Universität Trier
Zentrum für Informations-, Medien- und Kommunikationstechnologie

Bernhard Baltes-Götz

Analyse von hierarchischen linearen Modellen mit SPSS

2020 (Rev. 200407)

Herausgeber: Zentrum für Informations-, Medien- und Kommunikationstechnologie (ZIMK)
an der Universität Trier
Universitätsring 15
D-54286 Trier
WWW: zimk.uni-trier.de
E-Mail: zimk@uni-trier.de
Tel.: (0651) 201-3417, Fax.: (0651) 3921

Autor: Bernhard Baltes-Götz (E-Mail : baltes@uni-trier.de)

Vorwort

In diesem Manuskript geht es um die Analyse von hierarchischen linearen Modellen für Cluster- und Messwiederholungsdaten mit SPSS Statistics. Es wird die Programmversion 26 verwendet, doch sollten praktisch alle genutzten Prozeduren auch in älteren Versionen verfügbar sein.

Die aktuelle Version des Manuskripts ist als PDF-Dokument zusammen mit den im Kurs benutzen Dateien auf dem Webserver der Universität Trier von der Startseite (http://www.uni-trier.de/) ausgehend folgendermaßen zu finden:

IT-Services > Downloads & Broschüren > Statistik >
Analyse von hierarchischen linearen Modellen mit SPSS

Kritik und Verbesserungsvorschläge zum Manuskript werden dankbar entgegen genommen (z. B. unter der Mail-Adresse baltes@uni-trier.de).

Trier, im April 2020 Bernhard Baltes-Götz

Inhaltsverzeichnis

1 Einleitung

Im linearen Modell wird die *Unkorreliertheit* der Residuen vorausgesetzt, die bei den nicht selten anzutreffenden *Cluster-Stichproben* in der Regel verletzt ist. Werden z. B. 1000 Schüler aus insgesamt 100 Schulen untersucht, sind bei einem linearen Modell mit den *Schülern* als Untersuchungseinheiten abhängige Residuen zu erwarten. Schüler aus derselben Schule haben viele im Modell unberücksichtigte Bedingungen mit Relevanz für das Kriterium gemeinsam, was zu ähnlichen Residuen führt.

Cluster-Stichproben resultieren aus einer *mehrstufigen Stichprobenziehung* (siehe z. B. Eid et al. 2011, S. 700). Im Beispiel wurden nicht 1000 Schüler aus der Population aller Schüler gezogen, sondern in einem zweistufigen Prozess ...

- wurden zunächst 100 Schulen zufällig aus der Population alle Schulen gezogen,
- um dann aus jeder Schule zufällig ca. 10 Schüler auszuwählen.

Für die mehrstufige Stichprobenziehung sprechen oft praktische Erwägungen (Snijders & Bosker 2012, S. 7). 1000 Schüler aus insgesamt 100 Schulen zu untersuchen, ist erheblich kostengünstiger als die Untersuchung von 1000 zufällig gezogenen Schülern, die im Extremfall 1000 verschiedene Schulen besuchen. Der Rekrutierungsaufwand pro Schüler (Fall der Mikroebene) ist geringer. Über Merkmale auf der Schul- bzw. Makroebene sind bei einer Cluster-Stichprobe allerdings erheblich weniger Informationen vorhanden als bei einer einfachen Zufallsstichprobe von Schülern.

Für die mehrstufige Stichprobenziehung spricht außerdem, dass durch Aggregation entstehende Merkmale auf der Makroebene bei einer einfachen Stichprobenziehung von Individuen nur schwer zu erheben sind, z.B.:

- die durchschnittliche Intelligenz in der Klasse eines untersuchten Schülers
- die über alle Schüler in der Klasse gemittelte Beurteilung der Lehrerqualität

Vermutlich ist in vielen Mehrebenenstudien der reale Prozess der Stichprobengewinnung nur partiell als mehrstufige Zufallsziehung zu beschreiben. Häufig sind Bequemlichkeits- statt Zufallsstichproben im Spiel, und bei Klassen als Makroeinheiten wird man in der Regel *alle* Schüler einer Klasse einbeziehen (vgl. Hox 2010, S. 1). Es genügt allerdings, wenn in jeder Gruppe j der untersten Ebene n_j Beobachtungen vorliegen, für die das angenommene lineare Mikromodell mit unabhängig identisch normalverteilten Residuen gilt (siehe unten). Eine Zufallsziehung muss auf der Mikroebene nicht stattfinden.

Werden die Daten aus einer Cluster-Stichprobe mit zwei Ebenen (z. B. Zufallsauswahl von Schülern aus zufällig gewählten Schulen) wie unabhängige Beobachtungen behandelt (z. B. per POLS-Regressionsanalyse, *Pooled Ordinary Least Squares*), dann …

- sind für *Makroregressoren* (im Beispiel: Merkmale der Schulen wie Größe oder Ausstattung) *unterschätzte* Standardfehler zu erwarten. Damit steigt die Rate der Fehler erster Art für Regressoren auf der Makroebene (im Beispiel: auf der Schulebene) an, d.h. wahre Nullhypothesen werden mit einer Wahrscheinlichkeit größer als 0,05 verworfen.
- sind für *Mikroregressoren* (im Beispiel Merkmale der Schüler wie Motivation oder Begabung) *überschätzte* Standardfehler zu erwarten. Damit steigt die Rate der Fehler zweiter Art für Regressoren auf der Mikroebene an (zu konservative Testentscheidungen). Das liegt daran, dass die mögliche Varianzaufklärung durch die Cluster-Variable nicht ausgeschöpft wird.

Arbeitet man zur Vermeidung der Abhängigkeit von Beobachtungen mit *aggregierten Daten* (im Beispiel: auf Schulebene) und überträgt Befunde (z. B. zum Effekt von Intelligenz auf das akademische Selbstkonzept) auf die Individualebene, dann kann es zum sogenannten *ökologischen Fehlschluss* kommen (Robinson 1950; Eid et al. 2011, S. 701). Bei manchen Merkmalen (z. B. Intelligenz von Schülern) ist auf der Mikroebene (mit Individualdaten) ein anderer Effekt auf die abhänge Variable (z. B. akademisches Selbstkonzept) zu erwarten als auf der Makroebene (mit aggregierten Daten). Während die eigene Intelligenz eines Schülers einen positiven (Within-)Effekt auf das akademische Selbstkonzept hat, wird für die mittlere Intelligenz der Klasse regelmäßig ein negativer (Between-)Effekt auf das akademische Selbstkonzept beobachtet (Big-Fish-Little-Pond - Effekt, siehe z. B. Marsh et al., 2008). Offenbar muss bei manchen Regressoren zwischen

dem Within- und dem Between-Effekt unterschieden werden. Bei einer Analyse mit aggregierten Daten kann ausschließlich der Between-Effekt untersucht werden, und der oft relevantere Within-Effekt darf keinesfalls mit dem Between-Effekt gleichgesetzt werden.

Wenn man eine vorhandene Cluster-Struktur ignoriert und trotz verletzter Unabhängigkeit von Individualdaten ausgeht, kommt es nicht nur zu einer erhöhten Rate fehlerhafter Testentscheidungen, sondern man erhält bei Regressoren mit Varianz auf der Mikro- und der Makroebene auch verzerrte Effektschätzer (z.B. in einer POLS-Regressionsanalyse), wenn sich der Within- und der Between-Effekt unterschieden. Man erhält eine Mischung aus beiden Effekten und kommt mit großer Wahrscheinlichkeit zu falschen Schlüssen. In Abschnitt 2 werden sowohl der ökologische Fehlschluss als auch die gesteigerte α-Fehlerrate als mögliche Gefahren bei einer falschen Analyse von Cluster-Daten vorgeführt.

Hierarchische lineare Modelle (Mehrebenenmodelle) erlauben die methodisch korrekte Analyse von Cluster-Stichproben (mit unverzerrten Effektschätzungen und unverfälschter Inferenzstatistik). Das aufzuklärende Kriterium ist dabei auf der Mikroebene angesiedelt (z. B. Leistung oder akademisches Selbstkonzept von Schülern). Als Regressoren können Variablen auf der Mikro- *und* der Makroebene einbezogen werden (z. B. Motivation der Schüler, Ausstattung der Klassen). Ein hierarchisches lineares Modell (ab jetzt als *HLM* bezeichnet) geht von einer zweistufigen Stichprobenziehung aus und berücksichtigt die daraus resultierenden Abhängigkeiten von Beobachtungen auf der Mikroebene. Durch eine korrekte Modellspezifikation, d.h. durch die Berücksichtigung der benötigten Regressoren auf der Mikro- und der Makroebene kann ein ökologischer Fehlschluss vermieden werden.

Einige Anwendungsbeispiele mit Merkmalen auf der Makro- und der Mikroebene:

- Arbeitsgruppen und Mitarbeiter
  - Makroebene: Führungsstil
  - Mikroebene: Leistungsmotivation der Mitarbeiter
  - Kriterium: Produktivität der Mitarbeiter
- Kliniken und Patienten
  - Makroebene: Erfahrung und Personalausstattung der Kliniken
  - Mikroebene: Alter, Geschlecht, Kooperationsbereitschaft der Patienten
  - Kriterium: Behandlungserfolg der Patienten
- Richter und Angeklagte
  - Makroebene: Alter, Geschlecht und Arbeitsbelastung der Richter
  - Mikroebene: Alter, Geschlecht, Bildung, sozioökonomischer Status der Angeklagten
  - Kriterium: Härte des Urteils (z. B. Prozent der Höchststrafe)

Wechselwirkungen können in einem HLM sowohl innerhalb der Mikro- bzw. Makroebene als auch im Sinne einer Moderation von Effekten der Mikroebene durch Regressoren der Makroebene (Cross-Level - Interaktion) einbezogen werden.

Hierarchische lineare Modelle sind nicht auf *zwei* Ebenen begrenzt. Werden z. B. bei einer Untersuchung von Schülern aus verschiedenen Schulen mehrere Länder einbezogen, dann sind drei Ebenen beteiligt:

- Schüler
- Schulen
- Länder

In diesem Manuskript werden wir uns aber auf zwei Ebenen beschränken.

Hierarchische lineare Modelle bewähren sich nicht nur bei Cluster-Stichproben, sondern auch bei **Längsschnittdaten**, wobei sich z. B. die von einer Person zu mehreren Zeitpunkten abgelieferten Messungen auf der Mikroebene befinden und eine Person eine Makroeinheit darstellt. Im Vergleich zu traditionellen Auswertungsmethoden für Längsschnittdaten bietet die Mehrebenenmodelle eine höhere Flexibilität:

- Die Messzeitpunkte der Personen dürfen sich nach Anzahl zeitlicher Anordnung unterscheiden.
- Fehlende Werte stellen kein Problem dar, sofern mindestens die MAR-Bedingung (*Missing At Random*, siehe z. B. Baltes-Götz 2013) eingehalten wird.
- Es können zeitabhängige Kovariaten einbezogen werden.

Raudenbush & Bryk 2002, S. 5f listen folgende Alternativen für die von ihnen bevorzugte Bezeichnung *Hierarchisches Lineares Modell* (HLM) auf:

- Soziologie: *multilevel linear models* (z. B. Goldstein 2003)
- Biometrie: *mixed effects models*, *random effects models* (z. B. Singer 1998)
- Ökonometrie: *random coefficient regression models*
- Statistik: *covariance components models*

Seit den 90er Jahren des letzten Jahrhunderts wurden spezielle Programme für den neuen Analyseansatz entwickelt, wobei vor allem zu erwähnen sind:

- **HLM** (Raudenbush et al. 2011)
- **MLwiN** (Rasbash et al. 2019)

Mittlerweile haben die Hersteller der statistischen Universalprogramme (z. B. **Mplus**, **SAS**, **SPSS**, **Stata**) nachgezogen, und auch das freie Statistikentwicklungssystem **R** unterstützt Mehrebenenmodelle durch mehrere Pakete (z. B. **nlme**, **lme4**). SPSS bietet in den Prozeduren MIXED, die im vorliegenden Manuskript hauptsächlich benutzt wird, und GENLINMIXED flexible Analysemöglichkeiten für Mehrebenenmodelle:

- MIXED erlaubt die Analyse von *metrischen* Kriterien mit *normalverteilten* Residuen.
- GENLINMIXED erlaubt zusätzlich eine Mehrebenenanalyse von *kategorialen* Kriterien und bietet außerdem eine robuste Schätzung der Standardfehler zu Regressionskoeffizienten trotz verletzter Homoskedastizität. Wir werden GENLINMIXED im Kurs gelegentlich ergänzend einsetzen. Eine ausführliche Behandlung der Prozedur findet sich z. B. in Heck, Thomas & Tabata (2012).

2 Analyse von Cluster-Daten durch traditionelle lineare Modelle

In diesem Abschnitt werden drei Optionen zur Analyse von Cluster-Daten durch traditionelle lineare Modelle (mit einer Ebene) beschrieben. Die beiden ersten Verfahren sind fehlerhaft und damit zu vermeiden. Das dritte Verfahren ist unter bestimmten Umständen eine sinnvolle Option.

2.1 Fehlerhafte Inferenzstatistik zu Makroregressoren bei ignorierter Abhängigkeit

Wird bei der Analyse von Cluster-Daten eine bestehende Abhängigkeit der Residuen ignoriert, ist in Signifikanztests über Regressionskoeffizienten von Makroregressoren mit einer gesteigerten α-Fehlerrate zu rechnen. Weil in den Daten über die Regressionskoeffizienten zu den Makroregressoren weniger Information vorhanden sind als die Stichprobengröße vorgaukelt, werden die Standardfehler dieser Regressionskoeffizienten unterschätzt.

Das Problem soll mit Hilfe simulierter Daten demonstriert werden:[1]

- Aus einer Population werden 100 Gruppen (z. B. Schulen) zufällig gewählt, und aus jeder Gruppe werden 10 Fälle zufällig gezogen.
- Bei jedem Fall werden das Kriterium *Y* und der Mikroregressor *X* beobachtet. Außerdem ist auf Gruppenebene der Makroregressor *W* vorhanden. Beide Regressoren haben *keinen* Effekt.
- Die Gruppen erhalten einen Zufallseffekt auf ihren *Y*-Mittelwert (U_{0j}), den man als *random intercept* bezeichnet.
- Die Zufallseffekte folgen einer Normalverteilung und sind untereinander sowie mit den ebenfalls normalverteilten Residuen der Mikroebene (R_{ij}) unkorreliert.

Für die Beobachtungen Y_{ij} gilt folgendes Modell:

$$Y_{ij} = U_{0j} + R_{ij},\ i = 1, \ldots, n_j \text{ und } j = 1, \ldots, J$$

$$U_{0j} \sim \text{unabhängig } N(0, \tau_{00}),\ R_{ij} \sim \text{unabhängig } N(0, \sigma^2)$$

In einer Simulationsstudie wurde die beschriebene zweistufige Stichprobenziehung 1000-mal durchgeführt. Die Daten wurden jeweils mit der SPSS-Prozedur für Regressionsanalysen mit unkorrelierten Residuen (POLS-Regression, *Pooled Ordinary Least Squares*) sowie mit der in diesem Manuskript vorzustellenden Prozedur MIXED für Mehrebenenmodelle analysiert. Dabei ergaben sich die folgenden geschätzten α-Fehlerraten (korrekter Wert: 0,05):

	α-Fehlerrate für den Mikroregressor *X*	α-Fehlerrate für den Makroregressor *W*
Lineares Modell (ignorierte Abhängigkeit)	0,040	0,457
Hierarchisches lineares Modell	0,057	0,047

Bei der Beurteilung des Makroregressors per POLS-Regressionsanalyse steigt die α-Fehlerrate in einem inakzeptablen Maß an, während beim Mikroregressor das Testniveau von 0,05 nicht ganz ausgeschöpft wird. Die Prozedur MIXED liefert hingegen eine verlässliche Inferenzstatistik.

Das Ausmaß der Verfälschung der Inferenzstatistik durch ignorierte Gruppeneffekte hängt von der in Abschnitt 3.2.2 zu behandelnden Intraklassenkorrelation ab, für die in pädagogischen Studien Werte zwischen 0,10 und 0,25 beobachtet werden (Snijders & Bosker 2012, S. 50). In der Simulationsstudie wurde ein Wert von 0,2 realisiert.

Weil in der Simulationsstudie der Regressor mit Varianz auf der Mikro- und der Makroebene auf beiden Ebenen denselben Effekt null hat, kann es nicht zu einer verzerrten Effektschätzung kommen.

[1] An der im Vorwort vereinbaren Stelle finden Sie im Ordner **Alpha-Inflation** das SPSS-Programm **AlphaInflation.sps**, das die Simulationsdaten erzeugt und analysiert.

2.2 Aggregation und ökologischer Fehlschluss

Um das Risiko einer erhöhten α-Fehlerrate zu vermeiden, werden Cluster-Daten oft aggregiert, so dass die Gruppen zu den Beobachtungseinheiten werden (z. B. Schulen). Dieses Vorgehen ist akzeptabel, wenn nur Makroregressoren und Between-Effekte von aggregierten Mikroregressoren interessieren. Effekte auf der Mikroebene können aber nicht untersucht werden. Bei der Übertragung von Befunden der Makroebene auf die Mikroebene kann es zum ökologischen Fehlschluss kommen, was nun demonstriert werden soll.

