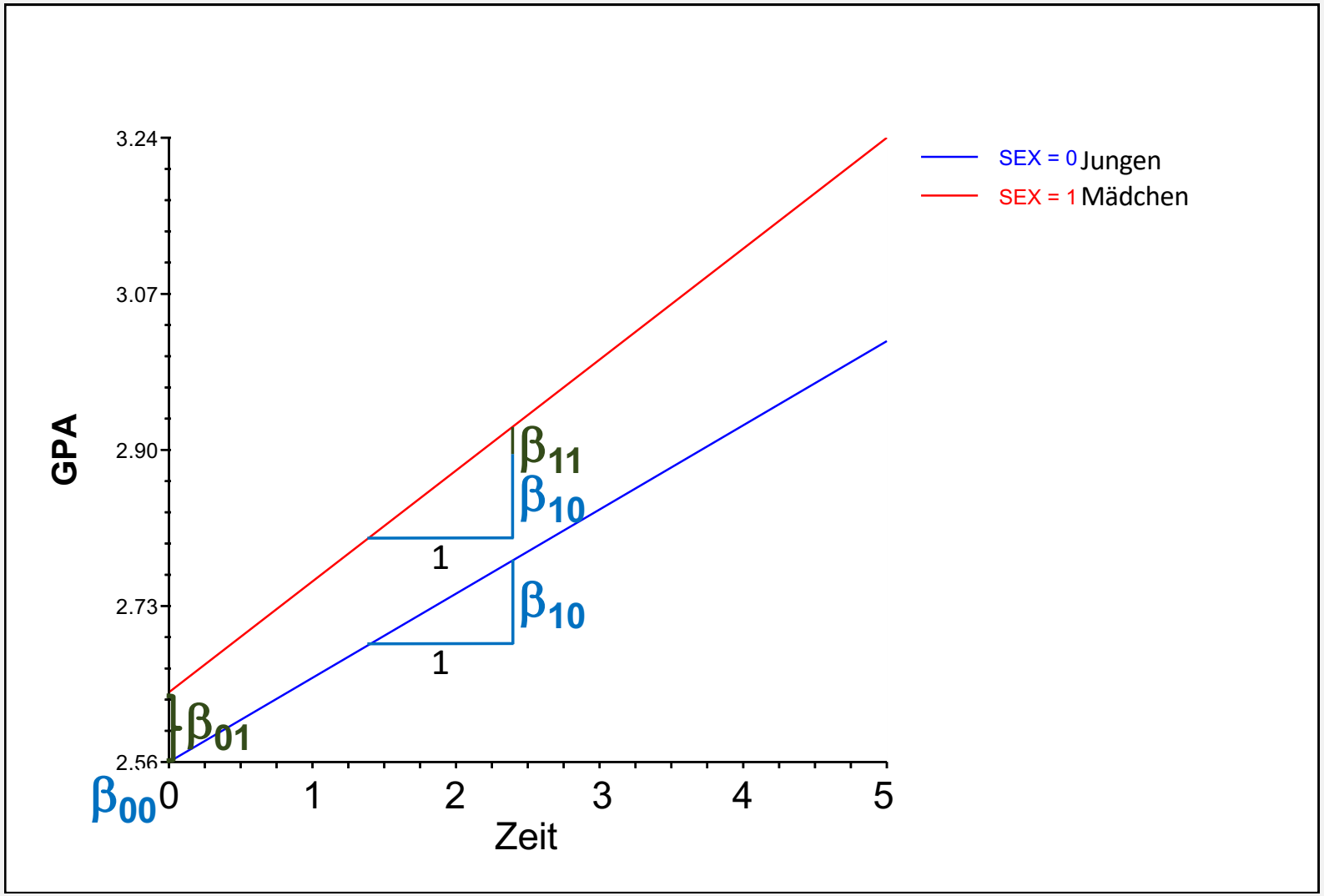
Durchschnittliche Veränderung über die Zeit bei Mädchen

$$\pi_{1i} = \overbrace{\beta_{10} + \beta_{11}}\, 1 + r_{1i}$$

Durchschnittliche Veränderung über die Zeit bei Jungen

Differenz von Jungen und Mädchen bzgl. der durchschnittlichen Veränderung über die Zeit

Beispiel „Grade Point Average Data“
Intercepts- and Slopes-as-Outcomes Modell



Strategien der Modellbildung

⇒ Modellbildungsstrategien:

- Top-down ⟶ komplex -> einfach
- Bottom-up ⟶ einfach -> komplex

⇒ Bei Multilevel Modeling Bevorzugung von Bottom-up Strategien:

- weniger Berechnungsprobleme
- erleichterte Interpretation

Strategien der Modellbildung
Analyseschritte nach Hox (2010)

Schritt 1

Level 1:

$$Y_{ti} = \pi_{0i} + e_{ti}$$

Level 2:

$$\pi_{0i} = \beta_{00} + r_{0i}$$

⇒ Intercept-Only Modell

⇒ Modell ohne Level 1 und Level 2 Prädiktoren

⇒ σ_{r0}^2 -> Varianz zwischen den Personen
σ_e^2 -> Varianz zwischen den Messungen innerhalb der Personen
Gesamtvarianz: VAR(Y_{ti}) = $\sigma_{r0}^2 + \sigma_e^2$

⇒ Kann für die Berechnung der Intraklassenkorrelation ρ benutzt werden:

$$\rho = \frac{\sigma_{r_0}^2}{\sigma_{r_0}^2 + \sigma_e^2} = \frac{\tau_{00}}{\underbrace{\tau_{00} + \sigma^2}_{\text{Gesamtvarianz}}}$$

Varianz zwischen den Personen

Varianz zwischen den Messungen innerhalb der Personen

Anteil der Varianz zwischen den Personen im Verhältnis zur Gesamtvarianz

Strategien der Modellbildung
Analyseschritte nach Hox (2010)

Schritt 2

Level 1:

$$Y_{ti} = \pi_{0i} + \pi_1 ZEIT_{ti} + e_{ti}$$

Level 2:

$$\pi_{0i} = \beta_{00} + r_{0i}$$

$$\pi_1 = \beta_{10}$$

⇒ Berücksichtigung von allen Level 1 Prädiktoren (hier: Prädiktorvariable *ZEIT*)

⇒ Fixierung der Slopes -> keine interindividuelle Variation in den Steigungskoeffizienten (Random Intercept, Fixed Slope Modell)

Schritt 3

Level 1:

$$Y_{ti} = \pi_{0i} + \pi_1 ZEIT_{ti} + e_{ti}$$

Level 2:

$$\pi_{0i} = \beta_{00} + \beta_{01} X_{1i} + r_{0i}$$

$$\pi_1 = \beta_{10}$$

⇒ Ggf. Berücksichtigung von allen Level 2 Prädiktoren zur Vorhersage der Intercepts (hier: X_1)

⇒ Slopes bleiben weiterhin fixiert

Strategien der Modellbildung
Analyseschritte nach Hox (2010)

Schritt 4

Level 1:

$$Y_{ti} = \pi_{0i} + \pi_{1i} ZEIT_{ti} + e_{ti}$$

Level 2:

$$\pi_{0i} = \beta_{00} + \beta_{01} X_{1i} + r_{0i}$$

$$\pi_{1i} = \beta_{10} + r_{1i}$$

⇒ Freie Schätzung der Slopes

⇒ Überprüfung, ob Varianz der Slopes σ_{r1}^2 bzw. τ_{11} signifikant ist

Schritt 5

Level 1:

$$Y_{ti} = \pi_{0i} + \pi_{1i} ZEIT_{ti} + e_{ti}$$

Level 2:

$$\pi_{0i} = \beta_{00} + \beta_{01} X_{1i} + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11} X_{1i} + r_{1i}$$

⇒ Ggf. Berücksichtigung von allen Level 2 Prädiktoren zur Vorhersage der Slopes (hier: X_1)

Literatur

Einführung

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications* (3rd ed.). Routledge.

Luke, D. A. (2019). *Multilevel Modeling* (2nd ed.). Sage.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press.

Analyse in SPSS, R und andere

Bickel, R. (2007). *Multilevel analysis for applied research: It's just regression!* Guilford Press.

Finch, W. H., Bolin, J. E., & Kelley, K. (2019). *Multilevel Modeling Using R* (2nd ed.). Chapman and Hall/CRC.

Garson, G. D. (2020). *Multilevel Modeling Applications in STATA®, IBM SPSS®, SAS®, R, & HLM™*. Sage.

Heck, R. H., & Thomas, S. L. (2020). *An introduction to multilevel modeling techniques: MLM and SEM approaches (*4th ed.). Routledge.

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2014). *Multilevel and longitudinal modeling with IBM SPSS* (2nd ed.). Routledge.