Auch dieses Problem soll mit Hilfe simulierter Daten demonstriert werden, wobei der oben erwähnte Big-Fish-Little-Pond - Effekt konstruiert wird:[1]

- Es wird eine Clusterstichprobe gezogen mit 30 Schulklassen (Makroeinheiten) und jeweils 30 Schülern (Mikroeinheiten).
- Als Kriterium *Y* wird das akademische Selbstkonzept der Schüler untersucht. Als Regressor *X* wird die Intelligenz der Schüler einbezogen.
- Während in der Kunstwelt die eigene Intelligenz eines Schülers einen positiven (Within-)Effekt auf sein akademische Selbstkonzept hat, besteht für die mittlere Intelligenz der Klasse ein negativer (Between-)Effekt auf das akademische Selbstkonzept. Diese Konstellation ist auch in vielen realen Stichproben beobachtet worden (siehe z. B. Marsh et al., 2008).

Wir definieren ....

- einen Makroregressor mit den Klassenmittelwerten: $\frac{1}{n_j}\sum_{i=1}^{n_j} X_{ij} =: \bar{X}_j$
- einen Mikroregressor mit den klassenintern zentrierten Werten: $X_{ij} - \bar{X}_j$

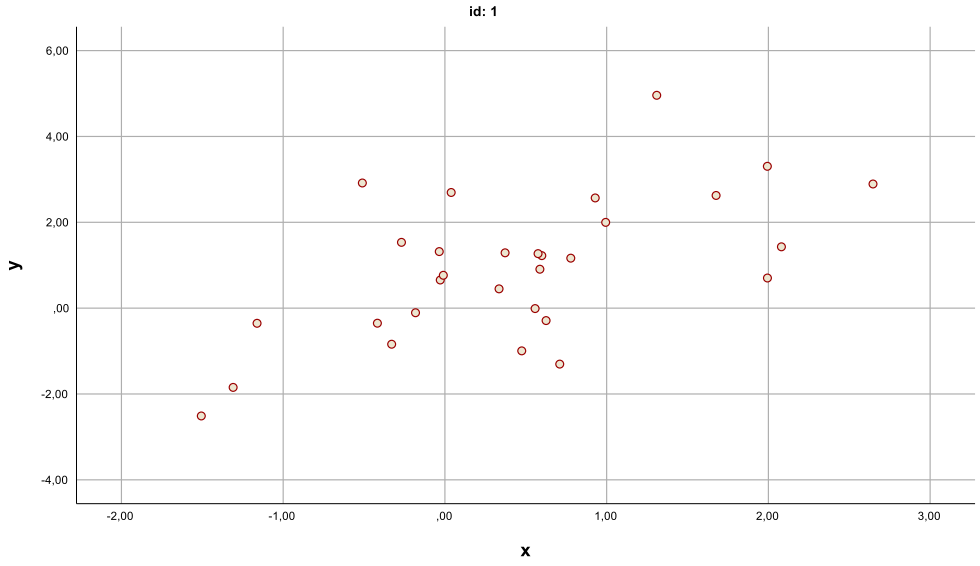
Beim Makroregressor wird das Globalmittel auf 0 gesetzt.

Für den Kriteriumswert Y_{ij} zum Schüler *i* in der Klasse *j* gilt folgendes Modell:

$$Y_{ij} = -1 \cdot \bar{X}_j + U_{0j} + 1 \cdot (X_{ij} - \bar{X}_j) + R_{ij},\ i = 1, \ldots, n_j \text{ und } j = 1, \ldots, J$$

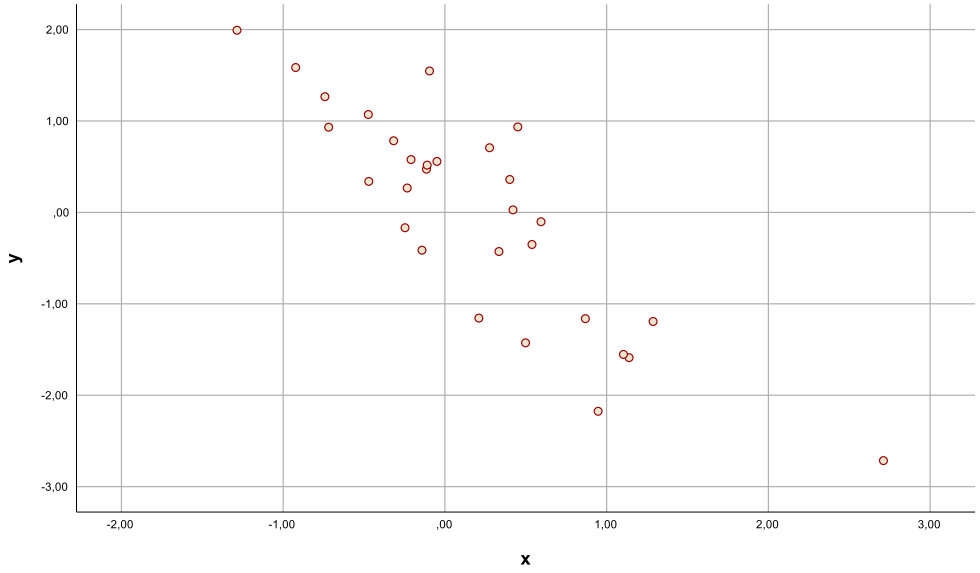
$$U_{0j} \sim \text{unabhängig } N(0, \tau_{00}),\ R_{ij} \sim \text{unabhängig } N(0, \sigma^2)$$

In den Klassen zeigt sich der positive Within-Effekt, z.B. für Klasse 1:



Ein Streudiagramm für die aggregierten Daten mit den 30 Klassen zeigt den negativen Between-Effekt:

[1] An der im Vorwort vereinbaren Stelle finden Sie im Ordner **BFLP** SPSS-Programm **BFLP.sps**, das die Simulationsdaten erzeugt und analysiert.



In einer linearen Regression von *Y* auf *X* mit den aggregierten Daten wird der negative Between-Effekt relativ präzise geschätzt und als signifikant beurteilt:

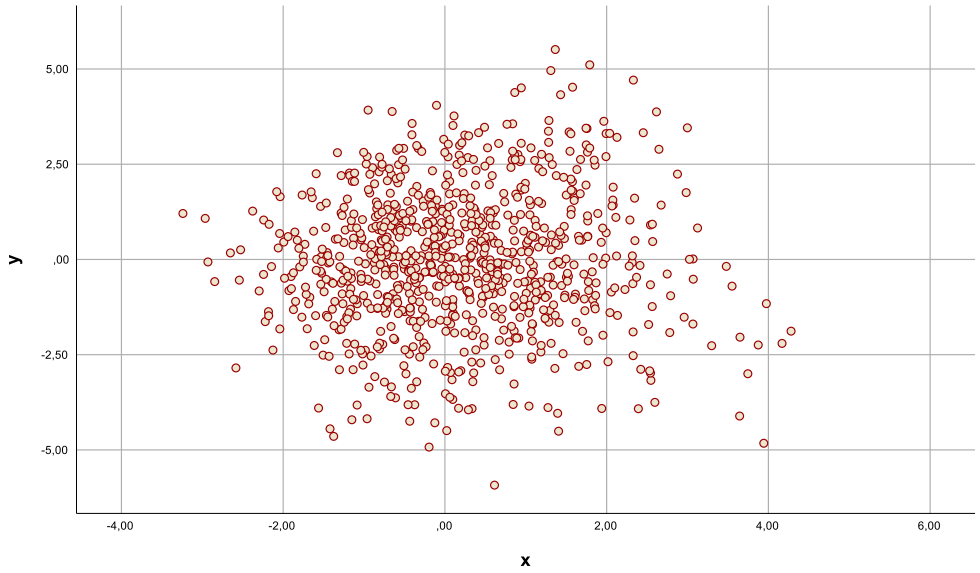
Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			95,0% Konfidenzintervalle für B	
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Untergrenze	Obergrenze
1	(Konstante)	,222	,116		1,908	,067	-,016	,459
	x	-1,264	,145	-,854	-8,701	,000	-1,562	-,966

a. Abhängige Variable: y

Solche Ergebnisse aus aggregierten Daten darf man offenbar nicht auf die Mikroebene übertragen.

Das Risiko der ökologischen Fehlinterpretation ist keinesfalls an das Aggregieren gekoppelt, sondern droht auch bei einer Analyse von Individualdaten mit Cluster-Struktur per POLS-Regressionsanalyse (*Pooled Ordinary Least Squares*). Im Beispiel erhält man das folgende Streudiagramm



und die folgende Ergebnistabelle zur (unzulässigen) POLS-Regression:

Koeffizienten [a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			95,0% Konfidenzintervalle für B	
		Regressions-koeffizient B	Std.-Fehler	Beta	T	Sig.	Untergrenze	Obergrenze
1	(Konstante)	-,028	,060		-,467	,640	-,146	,090
	x	,061	,048	,042	1,260	,208	-,034	,155

a. Abhängige Variable: y

Es resultiert eine wenig aufschlussreiche Mischung aus dem Within- und dem Between-Effekt.

2.3 Modelle mit fixierten Effekten

Unterschiedliche Kriteriumsmittelwerte der untersuchten Gruppen lassen sich in einem linearen Modell berücksichtigen, indem die Gruppenzugehörigkeitsvariable als Faktor mit fixierten Effekten einbezogen wird. Bei J Gruppen sind dazu $(J - 1)$ Kodiervariablen erforderlich.

Mit dem Test auf signifikante Gruppenunterschiede wird übrigens die in Abschnitt 3.2.2 zu behandelnde Intraklassenkorrelation auf Signifikanz beurteilt. Somit kann das lineare Modell mit Gruppenfaktor einen Beitrag zur Entscheidung über die Notwendigkeit eines Mehrebenenmodells leisten.

Snijders & Bosker (2012, S. 46f) nennen folgende Gesichtspunkte zur Entscheidung zwischen einem linearen Modell mit fixierten Gruppeneffekten und einem Mehrebenenmodell mit zufälligem Gruppeneffekten:

- Sind die beteiligten Gruppen von eigenständigem Interesse (z. B. Berufsgruppen, Parteizugehörigkeit), dann ist ein Modell mit fixierten Effekten naheliegend.
- Wurden die Gruppen zufällig aus einer Population gezogen (z. B. Schulen, Arbeitsgruppen), auf die man generalisieren möchte, dann ist ein Modell mit zufälligen Effekten angemessen.
- In einem Modell mit fixierten Gruppeneffekten lässt sich die Wirkung von Gruppenmerkmalen *nicht* untersuchen, weil alle Mittelwertsunterschiede zwischen den Gruppen bereits vom Gruppenfaktor aufgeklärt werden. In einer Studie mit Schulen als Makroeinheiten könnte man z. B. *nicht* untersuchen, wie der Makroregressor *Schultyp* (mit den Werten *Öffentlich* und *Privat*) die mittlere Mathematikleistung der Schulen oder die schulinterne Regression von Matheleistung auf den sozioökonomischen Status beeinflusst.
- Bei relativ schwach besetzten Gruppen (mit 2 bis 50 Fällen pro Gruppe) sind Modelle mit zufälligen Effekten überlegen, weil die Schätzung von gruppenspezifischen Parametern aufgrund einer schwachen empirischen Basis vermieden wird. Allerdings muss die meist enthaltene Annahme der zufälligen Ziehung aus einer Population mit normalverteilten Effekten gerechtfertigt sein (siehe Abschnitt 3). Ist diese Annahme erheblich verletzt, drohen unzuverlässige Ergebnisse.
- Wenn ausschließlich die Within-Effekte von Mikroregressoren interessieren, und die Gruppen angemessen stark besetzt sind ($n_j > 50$), dann ist das Modell mit fixierten Gruppeneffekten im Vorteil, weil es im Unterschied zum Modell mit zufälligen Gruppeneffekten weniger Annahmen benötigt:
  - Mögliche Korrelationen zwischen den Gruppeneffekten und den Mikroregressoren bewirken keine verzerrten Schätzungen der Regressionskoeffizienten. Dies ist aus kausalitätstheoretischer Sicht ein gravierender Vorteil, der sich allerdings auch in Mehrebenenmodellen realisieren lässt (siehe Beispiel in Abschnitt 2.2).
  - Die Verteilung der Gruppeneffekte ist irrelevant.

  Die Freiheit von kritischen Annahmen haben insbesondere in der ökonometrischen und soziologischen Methodenliteratur zu einer ausgeprägten Bevorzugung des Fixed-Effects - Modells geführt (siehe z. B. Allison 2009; Brüderl 2010; Wooldridge 2013).

Insgesamt geben Snijders & Bosker (2012, S. 48) für sozialwissenschaftliche Forschungsprojekte die folgenden Empfehlungen:

- Bei weniger als 10 Gruppen eignet sich meist ein Modell mit festen Effekten.
- Sind 20 oder mehr, relativ schwach besetzte Gruppen vorhanden (mit weniger als 100 Fällen), dann ist ein Modell mit zufälligen Effekten zu bevorzugen.
- Haben alle Gruppen mehr als 100 Fälle, kommen beide Ansätze in Frage, und eine Entscheidung muss anhand anderer Kriterien erfolgen.

Wenn keine Makroregressoren interessieren und nicht allzu viele Gruppen vorliegen, ist das Fixed-Effects - Modell zu bevorzugen, weil man es mit jeder Software für das lineare Modell analysieren und den deutlich höheren Aufwand für ein Mehrebenenmodell vermeiden kann. Wie sich in Abschnitt 3.6.5 zeigen wird, ist z. B. der Zeitaufwand für eine Fixed-Effects - Analyse mit 160 Gruppen auf einem modernen PC kein Problem.

3 Zweiebenenmodelle für Cluster-Daten

Wir betrachten in diesem Abschnitt Modelle für Daten zu Untersuchungseinheiten auf zwei Ebenen einer hierarchischen Struktur (z. B. Beschäftigte in Arbeitsgruppen, Schüler in Klassen). Zur Erklärung des Verhaltens der Mikroebenenfälle sollen Regressoren aus beiden Ebenen verwendet werden.

Die allgemeine Formulierung des Zweiebenenmodells nach Raudenbush & Bryk (2002, S. 56f) enthält auf der **Mikroebene** ein lineares Modell mit Q (Individual-)Regressoren und einem unabhängig identisch normalverteilten Residuum, wobei für jede Gruppe (mit Index j = 1, .., J) eigenständige Regressionskoeffizienten erlaubt sind:

$$\begin{aligned} Y_{ij} &= \beta_{0j} + \beta_{1j} X_{1ij} + \beta_{2j} X_{2ij} + \ldots + \beta_{Qj} X_{Qij} + R_{ij} \\ &= \beta_{0j} + \sum_{q=1}^{Q} \beta_{qj} X_{qij} + R_{ij} \end{aligned}$$

$$\text{für } i = 1, .., n_j \text{ und } j = 1, .., J; \text{ mit } R_{ij} \sim \text{unabhängig } N(0, \sigma^2)$$

Für die (Q + 1) Regressionskoeffizienten β_{qj} in den zufällig gewählten Gruppen wird jeweils das folgende **Makromodell** mit (Gruppen-)Regressoren W_s (s = 1, .., S) angesetzt:

$$\begin{aligned} \beta_{qj} &= \gamma_{q0} + \gamma_{q1} W_{1j} + \gamma_{q2} W_{2j} + \ldots + \gamma_{qS} W_{Sj} + U_{qj} \\ &= \gamma_{q0} + \sum_{s=1}^{S} \gamma_{qs} W_{sj} + U_{qj} \end{aligned}$$

Für jeden zu erklärenden Erstebenenparameter β_{qj} kann eine Teilmenge der γ_{qs} - Parameter auf null gesetzt werden, wenn für die zugehörigen Regressoren kein Einfluss anzunehmen ist.

Für die Residuen der Makroebene U_{qj} (q = 0, .., Q) wird eine multivariate Normalverteilung angenommen mit Varianzen

$$\text{Var}(U_{qj}) = \tau_{qq}$$

und Kovarianzen

$$\text{Cov}(U_{qj}, U_{q'j}) = \tau_{qq'}$$

Je nach Modell kann für die Kovarianzmatrix eine spezielle Struktur angenommen werden (z.B. Unkorreliertheit der Makroresiduen zu verschiedenen Erstebenparametern).

Die (Q+1) - dimensionale Zufallsvariable $(U_{0j}, U_{1j}, .., U_{Qj})$ zur Gruppe j ist von den Makroresiduen aus anderen Gruppen ebenso unabhängig wie von den Mikroresiduen R_{ij}.

Wir werden im Verlauf von Abschnitt 3 das allgemeine Zweiebenenmodell sukzessive aus einfacheren Partialmodellen entwickeln.

3.1 Anwendungsbeispiel

Als empirisches Anwendungsbeispiel dient ein Datensatz aus der amerikanischen *High School and Beyond* - Studie (Abkürzung: *HS&B - Studie*), der z.B. auch in Norušis (2008), Raudenbush & Bryk (2002) und Singer (1998) zur Demonstration von diversen HLM-Varianten verwendet wird. Es sind Informationen zu 7185 Schülern aus 160 Schulen vorhanden, wobei die einzelnen Schulen mit 14 bis 67 Schülern vertreten sind (mittlere Gruppenstärke: 44,91; Standardabweichung: 11,86).

Auf der Schüler- bzw. Mikroebene werden folgende Variablen betrachtet:

- MACH
  Als Kriterium wird die Mathematikleistung der Schüler anlysiert.
- CSES
  Der sozioökonomische Status der Schüler wird als Regressor auf der Mikroebene verwendet. Diese Variable ist *schulbezogen zentriert*, so dass die Werte innerhalb jeder Schule den Mittelwert null besitzen (vgl. Abschnitt 3.4.1).

Auf der Makroebene werden folgende Regressoren eingesetzt:

- MEANSES
  Der mittlere sozioökonomische Status einer Schule. Diese Variable ist global zentriert, so dass die 160 Schulen den MEANSES-Mittelwert null besitzen.
- SECTOR
  Hier werden öffentliche (Kodierung = 0) und katholische (Kodierung = 1) Schulen unterschieden.

Um die Beispieldaten aus dem **R**-Paket **mlmRev** in die SPSS-Datendatei **hsb82.sav** zu befördern, sind folgende Arbeitsschritte erforderlich:

- Es muss eine R-Installation zur Verfügung stehen. Starten Sie z. B. die R-Konsole, öffnen Sie mit dem Menübefehl

  Datei > Neues Skript

  ein Skriptfenster.
- Kopieren Sie die folgenden R-Kommandos in das Skriptfenster:

```
setwd("U:\\Eigene Dateien\\SPSS\\HLM\\HS&B")
install.packages("mlmRev")
library(mlmRev)
library(foreign)
data(Hsb82)
write.foreign(Hsb82, datafile="Hsb82.dat", codefile="Hsb82.sps", package="SPSS")
```

- Tragen Sie als Argument zur R-Funktion **setwd()** den gewünschten Ordner für die entstehenden Dateien als Arbeitsverzeichnis ein, wobei der Rückwärtsschrägstrich zur Trennung der Pfadsegmente verdoppelt werden muss. Das eingestellte Arbeitsverzeichnis muss existieren.
- Lassen Sie die Kommandos ausführen, indem Sie z. B. den folgenden Menübefehl benutzen:

  Bearbeiten > Alles ausführen

  Im eingestellten Arbeitsverzeichnis entstehen die Textdatendatei **Hsb82.dat** und die SPSS-Syntaxdatei **Hsb82.sps**.
- Öffnen Sie die Syntaxdatei **Hsb82.sps** in SPSS.
- Setzen Sie an den Anfang des Programms ein **CD**-Kommando, das den Ordner mit der Datei **Hsb82.dat** zum aktuellen Verzeichnis macht, z.B.:

```
cd "U:\Eigene Dateien\SPSS\HLM\HS&B".
```

- Lassen Sie ein Kommando folgen, das den Punkt zum Dezimaltrennzeichen erklärt:

```
SET DECIMAL=DOT.
```

- Setzen Sie das folgende **RECODE**-Kommando hinter das **DATA LIST** - Kommando, um die oben angegebene SECTOR-Kodierung (0 für die öffentlichen und 1 für die katholischen Schulen) zu erreichen:

```
recode sector (1=0) (2=1).
```

- Die neue SECTOR-Kodierung muss im **VALUE LABELS** - Kommando berücksichtigt werden:

```
value labels sector 0 "Public" 1 "Catholic".
```

- Die Daten nach der Schulnummer zu sortieren, erleichtert die Orientierung:

```
SORT CASES by school.
```

- Mit den folgenden Kommandos werden Variablenattribute gesetzt, was (wie die Sortierung) keinen Einfluss auf die Ergebnisse hat:

```
FORMATS school minrty sx sector (f8.0).
VARIABLE LEVEL school minrty sx sector(nominal)
 /ses mAch meanses cses (scale).
VARIABLE WIDTH school to cses (8).
```

- Speichern Sie das entstehende SPSS-Datenblatt in die SPSS-Datendatei **hsb82.sav**, was über das folgende **SAVE**-Kommando geschehen kann:

```
SAVE OUTFILE='hsb82.sav'
  /COMPRESSED.
```

- Lassen Sie alle Kommandos ausführen, indem Sie z. B. den folgenden Menübefehl benutzen:

Ausführen > Alle

Bei Verwendung des Ordners **U:\Eigene Dateien\SPSS\HLM\HS&B** erledigt das folgende SPSS-Programm die Erstellung der SPSS-Datendatei **hsb82.sav**:[1]

```
CD "U:\Eigene Dateien\SPSS\HLM\HS&B".
SET DECIMAL=DOT.
DATA LIST  FILE= "Hsb82.dat"  free (",")
 /school minrty sx ses mAch meanses sector cses.
recode sector (1=0) (2=1).

VARIABLE LABELS
 school "school"
 minrty "minrty"
 sx "sx"
 ses "ses"
 mAch "mAch"
 meanses "meanses"
 sector "sector"
 cses "cses".

VALUE LABELS
 /school
 1 "8367"
 2 "8854"
 3 "4458"
  . . .
160 "9586"
 /minrty  1 "No" 2 "Yes"
 /sx  1 "Male" 2 "Female"
 /sector  0 "Public" 1 "Catholic".

SORT CASES by school.
FORMATS school minrty sx sector (f8.0).
VARIABLE LEVEL school minrty sx sector(nominal)
 /ses mAch meanses cses (scale).
VARIABLE WIDTH school to cses (8).

SAVE OUTFILE='hsb82.sav'
  /COMPRESSED.
```

In den folgenden Abschnitten werden in Anlehnung an Raudenbush & Bryk (2002, Kap. 4) mehrere Modelle zu den HS&B - Daten vorgestellt, die sich von einem einfachen Basismodell hin zu einem komplexen Endprodukt entwickeln (Bottom-Up - Modellierungsstrategie, vgl. Abschnitt 4.4).

[1] Das Programm ist in der Datei **Hsb82.sps** an der im Vorwort vereinbarten Stelle im Ordner **HS&B** zu finden.

3.2 Basismodell (Varianzanalyse mit Zufallsfaktor)

Wir beschäftigen uns zunächst mit der Frage, wie stark sich die Schulen in der Mathematikleistung unterscheiden, d.h. welcher Anteil der Gesamtvariation auf die Unterschiede zwischen den Schulen zurückzuführen ist. Weil die Schulen zufällig in die Studie einbezogen worden sind, sollte man Schule als Zufallsfaktor behandeln, damit Schlüsse auf die Gesamtheit aller Schulen möglich sind. Nach den Kriterien von Snijders & Bosker (2012, S. 48) ist die Behandlung als Zufallsfaktor deutlich zu bevorzugen (siehe Abschnitt 2.3).

3.2.1 Modell

Auf der Mikroebene verwenden wir folgendes Modell, wobei β_{oj} als Mittelwert der Schule *j* interpretiert werden kann (vgl. Raudenbush & Bryk, 2002, S. 69f):

$$Y_{ij} = \beta_{0j} + R_{ij} \quad \text{für } i = 1, .., n_j \text{ und } j = 1, .., J$$

Für die Residuen R_{ij} wird angenommen, dass sie unabhängig identisch normalverteilt sind mit dem Erwartungswert null und der Varianz σ^2

$$R_{ij} \sim \text{unabhängig } N(0, \sigma^2)$$

Im Modell der Makroebene werden die Schulmittelwerte β_{oj} erklärt durch das Globalmittel γ_{00} und einen zufälligen Gruppeneffekt U_{0j}:

$$\beta_{0j} = \gamma_{00} + U_{0j} \quad \text{für } j = 1, .., J$$

Bei γ_{00} handelt es sich um die mittlere Mathematikleistung in der Population aller Schulen. Für die Makroresiduen U_{0j} wird angenommen, dass sie unabhängig identisch normalverteilt sind mit dem Erwartungswert null und der Varianz τ_{00}

$$U_{0j} \sim \text{unabhängig } N(0, \tau_{00})$$

Mit einer verletzten Unabhängigkeit der Makroresiduen U_{0j} ist z. B. dann zu rechnen, wenn eine dritte Ebene der Stichprobengewinnung (z. B. eine Gruppierung der Schulen nach Ländern) im Modell unberücksichtigt bleibt.

Durch Einsetzen des Makroebenenmodells für β_{oj} in das Mikroebenenmodell erhalten wir das kombinierte Modell mit dem festen Effekt γ_{00}, dem zufälligen Effekt U_{0j} und dem Mikroebenenresiduum R_{ij}:[1]

$$Y_{ij} = \gamma_{00} + U_{0j} + R_{ij}$$

Es wird angenommen, dass die Makroresiduen U_{0j} von den Mikroresiduen R_{ij} unabhängig sind.

Raudenbush & Bryk (2002, S. 24) bezeichnen das vorgestellte Modell auch als *fully unconditional*, weil es keine Regressoren auf der Mikro- oder Makroebene enthält. Snijders & Bosker (2012, S. 49) sprechen vom *empty model*. Es dient ...

- als Basismodell zur vergleichenden Beurteilung anderer Modelle (z. B. bzgl. der Erklärungsleistung)
- und zur Schätzung der Intraklassenkorrelation, die den Anteil der Schulen an der Gesamtvariation der Mathematikleistungen quantifiziert.

Wenn ein Mehrebenenmodell über den zufälligen Effekt U_{0j} für jede Gruppe *j* einen eigenen Mittelwert bzw. Ordinatenabschnitt erlaubt, aber für alle Mikroregressoren einen fixierten (gruppenunabhängigen)

---

[1] Weil ein fester und ein zufälliger Effekt vorhanden sind, spricht man von einem *gemischten Modell*. Genau genommen ist auch das Mikroebenenresiduum ein zufälliger Effekt. Es wird jedoch nicht durchgehend so bezeichnet.

Steigungskoeffizienten vorsieht, bezeichnet man es als *random intercept* - Modell. Das aktuell betrachtete Basismodell enthält keine Regressoren, und ist somit das einfachste mögliche random intercept - Modell.

Weil das random intercept - Modell für die gruppenspezifischen Mittelwerte bzw. Ordinatenabschnitte eine Verteilung unterstellt,

$$U_{0j} \sim \text{unabhängig } N(0, \tau_{00})$$

gehört es zu den *random coefficient* - Modellen. Diese stehen bei der Analyse von Cluster-Daten in Konkurrenz zu den Modellen mit fixierten Cluster-Effekten (vgl. Abschnitt 2.3).

3.2.2 Intraklassenkorrelation und Kovarianzmatrix der Beobachtungen

Aufgrund der zufälligen Ordinatenabschnitte sind folgende **Varianzkomponenten** im Basismodell enthalten:

- τ_{00} (Variabilität der Schulmittelwerte)
- σ^2 (Variabilität innerhalb der Schulen)

Für die Varianz von Y_{ij} gilt:

$$\begin{aligned} \text{Var}(Y_{ij}) &= \text{Var}(\gamma_{00} + U_{0j} + R_{ij}) \\ &= \text{Var}(U_{0j}) + \text{Var}(R_{ij}) \\ &= \tau_{00} + \sigma^2 \end{aligned}$$

Zur Beurteilung Frage nach dem Beitrag des Schulfaktors zur Variabilität der Mathematikleistungen eignet sich die folgendermaßen definierte **Intraklassenkorrelation**:

$$\rho := \frac{\tau_{00}}{\tau_{00} + \sigma^2}$$

Heck et al. (2010, S. 6 und S. 74) vertreten die Ansicht, dass bei einer Intraklassenkorrelation unterhalb von 0,05 kein zwingender Grund für eine Mehrebenenanalyse bestünde. Andererseits kann es auch bei einer Intraklassenkorrelation kleiner als 0,05 zu einer Steigerung des α-Fehlerniveaus kommen (Tabachnick & Fidell 2007, S. 822). Typische Werte aus der pädagogischen Forschung liegen zwischen 0,10 und 0,25 (Snijders & Bosker 2012, S. 50).

Für die kombinierten Residuen $(U_{0j} + R_{ij})$ und damit auch für die Beobachtungen Y_{ij} ergibt sich die folgende Kovarianzmatrix mit **zusammengesetzt-symmetrischer** Struktur:

$$\begin{pmatrix}
\tau_{00}+\sigma^2 & \tau_{00} & . & \tau_{00} & 0 & . & 0 & 0 & 0 & . & 0 \\
\tau_{00} & \tau_{00}+\sigma^2 & . & \tau_{00} & 0 & . & 0 & 0 & 0 & . & 0 \\
. & . & . & . & . & . & . & . & . & . & . \\
\tau_{00} & \tau_{00} & . & \tau_{00}+\sigma^2 & 0 & . & 0 & 0 & 0 & . & 0 \\
0 & 0 & . & 0 & \tau_{00}+\sigma^2 & . & 0 & 0 & 0 & . & 0 \\
. & . & . & . & . & . & . & . & . & . & . \\
0 & 0 & . & 0 & 0 & . & \tau_{00}+\sigma^2 & 0 & 0 & 0 & 0 \\
0 & 0 & . & 0 & 0 & . & 0 & \tau_{00}+\sigma^2 & \tau_{00} & . & \tau_{00} \\
0 & 0 & . & 0 & 0 & . & 0 & \tau_{00} & \tau_{00}+\sigma^2 & . & \tau_{00} \\
. & . & . & . & . & . & . & . & . & . & . \\
0 & 0 & . & 0 & 0 & . & 0 & \tau_{00} & \tau_{00} & . & \tau_{00}+\sigma^2
\end{pmatrix}$$

Zwei Beobachtungen aus derselben Schule *j* besitzen die Kovarianz τ_{00}, weil die Makroresiduen U_{0j} von den Mikroresiduen R_{ij} sowie die Mikroresiduen untereinander unabhängig sind:

$$\begin{aligned}\text{Cov}(Y_{1j}, Y_{2j}) &= \text{Cov}(\gamma_{00} + U_{0j} + R_{1j}, \gamma_{00} + U_{0j} + R_{2j}) \\ &= \text{Var}(U_{0j}) \\ &= \tau_{00}\end{aligned}$$

Für Beobachtungen aus verschiedenen Schulen beträgt die Kovarianz null, weil die Makroresiduen aus verschiedenen Schulen untereinander unabhängig sind.

Da die Korrelation von zwei Zufallsvariablen definiert ist als Quotient aus der Kovarianz der Variablen im Zähler und dem Produkt aus den Standardabweichungen der beiden Variablen im Nenner, ist die Intraklassenkorrelation tatsächlich gerade die Korrelation von zwei Zufallsvariablen zu Beobachtungen aus *derselben* Makroeinheit:

$$\frac{\tau_{00}}{\sqrt{(\tau_{00} + \sigma^2)(\tau_{00} + \sigma^2)}} = \frac{\tau_{00}}{\tau_{00} + \sigma^2}$$

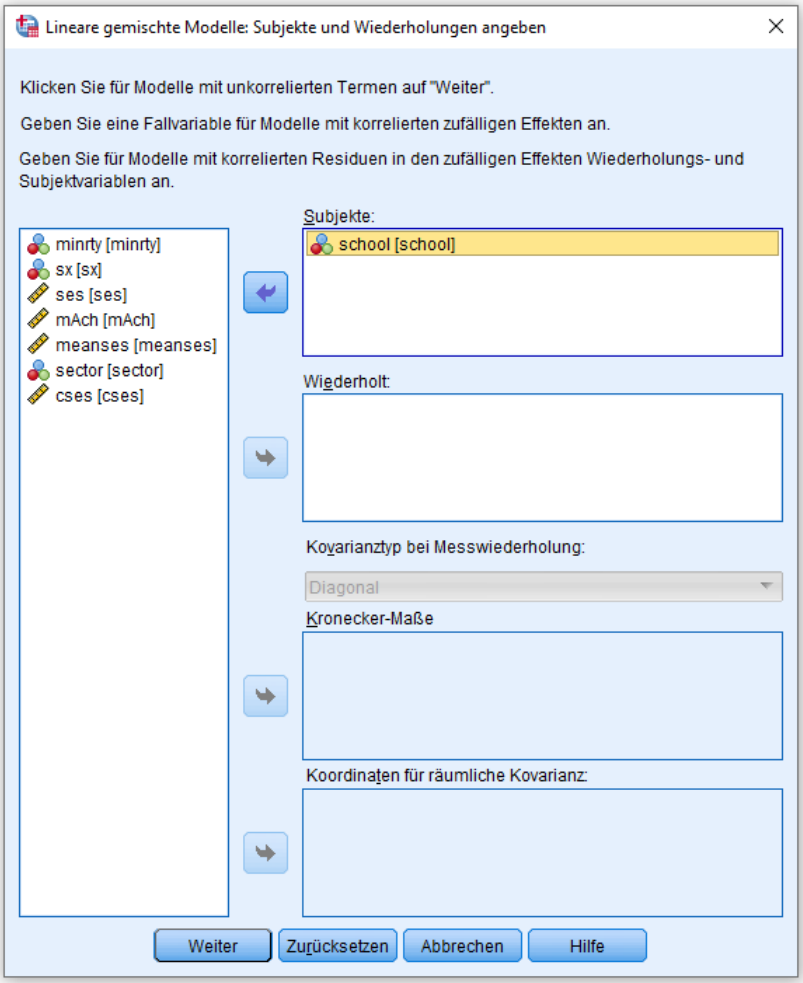
3.2.3 Anforderung der Analyse in SPSS

Weil die Spezifikation eines Mehrebenenmodells über den zuständigen SPSS-Dialog nicht ganz trivial ist, werden in diesem Manuskript alle Schritte durch Bildschirmfotos und Anleitungen sorgfältig beschrieben.[1]

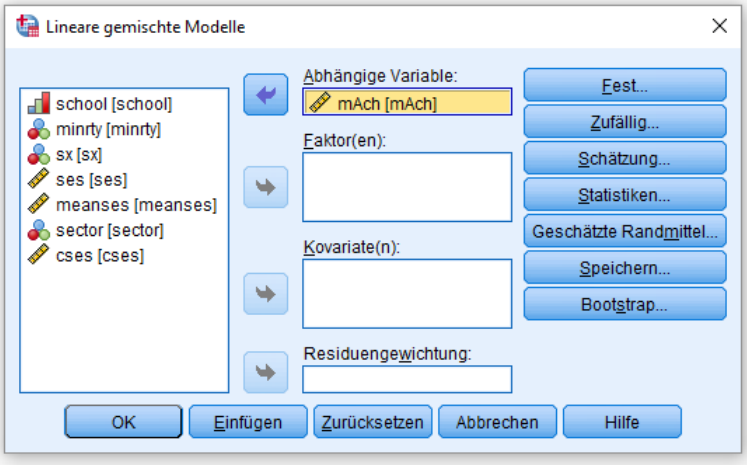
Gehen Sie folgendermaßen vor, um in SPSS das Basismodell auf die HS&B - Daten anzuwenden:

[1] Field (2013, S. 830) bezeichnet die Dialogbox zur Prozedur MIXED deutlich übertreibend als „completely indecipherable“.

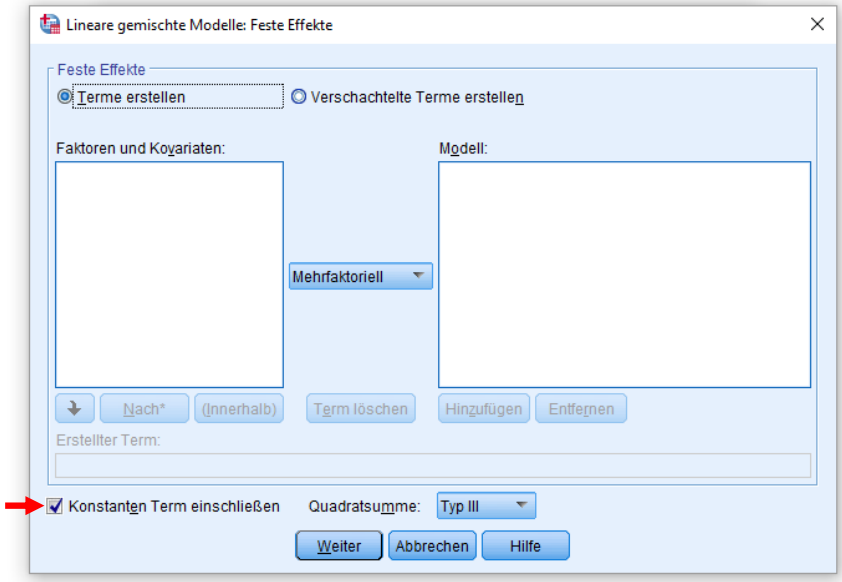
- Menübefehl **Analysieren > Gemischte Modelle > Linear**
- Legen Sie fest, dass die Variable SCHOOL für die spätere Definition der Makroeinheiten in Frage kommt:



- Wir machen **weiter** und wählen MACH als **abhängige Variable**:

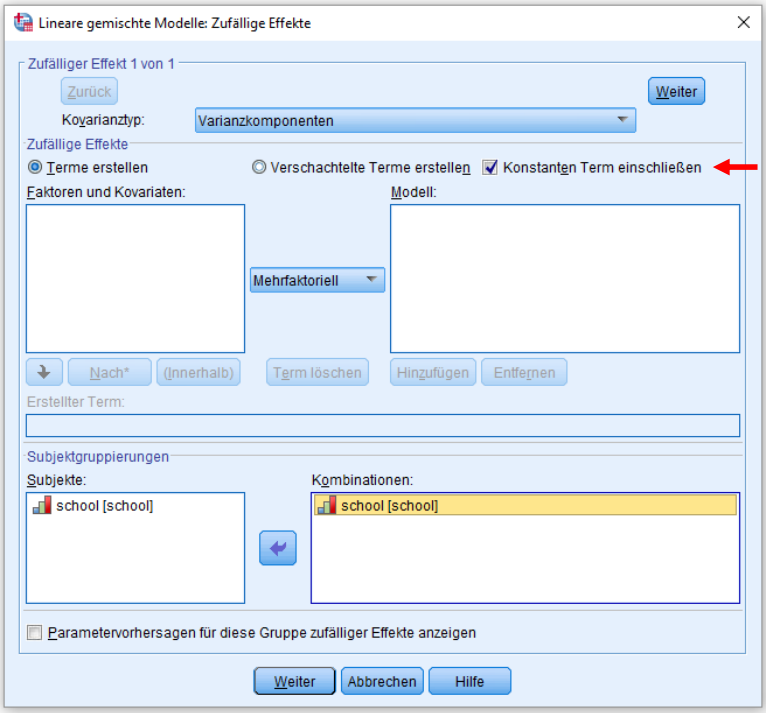


- Über den Schalter **Fest** erreicht man die Subdialogbox für die **festen Effekte**. Hier muss der **konstante Term eingeschlossen** werden (= Voreinstellung):



Anderenfalls würde im Modell der Parameter γ_{00} fehlen.

- Quittieren Sie die Subdialogbox für die festen Effekte mit **Weiter** und begeben Sie sich über den Schalter **Zufällig** in das Gegenstück für die **zufälligen Effekte**:



Hier sind folgende Einstellungen vorzunehmen:

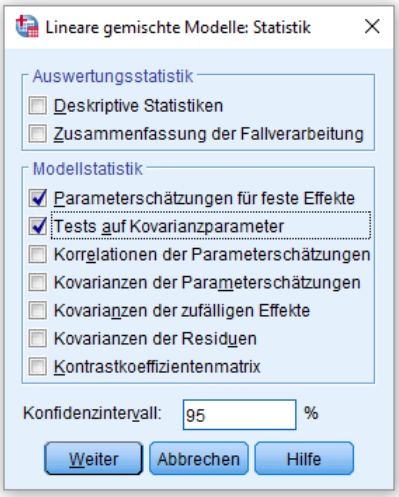
- Als **Kovarianztyp** kann die Voreinstellung **Varianzkomponenten** beibehalten werden. Hier geht es um die Struktur der Kovarianzmatrix der zufälligen Effekte. Momentan enthält unser Modell nur *einen* zufälligen Effekt (U_{0j}), so dass wir uns über die Kovarianzstruktur keine Gedanken machen müssen.
- Der **konstante Term** wird **eingeschlossen**.
  Ansonsten wird der Zufallseffekt der Schulen, also der schulbedingte Einfluss auf die Mathematikleistungen konstant gleich null gesetzt, was vermutlich zu einem unplausiblen kombinierten Modell mit verletzter Unabhängigkeitsannahme führen würde:

$$\tau_{00} = 0$$
$$Y_{ij} = \gamma_{00} + R_{ij}$$

- Die Variable SCHOOL wird in die Liste der **Kombinationen** befördert, womit die oben im Dialog **Subjekte und Wiederholungen angeben** vorbereitete Definition der Makroeinheiten abgeschlossen ist.

Statt SCHOOL als Subjektvariable anzumelden und zur Kombinationsbildung zu verwenden, kann man die Variable auch als zufälligen Effekt deklarieren. Man erhält dieselben Ergebnisse, wartet aber etwas länger auf die Fertigstellung (siehe Norušis 2008, S. 214).

- Machen Sie **weiter**, und wählen Sie in der **Statistik**-Subdialogbox als zusätzliche Ausgaben **Parameterschätzungen für feste Effekte** und **Tests auf Kovarianzparameter**:



Über den **Einfügen**-Schalter der Hauptdialogbox erhält man das folgende MIXED-Kommando als Syntax-Äquivalent zur Modellspezifikation:

```
MIXED mAch
  /CRITERIA=CIN(95) MXITER(100) MXSTEP(10) SCORING(1) SINGULAR(0.000000000001)
   HCONVERGE(0, ABSOLUTE) LCONVERGE(0, ABSOLUTE) PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=| SSTYPE(3)
  /METHOD=REML
  /PRINT=SOLUTION TESTCOV
  /RANDOM=INTERCEPT | SUBJECT(school) COVTYPE(VC).
```

Zur manuellen Erstellung einer äquivalenten Syntax ist durch Ausnutzung der passenden Voreinstellungen nur wenig Schreibarbeit erforderlich:

```
mixed mach
  /print = solution testcov
  /random = intercept | subject(school).
```

Anmerkungen zum MIXED-Kommando für das Basismodell:

- Im PRINT-Subkommando werden folgende Ausgaben angefordert:
  - SOLUTION
    Schätzer und asymptotische Standardfehler für die festen Effekte (im Beispiel: γ_{00})
  - TESTCOV
    Wald-Tests und Konfidenzintervalle für die Kovarianzparameter (im Beispiel: τ_{00} und σ^2)
- Im FIXED-Subkommando werden die festen Effekte spezifiziert. Es ist im Beispiel *nicht* erforderlich, weil der einzige feste Effekt (γ_{00}) per Voreinstellung vorhanden ist.

- Im RANDOM-Subkommando werden die zufälligen Effekte spezifiziert:
  - Das Mikroebenenresiduum (R_{ij}) ist stets vorhanden, muss also nicht angegeben werden.
  - Mit dem Schlüsselwort INTERCEPT wird zum Ausdruck gebracht, dass der konstante Term im Mikroebenenmodell als zufällig aufzufassen ist (als random intercept).
  - Die durch einen senkrechten Strich abgetrennte zweite Spezifikation im RANDOM-Subkommando definiert die Mehrebenenstruktur. Im Beispiel bestimmt die Variable SCHOOL, welche Fälle der Ebene 1 gemeinsam einen Fall der Ebene 2 bilden.

3.2.4 Diskussion der Ergebnisse

3.2.4.1 Modellspezifikation

In der Tabelle **Modelldimension**

Modelldimension[a]

		Anzahl Ausprägungen	Kovarianzstruktur	Anzahl Parameter	Subjektvariablen
Feste Effekte	Konstanter Term	1		1	
Zufallseffekte	Konstanter Term [b]	1	Varianz-Komponenten	1	school
Residuum				1	
Gesamt		2		3	

a. Abhängige Variable: mAch.

b. Ab Version 11.5 gelten andere Syntaxregeln für den Unterbefehl RANDOM. Die eingegebene Befehlssyntax führt zu abweichenden Ergebnissen im Vergleich zu früheren Versionen. Wenn Sie Version-11-Syntax verwenden, ziehen Sie das aktuelle Syntax-Referenzhandbuch zu Rate.

protokolliert SPSS, dass unser Modell die folgenden Parameter enthält:

- **Feste Effekte**: Das Globalmittel γ_{00} der Schulmittelwerte
- **Zufallseffekte**: Die Varianzkomponente τ_{00} (Variabilität der Schulmittelwerte)
- **Residuum**: Die Varianzkomponente σ^2 (Variabilität innerhalb der Schulen)

3.2.4.2 Likelihood und Informationskriterien zum Modellvergleich

In der folgenden Tabelle

Informationskriterien[a]

Eingeschränkte -2 Log Likelihood	47116,793
Akaike-Informationskriterium (AIC)	47120,793
Hurvich und Tsai (IC)	47120,795
Bozdogan-Kriterium (CAIC)	47136,553
Bayes-Kriterium von Schwarz (BIC)	47134,553

Die Informationskriterien werden in einem möglichst kleinen Format angezeigt.

a. Abhängige Variable: mathach.

präsentiert SPSS die mit -2 vormultiplizierte restringierte (REML-Schätzung) oder unrestringierte (ML-Schätzung) logarithmierte Likelihood des Modells (-2LL) sowie darauf basierende Informationskriterien, welche die Güte der Modellanpassung an die Daten quantifizieren und dabei in unterschiedlichem Ausmaß komplexe Modell mit vielen Parametern bestrafen. Diese Maße fallen umso kleiner aus, je besser ein Modell zu den Daten passt. Ohne Bezugsmodell sind sie nur schwer zu interpretieren, weil ihre Werte von der Stichprobengröße abhängen und nicht normiert sind.

In der Literatur wird die (-2) multiplizierte logarithmierte Likelihood auch als *Devianz* bezeichnet. Sie spielt eine wichtige Rolle bei Likelihood-Quotienten - Tests über Parameter oder Parametergruppen (siehe Abschnitt 4.2), wobei ein Vergleich von geschachtelten Modellen stattfindet.

Die Informationskriterien werden insbesondere zum Vergleich von nicht-geschachtelten Modellen verwendet.

Weil bei Verwendung der REML-Schätzung die Likelihood nur von den Varianzen und Kovarianzen der Zufallseffekte abhängt, dürfen in dieser Situation nur Modelle mit einer identischen Struktur fester Parameter über die Informationskriterien verglichen werden (Hox et al. 2018, S. 39). Damit sind insbesondere keine Likelihood-Quotienten - Tests zu Regressionskoeffizienten möglich.

Die beiden am häufigsten genutzten Kriterien sind:

- **Akaike-Informationskriterium** (**AIC**)
  Dieses Kriterium ist leicht zu bestimmen, indem zum -2LL - Wert die verdoppelte Anzahl *d* der Modellparameter addiert wird:

  $$AIC = -2LL + 2 \cdot d$$

  Dabei ist zu beachten, dass die Anzahl der Parameter von der Schätzmethode abhängt (vgl. Abschnitt 4.1):
  - REML
    d = Anzahl der Kovarianzparameter
  - ML
    d = Anzahl der festen Parameter (Regressionskoeffizienten) plus Anzahl der Kovarianzparameter

  Bei Verwendung der voreingestellten REML-Schätzmethode erhalten wir im Beispiel (mit zwei Varianzparametern):

  $$AIC = 47116{,}793 + 2 \cdot 2 = 47120{,}793$$

- **Bayes-Kriterium von Schwarz** (**BIC**)
  Bei diesem Kriterium fällt der Malus für die Zahl der aufgewendeten Parameter im Vergleich zum AIC etwas stärker aus. Mit der eben erläuterten Parameterzahl *d* resultiert die folgende Formel:

  $$BIC = -2LL + d \cdot \ln(N)$$

  Auch bei der Stichprobengröße *N* ist ein Unterschied zwischen den Schätzmethoden zu beachten (nach Norušis 2008, S. 203):
  - REML
    N = Anzahl der Mikroeinheiten minus Anzahl der festen Parameter (Regressionskoeffizienten)
  - ML
    N = Anzahl der Mikroeinheiten

  Während der BIC in der Modellbeurteilung generell gegenüber dem AIC bevorzugt wird, plädieren Hox et. al. (2018, S. 39) bei Mehrebenenmodellen für den AIC, weil bei dieser Modellklasse die zur BIC-Berechnung benötigte Stichprobengröße unklar sei:

  > Since multilevel data have a different sample size at different levels, the AIC is more straightforward than the BIC, and therefore the recommended choice.

  Bei Verwendung der voreingestellten REML-Schätzmethode erhalten wir im Beispiel:

  $$BIC = 47116{,}793 + 2 \cdot \ln(7185-1) = 47134{,}553$$

3.2.4.3 Beurteilung der festen Effekte

Für den einzigen festen Effekt in unserem Modell (den konstanten Term γ_{00} im Makromodell) erhalten wir einen Schätzwert von 12,64:

Schätzungen fester Parameter[a]

Parameter	Schätzung	Std.-Fehler	Freiheitsgrade	T-Statistik	Sig.	Konfidenzintervall 95% Untergrenze	Konfidenzintervall 95% Obergrenze
Konstanter Term	12,636974	,244394	156,647	51,707	,000	12,154242	13,119706

a. Abhängige Variable: mAch.

Es handelt sich um den *ungewichteten* Mittelwert der MACH-Schulmittelwerte, der im Allgemeinen (bei ungleichen Gruppenstärken) verschieden ist vom MACH-Mittelwert der Gesamtstichprobe.

Für den Test zur Hypothese

$$H_0: \gamma_{00} = 0$$

bietet SPSS zwei äquivalente Prüfstatistiken an. Die **T-Statistik** (siehe obige Tabelle mit den **Schätzungen fester Parameter**) ergibt sich als Quotient aus der Schätzung und dem zugehörigen geschätzten Standardfehler. Der **F-Wert** in der Tabelle mit den **Tests auf feste Effekte** ergibt sich durch Quadrieren der T-Statistik:

Tests auf feste Effekte, Typ III[a]

Quelle	Zähler-Freiheitsgrade	Nenner-Freiheitsgrade	F-Wert	Signifikanz
Konstanter Term	1	156,647	2673,663	,000

a. Abhängige Variable: mathach.

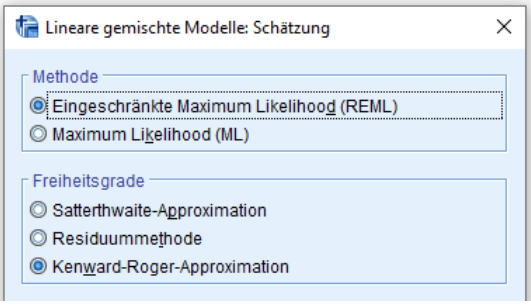
Bei Verwendung der voreingestellten REML-Schätzmethode (vgl. Abschnitt 4.1) handelt es sich im Wesentlichen um einen Einstichproben - T-Test für die aggregierten Daten (160 Fälle mit den MACH-Mittelwerten pro Schule). Aufgrund der verschiedenen Stichprobenumfänge besitzen die Mittelwerte zu den einzelnen Schulen allerdings unterschiedliche Fehlervarianzen, was bei der Berechnung der Freiheitsgrade zu berücksichtigen ist. Bei identischen Stichprobenumfängen in den Schulen (also bei balancierten Daten) würde (bei Verwendung der voreingestellten REML-Schätzmethode) exakt der Einstichproben - T-Test mit $N - 1 = 159$ Freiheitsgraden resultieren.

Die separate **Tests**-Tabelle ist *dann* eine notwendige Ergänzung der Tabelle mit den Schätzwerten, wenn ein Effekt (z.B. zu einem nominalskalierten Regressor) durch *mehrere* Kodiervariablen repräsentiert wird. Dann liefert die **Tests**-Tabelle den gemeinsamen Test zu *allen* Kodiervariablen des Effekts.

Im generellen MIXED-Modell werden bei der Prüfung eines festen Effekts die aus modellgemäßen Abhängigkeiten resultierenden Varianzen und Kovarianzen der Beobachtungen berücksichtigt ...

- bei der Berechnung der F- bzw. T-Statistik
- und bei der Anpassung der Freiheitsgrade.

Zur Berechnung der Freiheitsgrade steht neben der per Voreinstellung verwendeten **Satterthwaite-Approximation** in der Prozedur MIXED seit der SPSS-Version 26 im Subdialog **Schätzung**



auch die präzisere **Kenward-Roger - Approximation** zur Verfügung. Bei großem Stichprobenumfang (wie im HS&B - Beispiel) sind die Unterschiede zwischen den beiden Verfahren allerdings vernachlässigbar.

3.2.4.4 Beurteilung der zufälligen Effekte bzw. Varianzkomponenten

In der folgenden Tabelle liefert die Prozedur MIXED Schätzwerte zu den Varianzkomponenten:

Schätzungen von Kovarianzparametern[a]

Parameter		Schätzung	Standard Fehler	Wald Z	Sig.	Konfidenzintervall 95% Untergrenze	Konfidenzintervall 95% Obergrenze
Residuum		39,148322	,660645	59,258	,000	37,874662	40,464813
Konstanter Term [Subjekt = school]	Varianz	8,614014	1,078801	7,985	,000	6,739115	11,010532

a. Abhängige Variable: mAch.

Der Schätzwert zur Varianz τ_{00} der Schulmittelwerte (8,61) fällt deutlich kleiner aus als der Schätzwert zur Residualvarianz σ^2 innerhalb der Schulen (39,15). In späteren Modellen wird versucht, durch Regressoren auf der Makroebene einen möglichst großen Anteil der Makrovarianz von 8,61 und durch Regressoren auf der Mikroebene einen möglichst großen Anteil der Mikrovarianz von 39,15 zu erklären. Grundsätzlich kann ein Mikroregressor auch zur Aufklärung der Makrovarianz beitragen, sofern sich die Gruppen bei diesem Mikroregressor unterscheiden. Demgegenüber kann ein Makroregressor keinen Beitrag zur Erklärung der Residualvarianz σ^2 auf der Mikroebene leisten, weil er innerhalb der Gruppen konstant ist.

Für die Intraklassenkorrelation im Beispiel erhalten wir den folgenden Schätzwert:

$$\hat{\rho} = \frac{\hat{\tau}_{00}}{\hat{\tau}_{00} + \hat{\sigma}^2} = \frac{8{,}614}{8{,}614 + 39{,}148} = 0{,}18$$

Ca. 18% der gesamten MACH-Varianz (bei zufälliger Ziehung eine Schülers aus einer zufällig gewählten Schule) werden durch die Unterschiede zwischen den Schulen erklärt. Folglich darf nach der oben zitierten Empfehlung von Heck et al. (2010, S. 6 und S. 74) die Abhängigkeit zwischen Schülern aus derselben Schule bei der statistischen Analyse nicht ignoriert werden.

Zur Beurteilung der Nullhypothesen

$$H_0: \tau_{00} = 0 \quad \text{bzw.} \quad H_0: \sigma^2 = 0$$

rechnet SPSS auf Wunsch **Wald-Tests**, obwohl diese zumindest in kleinen Stichproben mit wenigen Gruppen als fragwürdig gelten. Ihre Prüfgröße wird als Quotient aus dem Schätzer und seinem Standardfehler errechnet und mit den Quantilen der Standardnormalverteilung verglichen.

Als bessere Alternative gelten die in Abschnitt 4.2 vorzustellenden Likelihood-Quotienten - Tests, die durch den Vergleich von zwei geschachtelten Modellen zu einer χ^2-verteilten Prüfgröße gelangen.

Ist die Anzahl der Gruppen nicht allzu groß, kann die Signifikanz von τ_{00} auch über den Gruppenfaktor im linearen Modell mit fixierten Effekten beurteilt werden (siehe Abschnitt 2.3).

Im Beispiel kann man davon ausgehen, dass beide Varianzen größer null sind, und mit τ_{00} ist auch die Intraklassenkorrelation signifikant von null verschieden.

3.3 Regressoren auf der Makroebene

Nachdem in Abschnitt 3.2 ein Basismodell für die HS&B - Daten vorgestellt wurde, machen wir uns nun an den Ausbau zu einem realistischen Erklärungsansatz für die Mathematikleistungen der Schüler. Wir lassen zunächst das Mikromodell unverändert und nehmen in das Makromodell eine potentielle Bedingung für die mittlere Mathematikleistung einer Schule auf: den durchschnittlichen sozioökonomischen Status ihrer Schüler (Variable MEANSES). Eine Besonderheit dieses Makromerkmals besteht darin, dass es durch Aggregation aus einem Mikromerkmal hervorgegangen ist.

Beim Ausbau des Basismodells in Anlehnung an Raudenbush & Bryk (2002, Kap. 4) zunächst einen Makroregressor aufzunehmen und erst später einen Mikroregressor, ist nicht unbedingt die häufigste Vor-

gehensweise, aber auch nicht fehlerhaft. In Abschnitt 4.4 werden wir uns mit Modellierungsstrategien beschäftigen.

Raudenbush & Bryk (2002, S. 72ff) verwenden für die nun betrachtete Modellklasse die Bezeichnung *Regression with Means-as-Outcomes*.

3.3.1 Modell

Das Mikromodell erklärt die *i*-te Beobachtung in der Makroeinheit *j* wie bei der einfaktoriellen Varianzanalyse mit zufälligem Schulfaktor (siehe Abschnitt 3.2.1) durch den Schulmittelwert β_{0j} und ein unabhängig identisch normalverteiltes Residuum:

$$Y_{ij} = \beta_{0j} + R_{ij} \quad \text{für } i = 1, .., n_j \text{ und } j = 1, .., J$$

Im Makroebenenmodell wird der Regressor MEANSES (Abkürzung *W*) eingesetzt, um den Schulmittelwert β_{0j} teilweise zu erklären:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + U_{0j} \quad \text{für } j = 1, .., J \text{ mit } U_{0j} \sim \text{unabhängig } N(0, \tau_{00})$$

Um die Interpretation des Parameters γ_{00} (prognostizierter MACH-Mittelwert für eine Schule mit dem MEANSES-Wert 0) zu erleichtern, wurde die Variable MEANSES zentriert. Von jedem Wert wurde der Mittelwert der 160 Schulen abgezogen. Somit steht γ_{00} für den prognostizierten MACH-Mittelwert bei einer Schule mit mittlerem MEANSES-Wert.

In der Varianz τ_{00} von U_{0j} kommt jetzt zum Ausdruck, wie stark sich die MACH-Schulmittelwerte nach Kontrolle der Variablen MEANSES noch unterscheiden. Man kann mit Raudenbush & Bryk (2002, S. 73) hier von einer *residualen* bzw. *bedingten Varianz* sprechen.

Um die Erklärungsleistung des Makroregressors *W* in Bezug auf die β_{0j} - Varianz zu beurteilen, setzt man die Differenz zwischen der Varianz $\tau_{00|B}$ aus dem Basismodell ohne Regressoren (*fully unconditional* - Modell, siehe Abschnitt 3.2) und der Residualvarianz $\tau_{00|M}$ aus dem Modell mit Makroregressor ins Verhältnis zum Ausgangswert $\tau_{00|B}$. Diese mit dem Determinationskoeffizienten der multiplen Regression vergleichbare Statistik soll mit $R^2_{\beta_0}$ bezeichnet werden (erklärter Anteil der β_0 - Varianz):

$$R^2_{\beta_0} = \frac{\tau_{00|B} - \tau_{00|M}}{\tau_{00|B}}$$

Das kombinierte Modell für die abhängige Variable Y_{ij}

$$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + U_{0j} + R_{ij}$$

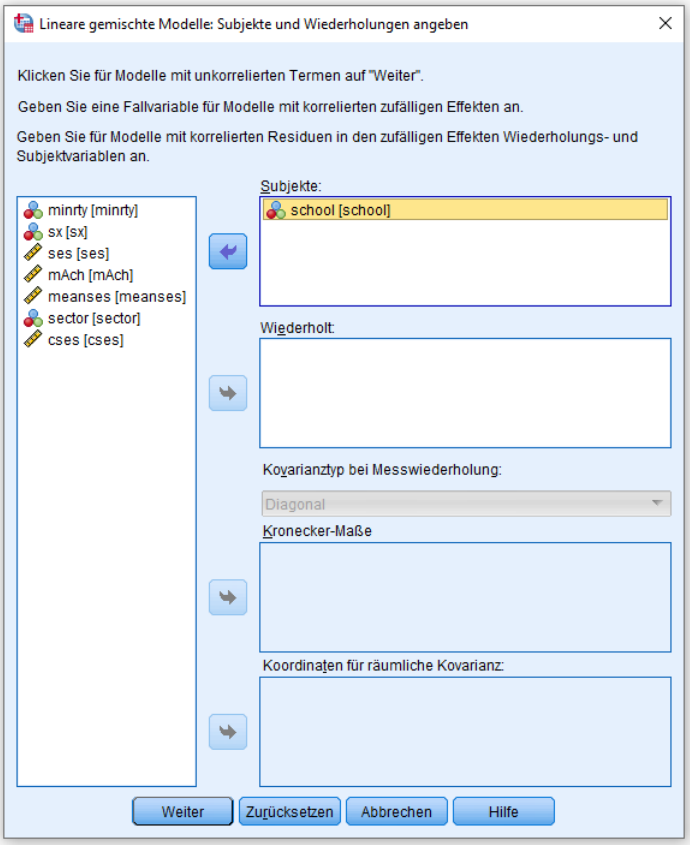
enthält die festen Effekte γ_{00} und γ_{01}, den zufälligen Effekt U_{0j} mit der Varianz τ_{00} und das Mikroresiduum R_{ij} mit der Varianz σ^2.

Wie bei allen Mehrebenenmodellen wird angenommen, dass die Makroresiduen U_{0j} untereinander, die Mikroresiduen R_{ij} untereinander sowie die Makro- von den Mikroresiduen unabhängig sind.

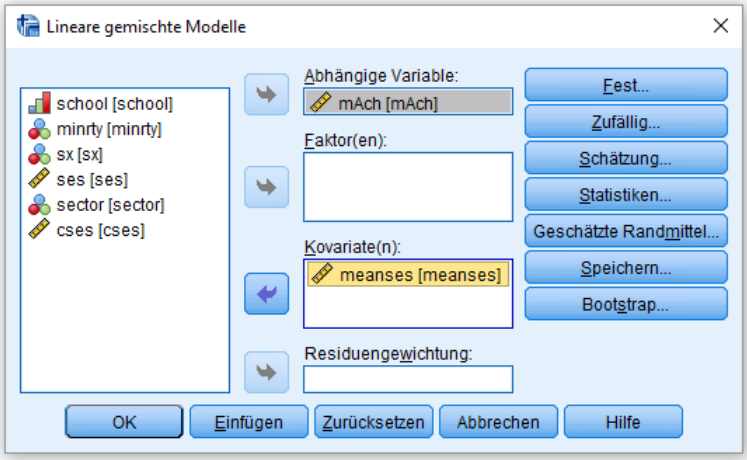
3.3.2 Anforderung der Analyse in SPSS

Gehen Sie folgendermaßen vor, um in SPSS das Modell mit Makroregressor auf die HS&B - Daten anzuwenden:

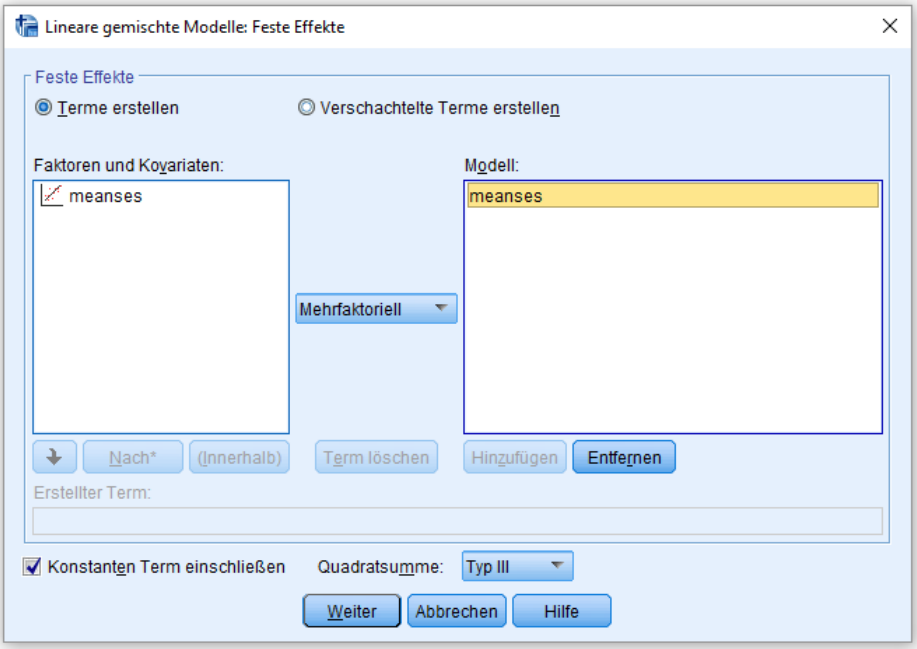
- Menübefehl **Analysieren > Gemischte Modelle > Linear**
- Wir legen fest, dass die Variable SCHOOL für die spätere Definition der Makroeinheiten in Frage kommt:



- Wir machen **weiter** und wählen MACH als **abhängige Variable** sowie MEANSES als **Kovariate**:

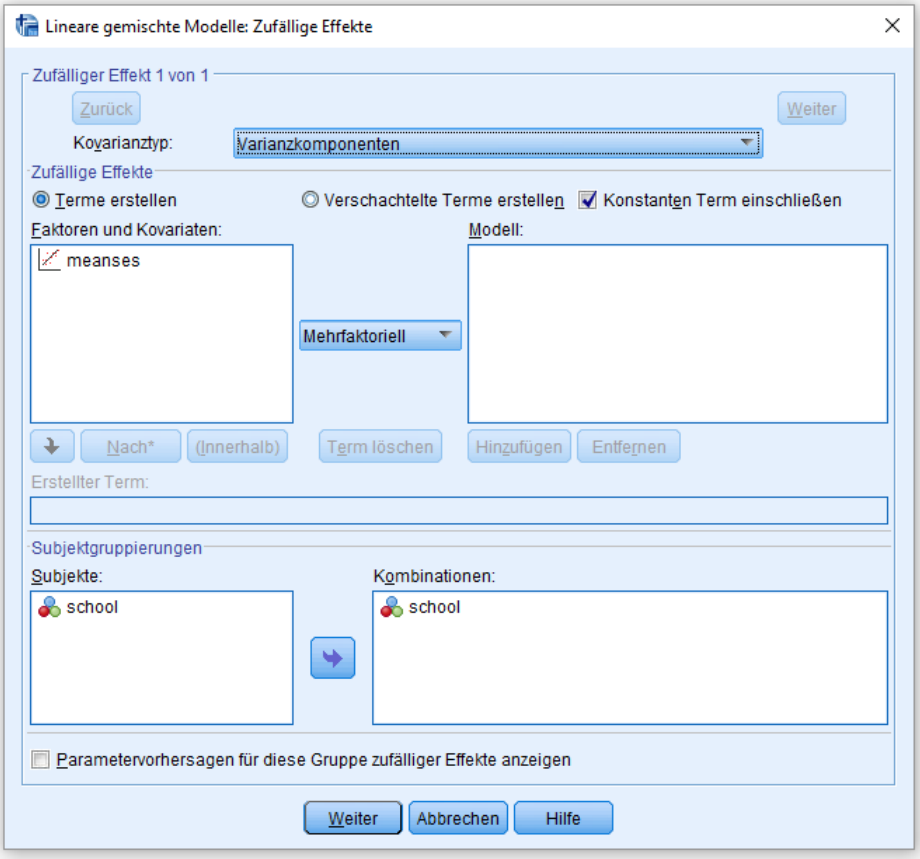


- Wir fügen in folgender Subdialogbox für die **festen Effekte** die Variable MEANSES zum **Modell** hinzu (entspricht dem Parameter γ_{01} aus der Modellformulierung):



Das Kontrollkästchen für den **konstanten Term** muss auf jeden Fall markiert sein (= Voreinstellung), weil ansonsten ein Modell ohne den Parameter γ_{00} geschätzt wird.

- Wir quittieren die Subdialogbox für die festen Effekte mit **Weiter** und begeben uns in das Gegenstück für die **zufälligen Effekte**, wo die Einstellungen aus Abschnitt 3.2.3 übernommen werden können:



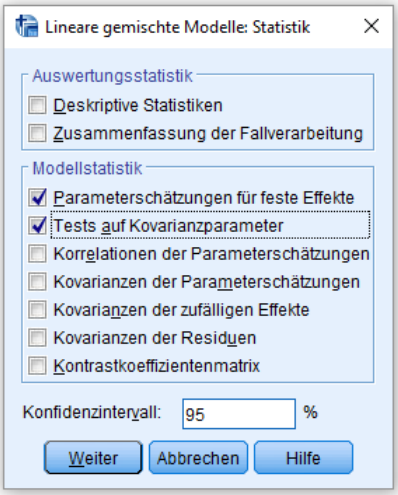
Erläuterungen:

- Als **Kovarianztyp** können wir die Voreinstellung **Varianzkomponenten (VK)** beibehalten (siehe Abschnitt 3.2.3).
- Der **konstante Term** wird **eingeschlossen**.
  Ansonsten fehlt im kombinierten Modell der zufällige Effekt auf die Schulmittelwerte:

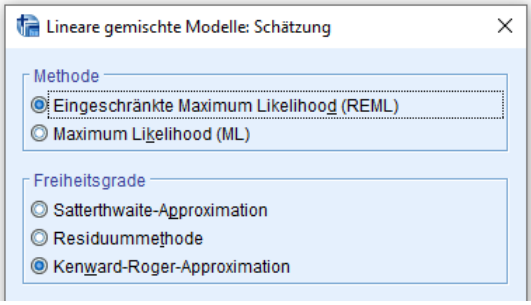
$$Y_{ij} = \gamma_{00} + \gamma_{01} W + R_{ij}$$

  Damit wären die folgenden, wenig realistischen Annahmen verbunden:
  - Die Residualvarianz τ_{00} der Schulmittelwerte ist nach Kontrolle der Variablen MEANSES gleich null.
  - Das lineare Modell mit den Schülern als Beobachtungseinheiten und der Schuleigenschaft MEANSES als Regressor besitzt unabhängige Residuen.
- Die Variable SCHOOL wird in die Liste der **Kombinationen** befördert, womit die oben im Dialog **Subjekte und Wiederholungen** vorbereitete Definition der Makroeinheiten abgeschlossen ist.

- Wir fordern in der **Statistik**-Subdialogbox **Parameterschätzer für feste Effekte** und **Tests auf Kovarianzparameter** an:



- Schließlich fordern wir im Subdialog **Schätzung** die **Kenward-Roger-Approximation** an, um die Präzision der Signifikanztests zu den festen Effekten zu verbessern, wobei in einer großen Stichprobe allerdings keine bedeutsamen Unterschiede im Vergleich zur voreingestellten Satterthwaite-Approximation zu erwarten sind:



Über den **Einfügen**-Schalter der Hauptdialogbox erhält man folgendes Syntax-Äquivalent zu der Modellspezifikation:

```
MIXED mAch WITH meanses
  /CRITERIA=DFMETHOD(KENWARDROGER) CIN(95) MXITER(100) MXSTEP(10) SCORING(1)
    SINGULAR(0.000000000001) HCONVERGE(0, ABSOLUTE) LCONVERGE(0, ABSOLUTE)
    PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=meanses | SSTYPE(3)
  /METHOD=REML
  /PRINT=SOLUTION TESTCOV
  /RANDOM=INTERCEPT | SUBJECT(school) COVTYPE(VC).
```

Zur manuellen Erstellung einer äquivalenten Syntax ist durch Ausnutzung der passenden Voreinstellungen nur wenig Schreibarbeit erforderlich:

```
mixed mach with meanses
  /criteria=dfmethod(kenwardroger)
  /fixed = meanses
  /random = intercept | subject(school)
  /print = solution testcov.
```

Im Vergleich zur Syntax für das Basismodell (siehe Abschnitt 3.2.3) sind hinzugekommen:

- das CRITERIA-Subkommando zur Wahl der Kenward-Roger - Approximation
- das FIXED-Subkommando zur Spezifikation fester Effekte

3.3.3 Diskussion der Ergebnisse

In der Modellspezifikation taucht im Vergleich zum Basismodell der feste Effekt von MEANSES (γ_{01}) als zusätzlicher Parameter auf:

Modelldimension[a]

		Anzahl Ausprägungen	Kovarianzstruktur	Anzahl Parameter	Subjektvariablen
Feste Effekte	Konstanter Term	1		1	
	meanses	1		1	
Zufallseffekte	Konstanter Term	1	Varianz-Komponenten	1	school
Residuum				1	
Gesamt		3		4	

a. Abhängige Variable: mAch.

Die Informationskriterien zeigen im Vergleich zum Basismodell (durch kleinere Werte) eine verbesserte Anpassung an die Daten:

Informationskriterien[a]

Eingeschränkte -2 Log Likelihood	46961,291
Akaike-Informationskriterium (AIC)	46965,291
Hurvich und Tsai (IC)	46965,293
Bozdogan-Kriterium (CAIC)	46981,050
Bayes-Kriterium von Schwarz (BIC)	46979,050

Die Informationskriterien werden in einem möglichst kleinen Format angezeigt.

a. Abhängige Variable: mAch.

Für die Parameter γ_{00} und γ_{01} aus dem Makromodell erhalten wir die folgenden Schätzwerte und Signifikanztests:

Schätzungen fester Parameter[a]

Parameter	Schätzung	Standardfehler	Freiheitsgrade	T-Statistik	Sig.	Konfidenzintervall 95% Untergrenze	Obergrenze
Konstanter Term	12,684609	,149302	153,704	84,960	,000	12,389662	12,979557
meanses	5,863539	,361498	153,410	16,220	,000	5,149382	6,577695

a. Abhängige Variable: mAch.

Von besonderem Interesse ist der Steigungsparameter γ_{01} zu MEANSES:

- Unterscheiden sich zwei Schulen beim Regressor MEANSES um eine Einheit, dann ist bei MACH ein Unterschied von 5,86 Punkten zu erwarten.
- Im T-Test mit Kenward-Roger - Approximation wird der Effekt als signifikant beurteilt.
- SPSS teilt das Ergebnis eines *zweiseitigen* Tests mit. Unter den üblichen Voraussetzungen ist durchaus auch ein einseitiger Test möglich (Snijders & Bosker 2012, S. 94). Dazu ist die von SPSS mitgeteilte Überschreitungswahrscheinlichkeit zu halbieren.

Der konstante Term hat sich im Vergleich zum Basismodell mit zufälligem Schulfaktor kaum geändert, weil der Regressor MEANSES zentriert ist.[1]

In der Tabelle mit den **Kovarianzparametern** interessiert vor allem die bedingte Varianz der Schulmittelwerte nach der Erweiterung des Modells um den Makroregressor MEANSES (Parameter τ_{00}):

Schätzungen von Kovarianzparametern[a]

Parameter		Schätzung	Std.-Fehler	Wald Z	Sig.	Konfidenzintervall 95% Untergrenze	Obergrenze
Residuum		39,157077	,660801	59,257	,000	37,883115	40,473881
Konstanter Term [Subjekt = school]	Varianz	2,638878	,404358	6,526	,000	1,954289	3,563279

a. Abhängige Variable: mAch.

Die bedingte Varianz von 2,64 ist erheblich kleiner als die unbedingte Varianz von 8,61 (siehe Abschnitt 3.2.4.4). Damit klärt MEANSES immerhin 69% der Schulunterschiede in der Matheleistung auf:

$$\hat{R}^2_{\beta_0} = \frac{8{,}614 - 2{,}639}{8{,}614} = 0{,}694$$

Nach dem (allerdings nicht sonderlich zuverlässigen) Wald-Test ist die bedingte Varianz der Makroebene immer noch signifikant von null verschieden, so dass die Suche nach weiteren Makroregressoren lohnenswert erscheint.

Bei der Residualvarianz innerhalb der Schulen ergibt sich erwartungsgemäß praktisch keine Änderung im Vergleich zum Basismodell (ohne Makroregressor).

Für die von Singer (1998, S. 333) vorgeschlagene *bedingte* bzw. *partielle Intraklassenkorrelation* für MACH nach Kontrolle von MEANSES ergibt sich:

$$\frac{2{,}639}{2{,}639 + 39{,}157} = 0{,}063$$

Durch die Aufnahme des Makroregressors MEANSES in das Modell wurde das Problem abhängiger Mikroresiduen also erheblich reduziert.

[1] Beim Ordinatenabschnitt handelt es sich um die Modellprognose für den Wert null des Regressors. Dies ist aber der Mittelwert des Regressors, und für den Mittelwert des Regressors prognostiziert das lineare Modell den Mittelwert des Kriteriums, also den mittleren MACH-Wert aller Schulen. Dabei gehen alle Schulen (unabhängig von der Schülerzahl) mit gleichem Gewicht ein.

3.4 Regressoren auf der Mikroebene

Das nächste Modell verzichtet auf einen Makroregressor und enthält stattdessen einen Regressor auf der Mikroebene, wobei in jeder (zufällig gezogenen!) Makroeinheit für die Regression des Kriteriums auf den Mikroregressor eigenständige Koeffizienten (Ordinatenabschnitt und Steigung) erlaubt sind. Damit enthält das Modell neben einem *random intercept* auch einen *random slope* für den Mikroregressor. Raudenbush & Bryk (2002, S. 75ff) verwenden für die betrachtete Modellklasse die Bezeichnung *random coefficient model*.

Im Beispiel wird als Mikroregressor die Variable CSES einbezogen, die für jeden Schüler die Abweichung seiner sozioökonomischen Statuseinschätzung vom Mittelwert seiner Schule enthält. Damit lässt sich untersuchen, welchen Effekt *innerhalb* der Schulen der soziökonomische Status der Schüler auf ihre Mathematikleistung hat. Es lassen sich u.a. folgende Fragestellungen untersuchen:

- Wie groß ist der Statuseffekt auf die Mathematikleistung innerhalb der Schulen im Durchschnitt (random slope - Mittelwert von CSES)?
- Unterscheiden sich die Schulen im Statuseffekt auf die Mathematikleistung (random slope - Varianz)?
- Wir haben schon festgestellt, dass 69% der MACH-Varianz *zwischen* den Schulen durch MEAN-SES aufgeklärt werden, und können uns nun fragen, wieviel Prozent der schul*internen* MACH-Varianz von schulinternen Unterschieden im sozioökonomischen Status (Variable CSES) erklärt werden.
- Gibt es einen Zusammenhang zwischen dem Niveau einer Schule (Ordinatenabschnitt) und dem CSES-Steigungskoeffizienten? Hier geht es um die Korrelation zwischen dem random intercept - und dem random slope - Zufallseffekt.

3.4.1 Zentrierung von Mikroregressoren

3.4.1.1 Metrische Mikroregressoren

Weil die Koeffizienten des Mikromodells als Explanandum des Makromodells auftreten, ist es bei Mikroregressoren noch weit wichtiger als bei den Makroregressoren (vgl. Abschnitt 3.3.1), für sinnvoll interpretierbare Koeffizienten zu sorgen. Bei einem metrischen Mikroregressor kommt vor allem das Zentrieren um ein Maß der zentralen Tendenz (z. B. Mittelwert oder Median der Stichprobe oder Population) in Frage.

Das Zentrieren von metrischen Regressoren kann nicht nur zu besser interpretierbaren Koeffizienten führen (siehe unten), sondern auch die Konvergenz des Schätzverfahrens erleichtern (Hox 2010, S. 63).

3.4.1.1.1 Globalmittel-Zentrierung

Bei der Globalmittel-Zentrierung eines Mikroregressors *X* subtrahiert man bei jedem Fall den Gesamtmittelwert der Stichprobe:

$$X_{ij} \rightarrow (X_{ij} - \overline{X}_{..})$$

Als Ordinatenabschnitt β_{0j} in der Gruppe *j* erhält man ihre Modellprognose für ein Individuum mit durchschnittlichem *X*-Wert (bezogen auf die Gesamtstichprobe). Sind mehrere Mikroregressoren beteiligt und zentriert, ist β_{0j} die Modellprognose für ein Durchschnittsindividuum (bezogen auf die Gesamtstichprobe).

Durch die Globalmittel-Zentrierung erfährt das Modell *keine wesentliche* Veränderung:

- Bei Beteiligung des zentrierten Regressors an Wechselwirkungen ändern einige Parameter ihre Schätzwerte, Bedeutungen und Signifikanzbeurteilungen, doch handelt es sich dabei lediglich um Anpassungen an die Skalenverschiebung.
- Die Fit-Indikatoren des Modells bleiben gleich.

Wenn man bei einem Mikroregressor sicher ist, dass sein Within- und sein Between-Effekt (siehe nächsten Abschnitt) identisch sind, dann ist die Globalmittel-Zentrierung sinnvoll.

3.4.1.1.2 Gruppeninterne Zentrierung

Bei der *gruppeninternen* Zentrierung eines Mikroregressors *X* subtrahiert man bei jedem Fall den Mittelwert seiner Gruppe (Makroeinheit), so dass die transformierte Variable in jeder Gruppe den Mittelwert null besitzt:

$$X_{ij} \rightarrow (X_{ij} - \overline{X}_{.j})$$

Bei den gruppenspezifischen Regressionen des Kriteriums auf $(X_{ij} - \overline{X}_{.j})$ tritt als Ordinatenabschnitt β_{0j} gerade der gruppeneigene Kriteriumsmittelwert auf. Damit ist Var(β_{0j}) = τ_{00} (wie im Basismodell) die Varianz der Kriteriumsmittelwerte zwischen den Makroeinheiten (vgl. Raudenbush & Bryk 2002, S. 33).

Während die Globalmittelzentrierung zu einem äquivalenten Modell führt, dessen Koeffizienten in einer einfachen Relation zu den Koeffizienten im Modell für die Originaldaten stehen, resultiert aus der gruppeninternen Zentrierung ein Modell mit alternativem Aussagegehalt.

Beim gruppeninternen Zentrieren geht die Information über die mittleren Niveaus der Gruppen verloren. Daher ist es in der Regel empfehlenswert, die Gruppenmittelwerte über einen Makroregressor ins Modell zurückzubringen (Field 2013, S. 830; Tabachnick & Fidell 2007, S. 824).

Für dieses intuitiv plausible Vorgehen bestehen außerdem starke methodische Argumente, weil es die wichtige Unterscheidung zwischen ...

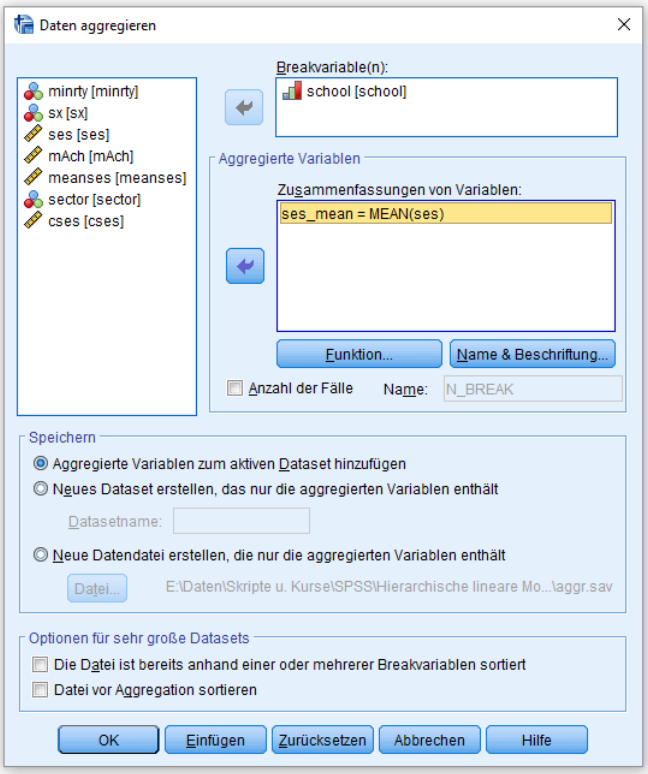
- dem Effekt eines Regressors *innerhalb der Gruppen* (Within-Effekt)
- und dem Effekt desselben Regressors *zwischen den Gruppen* (Between-Effekt)

ermöglicht. Wir haben das Thema im Abschnitt 2.2 über den ökologischen Fehlschluss bereits adressiert (siehe auch Baltes-Götz 2016).

Die im Abschnitt 3.1 beschriebene Datendatei enthält unter dem Namen CSES bereits den aktuell interessierenden Mikroregressor in der gruppenintern zentrierten Form und außerdem unter dem Namen MEANSES eine Variable mit den Gruppenmittelwerten. Trotzdem soll erläutert werden, wie die gruppeninterne Zentrierung in SPSS ausgehend von der Variablen SES mit den ursprünglichen Statuswerten der Schüler durchgeführt wird. Zunächst erzeugt man eine Variable mit den Gruppenmittelwerten, was bequem nach dem Menübefehl:

Daten > Aggregieren

über den folgenden Dialog gelingt:



Im Beispiel wird in der Arbeitsdatei die Variable SES_MEAN erzeugt, die anschließend von der Ausgangsvariablen SES subtrahiert werden muss, um die gruppenintern zentrierte Variante zu erzeugen. Die Transformationen zum Erstellen der beiden neuen Variablen lassen sich durch die folgende Syntax erledigen:

```
AGGREGATE
  /OUTFILE=* MODE=ADDVARIABLES OVERWRITEVARS=YES
  /BREAK=school
  /ses_mean=MEAN(ses).

COMPUTE icses = ses - ses_mean.
EXECUTE.
```

3.4.1.2 Nominalskalierte Mikroregressoren

Bei einem nominalskalierten Mikroregressoren (z. B. Geschlecht) ist die Indikatorkodierung oft akzeptabel (z. B. 1 für Frauen, 0 für Männer). Man erhält als Ordinatenabschnitt β_{0j} in der Gruppe *j* ihre Modellprognose für Fälle in der Null-Kategorie des Regressors, die gleichzeitig einen Durchschnittswert für die (zentrierten) metrischen Mikroregressoren besitzen.

Soll ein dichotom-kategorialer Mikroregressor „zentriert“ werden, dann bietet sich die Verwendung der *ungewichteten Effektkodierung* an (im Beispiel: 0,5 für Frauen und -0,5 für Männer). So erhält man als Ordinatenabschnitt β_{0j} der Gruppe *j* ihre ungewichtet über beide Kategorien gemittelte Modellprognose. Bei ansonsten zentrierten Mikroregressoren ist β_{0j} die geschlechtsunabhängige Modellprognose für ein durchschnittliches Gruppenmitglied (Hox 2010, S. 61).

3.4.2 Modell

Wir wollen in unserem Anwendungsbeispiel für den soziökonomischen Status der Schüler (abgekürzt bezeichnet durch *X*) eine gruppeninterne Zentrierung vornehmen, so dass folgendes Mikromodell resultiert:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}_{.j}) + R_{ij} \quad \text{für } i = 1, .., n_j \text{ und } j = 1, .., J$$

Für die Residuen R_{ij} wird angenommen, dass sie unabhängig identisch normalverteilt sind mit dem Erwartungswert null und der konstanten Residualvarianz σ^2:

$$R_{ij} \sim \text{unabhängig } N(0, \sigma^2)$$

Im aktuellen Abschnitt 3.4 verzichten wir (in Anlehnung an Raudenbush & Bryk 2002, S. 75ff) auf Makroregressoren, also insbesondere auf den Makroregressor mit den Schulmittelwerten beim sozioökonomischen Status. Die gruppenspezifischen Koeffizienten β_{0j} und β_{1j} werden also im Makromodell jeweils durch ein Populationsmittel und ein Residuum (einen zufälligen Effekt) erklärt:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + U_{0j} \\ \beta_{1j} &= \gamma_{10} + U_{1j} \end{aligned} \quad \text{für } j = 1, .., J$$

Für U_{0j} und U_{1j} wird eine bivariate Normalverteilung angenommen:

$$\begin{pmatrix} U_{0j} \\ U_{1j} \end{pmatrix} \sim \text{unabhängig } N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}\right]$$

Die bivariate Zufallsvariable (U_{0j}, U_{1j}) ist von den Makroresiduen aus anderen Gruppen ebenso unabhängig wie von den Mikroresiduen.

Durch Einsetzen des Makromodells in das Mikromodell erhält man das kombinierte Modell mit den beiden festen Parametern γ_{00} und γ_{10}, den beiden zufälligen Effekten U_{0j} und U_{1j} sowie dem Mikroebenenresiduum R_{ij}:

$$\begin{aligned} Y_{ij} &= \gamma_{00} + U_{0j} + \gamma_{10}(X_{ij} - \bar{X}_{.j}) + U_{1j}(X_{ij} - \bar{X}_{.j}) + R_{ij} \\ &= \gamma_{00} + \gamma_{10}(X_{ij} - \bar{X}_{.j}) \quad + \quad U_{0j} + U_{1j}(X_{ij} - \bar{X}_{.j}) + R_{ij} \end{aligned}$$

Um die Erklärungsleistung von Mikroregressoren in Bezug auf die Mikroebenenvarianz zu beurteilen, setzt man die Differenz zwischen der Varianz σ_B^2 aus dem Basismodell ohne Regressoren (*fully unconditional* - Modell, siehe Abschnitt 3.2) und der Residualvarianz σ_M^2 aus dem Modell mit Mikroregressoren ins Verhältnis zum Ausgangswert σ_B^2. Diese mit dem Determinationskoeffizienten der multiplen Regression vergleichbare Statistik soll hier in Anlehnung an Hox (2010, S. 71) mit R_1^2 bezeichnet werden (erklärter Varianzanteil auf der Ebene 1):

$$R_1^2 = \frac{\sigma_B^2 - \sigma_M^2}{\sigma_B^2}$$

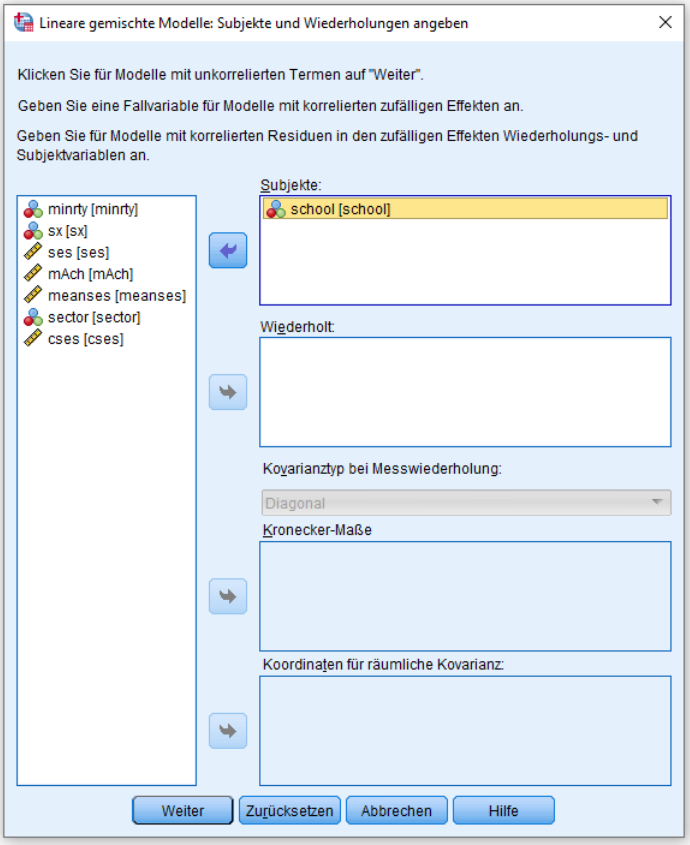
Mikroregressoren können auch einen Beitrag zur Erklärung der Makroebenenvarianz τ_{00} leisten, wenn bei den Mikroregressoren bedeutsame Gruppenunterschiede bestehen, was bei gruppeninterner Zentrierung allerdings ausgeschlossen ist. Um die Erklärungsleistung von Mikroregressoren in Bezug auf die β_{0j} - Varianz zu beurteilen, setzt man die Differenz zwischen der Varianz $\tau_{00|B}$ aus dem Basismodell ohne Regressoren (*fully unconditional* - Modell, siehe Abschnitt 3.2) und der Residualvarianz $\tau_{00|M}$ aus dem Modell mit Mikroregressoren ins Verhältnis zum Ausgangswert $\tau_{00|B}$. Diese mit dem Determinationskoeffizienten der multiplen Regression vergleichbare Statistik soll (wie in Abschnitt 3.3.1) mit $R_{\beta_0}^2$ bezeichnet werden:

$$R^2_{\beta_0} = \frac{\tau_{00|B} - \tau_{00|M}}{\tau_{00|B}}$$

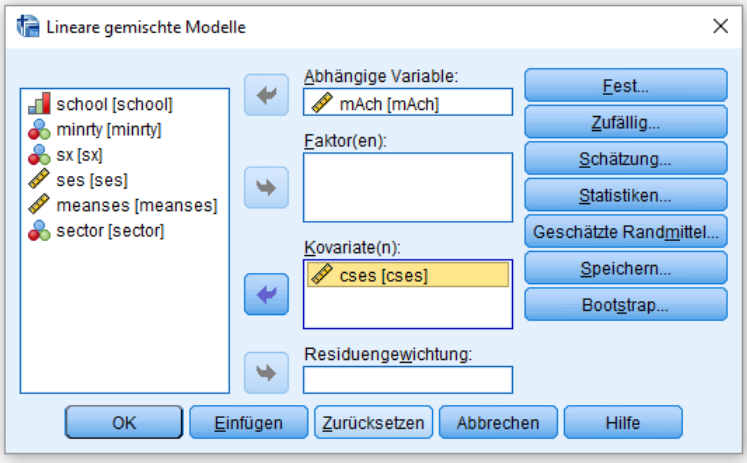
3.4.3 Anforderung der Analyse in SPSS

Die in Abschnitt 3.1 beschriebene Datei mit den HS&B - Daten enthält die Variable CSES mit dem gruppenintern zentrierten individuellen sozioökonomischen Status. Gehen Sie folgendermaßen vor, um in SPSS das eben beschriebene Modell mit zufälligem Ordinatenabschnitt und zufälliger Steigung für die gruppeninterne Regression von MACH auf CSES anzufordern:

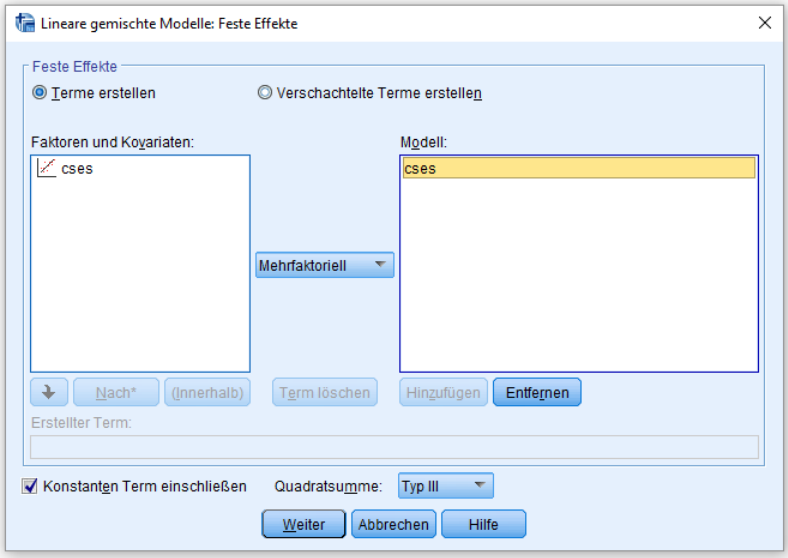
- Menübefehl **Analysieren > Gemischte Modelle > Linear**
- Wir legen fest, dass die Variable SCHOOL für die spätere Definition der Makroeinheiten in Frage kommt:



- Wir machen **weiter** und wählen MACH als **abhängige Variable** sowie CSES als **Kovariate**:

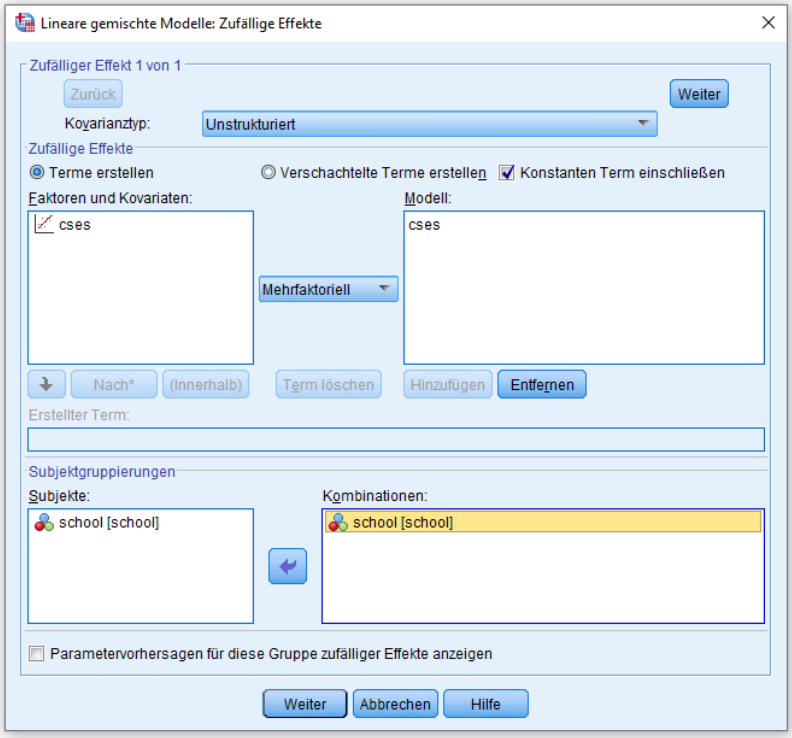


- Wir fügen in folgender Subdialogbox für die **festen Effekte** CSES zum **Modell** hinzu:



Das Kontrollkästchen für den **konstanten Term** muss auf jeden Fall markiert sein (= Voreinstellung), weil ansonsten ein Modell ohne den Parameter γ_{00} geschätzt wird.

- Wir quittieren die Subdialogbox für die festen Effekte mit **Weiter** und begeben uns in das Gegenstück für die **zufälligen Effekte**:



Hier sind folgende Einstellungen vorzunehmen:

- Als **Kovarianztyp** wählen wir **Unstrukturiert (UN)**.
  Hier geht es um die Struktur der (von SPSS mit **G** bezeichneten[1]) Kovarianzmatrix der zufälligen Effekte. Mit der Voreinstellung **Varianzkomponenten (VK)** wäre die Annahme verknüpft, dass U_{0j} (die zufällige Abweichung des Schulmittelwerts von γ_{00}) und U_{1j} (die zufällige Abweichung des schulinternen CSES-Effekts von γ_{10}) unabhängig seien, dass also τ_{01} gleich null sei (Varianzkomponentenmodell). Bei Analysen mit Cluster-Daten wird in der Regel **Unstrukturiert (UN)** als **Kovarianztyp** verwendet. Weil in unseren vorherigen Modellen nur *ein* Zufallseffekt vorhanden war, mussten wir dem Thema bisher keine Aufmerksamkeit widmen.
- Der **konstante Term** wird **eingeschlossen**.
  Ansonsten wird die Varianz τ_{00} der Schulmittelwerte gleich null gesetzt:

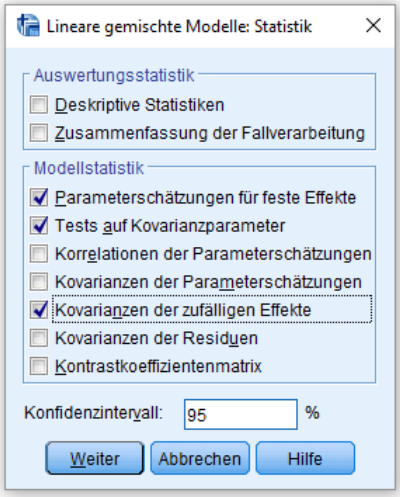
$$\tau_{00} = 0$$
$$\beta_{0j} = \gamma_{00}$$

- Die Variable CSES wird ins Modell befördert.
  Ansonsten wird die Varianz τ_{11} der schulinternen Steigungen gleich null gesetzt:

$$\tau_{11} = 0$$
$$\beta_{1j} = \gamma_{10}$$

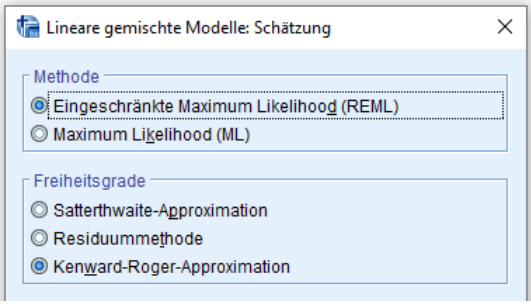
- Die Variable SCHOOL wird in die Liste der **Kombinationen** befördert, womit die oben im Dialog **Subjekte und Wiederholungen** vorbereitete Definition der Makroeinheiten abgeschlossen ist.

- Wir fordern in der **Statistik**-Subdialogbox **Parameterschätzer für feste Effekte**, **Tests auf Kovarianzparameter** und **Kovarianzen der zufälligen Effekte** an:



- Schließlich fordern wir im Subdialog **Schätzung** die **Kenward-Roger-Approximation** an, um die Präzision der Signifikanztests zu den festen Effekten zu verbessern, wobei in einer großen Stichprobe allerdings keine bedeutsamen Unterschiede im Vergleich zur voreingestellten Satterthwaite-Approximation zu erwarten sind:

[1] Die Bezeichnung **G** für die Kovarianzmatrix der zufälligen Effekte stammt aus der üblichen Notation für Modelle mit festen und zufälligen Effekten (siehe Abschnitt 4.7). Bei Raudenbush & Bryk (2002, z. B. S. 76) wird diese Matrix mit dem Buchstaben **T** bezeichnet.



Über den **Einfügen**-Schalter der Hauptdialogbox erhält man folgendes Syntax-Äquivalent zu der Modellspezifikation:[1]

```
MIXED mAch WITH cses
  /CRITERIA=DFMETHOD(KENWARDROGER) CIN(95) MXITER(100) MXSTEP(10) SCORING(1)
    SINGULAR(0.000000000001) HCONVERGE(0, ABSOLUTE) LCONVERGE(0, ABSOLUTE)
    PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=cses | SSTYPE(3)
  /METHOD=REML
  /PRINT=G  SOLUTION TESTCOV
  /RANDOM=INTERCEPT cses | SUBJECT(school) COVTYPE(UN).
```

Zur manuellen Erstellung einer äquivalenten Syntax ist durch Ausnutzung der passenden Voreinstellungen nur wenig Schreibarbeit erforderlich:

```
mixed mAch with cses
  /criteria=dfmethod(kenwardroger)
  /fixed = cses
  /random = intercept cses | subject(school) covtype(un)
  /print = g solution testcov.
```

Änderungen im Vergleich zur Syntax für das Basismodell (siehe Abschnitt 3.2.3):

- Das CRITERIA-Subkommando fordert die Kenward-Roger - Approximation für die Signifikanztests zu festen Effekten an.
- Per FIXED-Subkommando wird CSES in das Modell aufgenommen.
- Im RANDOM-Subkommando wird festgelegt, dass für die einzelnen Schulen ein zufälliger Ordinatenabschnitt *und* ein zufälliger CSES-Effekt zu schätzen ist.
- Außerdem wird im RANDOM-Subkommando mit COVTYPE(UN) für die (von SPSS mit **G** bezeichnete) Kovarianzmatrix der zufälligen Effekte festgelegt, dass die drei Kovarianzparameter τ_{00}, τ_{11} und τ_{01} frei aus den Daten geschätzt werden sollen.
- Im PRINT-Subkommando wird zusätzlich die Ausgabe der **G**-Matrix angefordert.

[1] Von der SPSS-Version 26.0.0.0 (ohne FixPack 1 zu SPSS 26) erhält man per **Einfügen**-Schalter ein fehlerhaftes RANDOM-Subkommando mit ID statt UN als Kovarianztyp:

```
/RANDOM=INTERCEPT cses | SUBJECT(school) COVTYPE(ID).
```

Weil damit die Kovarianz der beiden Zufallseffekte aus dem Modell verschwindet, muss RANDOM-Subkommando unbedingt korrigiert werden. In SPSS 26.0.0.1 (*mit* FixPack 1) ist der Fehler behoben.

3.4.4 Diskussion der Ergebnisse

In der Modellspezifikation

Modelldimension[a]

		Anzahl Ausprägungen	Kovarianzstruktur	Anzahl Parameter	Subjektvariablen
Feste Effekte	Konstanter Term	1		1	
	cses	1		1	
Zufallseffekte	Konstanter Term + cses [b]	2	Unstrukturiert	3	school
Residuum				1	
Gesamt		4		6	

a. Abhängige Variable: mAch.

b. Ab Version 11.5 gelten andere Syntaxregeln für den Unterbefehl RANDOM. Die eingegebene Befehlssyntax führt zu abweichenden Ergebnissen im Vergleich zu früheren Versionen. Wenn Sie Version-11-Syntax verwenden, ziehen Sie das aktuelle Syntax-Referenzhandbuch zu Rate.

tauchen sechs Parameter auf:

- Das Globalmittel γ_{00} der Mikroebenen-Ordinatenabschnitte, also der MACH-Schulmittelwerte
- Das Globalmittel γ_{10} der schulinternen Steigungskoeffizienten aus der Regression von MACH auf CSES
- Die Varianzkomponente τ_{00} (Variabilität der MACH-Schulmittelwerte)
- Die Varianzkomponente τ_{11} (Variabilität der schulspezifischen Steigungskoeffizienten)
- Die Kovarianz τ_{01} der schulspezifischen Mittelwerte und Steigungskoeffizienten
- Die Varianzkomponente σ^2 (Varianz der Mikroresiduen)

Aufgrund der bei CSES eingesetzten schulinternen Zentrierung steht γ_{00} für das Globalmittel der MACH-Schulmittelwerte. Folglich liegt der Schätzwert im Vergleich zu den früheren Modellen praktisch unverändert bei 12,64 (siehe z. B. Abschnitt 3.2.4.3):

Schätzungen fester Parameter[a]

						Konfidenzintervall 95%	
Parameter	Schätzung	Standardfehler	Freiheitsgrade	T-Statistik	Sig.	Untergrenze	Obergrenze
Konstanter Term	12,636193	,244508	156,753	51,680	,000	12,153238	13,119148
cses	2,193196	,128412	155,217	17,079	,000	1,939536	2,446857

a. Abhängige Variable: mAch.

Schulintern geht ein CSES-Unterschied von einer Einheit durchschnittlich mit einem MACH-Unterschied von 2,19 Einheiten einher. Die Signifikanztests zu den beiden festen Effekten lehnen jeweils ihre (zweiseitige!) Nullhypothese ab.

In der Tabelle mit den geschätzten Kovarianzparametern

Schätzungen von Kovarianzparametern[a]

						Konfidenzintervall 95%	
Parameter		Schätzung	Std.-Fehler	Wald Z	Sig.	Untergrenze	Obergrenze
Residuum		36,700200	,625744	58,650	,000	35,494030	37,947358
Konstanter Term + cses [Subjekt = school]	UN (1,1)	8,680969	1,079535	8,041	,000	6,803240	11,076962
	UN (2,1)	,046792	,406372	,115	,908	-,749683	,843267
	UN (2,2)	,693989	,280786	2,472	,013	,314021	1,533721

a. Abhängige Variable: mAch.

zeigt sich:

- Für die Varianz der Schulmittelwerte ergibt sich der im Vergleich zum Basismodell (vgl. Abschnitt 3.2.1) praktisch unveränderte Schätzwert 8,681. Den Schätzwert $\hat{\tau}_{00}$ bezeichnet SPSS mit `UN(1,1)`. Das war zu erwarten, weil der einbezogenen Mikroregressor gruppenintern zentriert ist, also für alle Gruppen den Mittelwert null hat, sodass keine Varianz zwischen den Gruppen besteht. Ein Makroregressor ist aktuell nicht im Modell vorhanden.
- Für die Varianz der schulinternen Steigungskoeffizienten erhalten wir den Schätzwert 0,694. Den Schätzwert $\hat{\tau}_{11}$ bezeichnet SPSS mit `UN(2,2)`.
- Für Kovarianz der Mittelwerts- und Steigungskoeffizienten erhalten wir den Schätzwert 0,047. Den Schätzwert $\hat{\tau}_{10}$ bezeichnet SPSS mit `UN(2,1)`.
- Während die beiden Varianzparameter in den Wald-Tests als signifikant beurteilt werden ($p < 0{,}001$ bzw. $p = 0{,}013$), kann für den Kovarianzparameter die Nullhypothese nicht abgelehnt werden ($p = 0{,}901$). Daraus lässt sich die Vermutung ableiten, dass der schulspezifische CSES-Effekt auf MACH nicht vom durchschnittlichen Mathematik-Niveau einer Schule abhängt.

SPSS teilt für die Tests zu den Kovarianzparametern τ_{00}, τ_{11} und τ_{01} generell zweiseitige *p*-Werte mit, obwohl bei τ_{00} und τ_{11} einseitige Test angemessen sind mit Alternativhypothesen, die eine positive Varianz behaupten (siehe z. B. Hox 2010, S. 47; Tabachnick & Fidell 2007, S. 798). So verfährt z. B. das Statistikprogramm SAS. Vergleiche der Ergebnisse von SPSS und SAS bei identischen Daten und Modellen zeigen, dass sich die von SAS berichteten einseitigen *p*-Werte durch Halbieren aus den von SPSS berichteten zweiseitigen *p*-Werten ergeben (Tabachnick & Fidell 2007, S. 803). Bei der Kovarianz τ_{01} ist der von SPSS durchgeführte zweiseitige Test angemessen, wenn keine gerichtete Hypothese bestand. War eine Richtungshypothese vorhanden, sollte auch der Test zu τ_{01} einseitig durchgeführt werden, wobei der zweiseitige *p*-Wert zu halbieren ist, um die korrekte Überschreitungswahrscheinlichkeit zu ermitteln.

Zur Beurteilung der Varianzen und Kovarianzen von Zufallseffekten sind die Abschnitt 4.2 beschriebenen Likelihood-Quotienten - Tests gegenüber den Wald-Tests zu bevorzugen. Im Beispiel sorgt allerdings der große Stichprobenumfang für verlässliche Wald-Tests.

Aufgrund unserer Zusatzanforderung liefert SPSS noch eine übersichtliche Tabelle mit der geschätzten Kovarianzmatrix **G** der zufälligen Effekte:

Kovarianzstruktur mit zufälligen Effekten (G)[a]

	Konstanter Term | school	cses | school
Konstanter Term | school	8,680969	,046792
cses | school	,046792	,693989

Unstrukturiert

a. Abhängige Variable: mAch.

Für die schulinterne Residualvarianz nach Kontrolle der (mit einem zufälligen Effekt ausgestatteten) Variablen CSES erhalten wir den Schätzwert 36,70. Zusammen mit dem Wert 39,15 des Basismodells (vgl. Abschnitt 3.2.4.4) lässt sich der CSES-Beitrag zur Aufklärung der schulinternen Varianz berechnen:

$$\hat{R}_1^2 = \frac{39{,}148 - 36{,}700}{39{,}148} = 0{,}063$$

Ein Mikroregressor kann einen Beitrag zur Erklärung der Makroebenenvarianz τ_{00} leisten, wenn sich die Gruppen bei diesem Regressor bedeutsam unterscheiden. Beim gruppenintern zentrierten Regressor CSES bestehen definitionsgemäß keine Gruppenunterschiede, und die durch Vergleich mit dem Basismodell in Abschnitt 3.2.4.4 geschätzte Varianzaufklärung auf der Makroebene wird sogar leicht negativ:

$$R_{\beta_0}^2 = \frac{8{,}614 - 8{,}681}{8{,}614} = -0{,}008$$

Hox (2010, S. 74ff) beschäftigt sich ausführlich mit der Frage, unter welchen Umständen es bei einer Mehrebenenanalyse zu negativen Varianzanteilsschätzungen kommen kann.

3.5 Modelle mit Regressoren auf der Makro- und der Mikroebene

Nun kommen wir zu einem voll entwickelten hierarchischen linearen Modell mit Regressoren auf der Makro- und der Mikroebene. In Anlehnung an Norušis (2008, S. 218ff), Raudenbush & Bryk (2002, S. 80ff) und Singer (1998, S. 336ff) nehmen wir im HS&B - Beispiel nicht nur die bisher diskutierten Regressoren (MEANSES auf der Makro- und CSES auf der Mikroebene) auf, sondern ergänzen mit dem Schultyp (Variable SECTOR, siehe Seite 15) einen weiteren Makroregressor. Außerdem werden Einflüsse der Makroregressoren auf die schulbedingten Effekte von CSES in das Modell zugelassen. Wir nehmen also die Cross-Level - Wechselwirkungen von CSES mit MEANSES und SECTOR in das Modell auf.

Wesentliche Forschungsfragestellungen sind:

- Haben MEANSES und SECTOR einen Einfluss auf das Mathematikleistungsniveau der Schulen?
- Moderieren MEANSES und SECTOR den schulinternen CSES-Effekt auf MACH?

Die beim gruppeninternen Zentrieren von SES verloren gegangene Information über den mittleren sozioökonomischen Status der Schulen ist über den Makroregressor MEANSES nun wieder im Modell enthalten. Nach dem gruppeninternen Zentrieren eines Mikroregressors sollte in der Regel ergänzend ein Makroregressor mit den Gruppenmittelwerten einbezogen werden (Field 2013, S. 830; Tabachnick & Fidell 2007, S. 824). Eine perfekte Kompensation gelingt nur dann, wenn für den gruppenintern zentrierten Mikroregressor ein fester, für alle Makroeinheiten identischer Parameter angenommen wird (Kreft & de Leeuw 1998, S. 109). In unserem Beispiel ist allerdings für CSES ein zufälliger, von der Makroeinheit abhängiger Steigungskoeffizient vorgesehen.

3.5.1 Modell

Das Mikromodell enthält wie in der vorherigen Modelvariante den schulintern zentrierten Regressor CSES ($X_{ij} - \overline{X}_{.j}$, vgl. Abschnitt 3.4.1):

$$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \overline{X}_{.j}) + R_{ij}\text{, } i = 1, .. , n_j \text{ und } j = 1, .., J$$
$$R_{ij} \sim \text{unabhängig } \mathrm{N}(0, \sigma^2)$$

Im Makroebenenmodell werden der MACH-Schulmittelwert β_{0j} und die schulinterne Steigung β_{1j} der Regression von MACH auf CSES folgendermaßen durch MEANSES (W_1) und SECTOR (W_2) erklärt:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_{1j} + \gamma_{02}W_{2j} + U_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_{1j} + \gamma_{12}W_{2j} + U_{1j}$$

Für U_{0j} und U_{1j} wird eine bivariate Normalverteilung angenommen:

$$\begin{pmatrix} U_{0j} \\ U_{1j} \end{pmatrix} \sim \text{unabhängig } \mathrm{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}\right]$$

Die bivariate Zufallsvariable (U_{0j}, U_{1j}) ist von den Makroresiduen aus anderen Gruppen ebenso unabhängig wie von den Mikroresiduen R_{ij}.

Um die Erklärungsleistung der Makroregressoren in Bezug auf die β_{0j}-Varianz zu beurteilen, setzt man die Differenz zwischen der Varianz $\tau_{00|B}$ aus dem *random coefficient* - Modell (siehe Abschnitt 3.4) und der Residualvarianz $\tau_{00|M}$ aus dem Modell mit Makroregressoren ins Verhältnis zum Ausgangswert $\tau_{00|B}$.

Diese mit dem Determinationskoeffizienten der multiplen Regression vergleichbare Statistik soll (wie in Abschnitten 3.3.1 und 3.4.2) mit $R^2_{\beta_0}$ bezeichnet werden:

$$R^2_{\beta_0} = \frac{\tau_{00|B} - \tau_{00|M}}{\tau_{00|B}}$$

Um die Erklärungsleistung der Makroregressoren in Bezug auf die β_{1j}-Varianz zu beurteilen, setzt man die Differenz zwischen der Varianz $\tau_{11|B}$ aus dem *random coefficient* - Modell (siehe Abschnitt 3.4) und der Residualvarianz $\tau_{11|M}$ aus dem Modell mit Makroregressoren ins Verhältnis zum Ausgangswert $\tau_{11|B}$. Diese mit dem Determinationskoeffizienten der multiplen Regression vergleichbare Statistik soll mit $R^2_{\beta_1}$ bezeichnet werden:

$$R^2_{\beta_1} = \frac{\tau_{11|B} - \tau_{11|M}}{\tau_{11|B}}$$

Durch Einsetzen des Makromodells in das Mikromodell erhält man das kombinierte Modell mit sechs festen Effekten ($\gamma_{00}, \gamma_{01}, \gamma_{02}, \gamma_{10}, \gamma_{11}, \gamma_{12}$), zwei zufälligen Effekten (U_{0j}, U_{1j}) und dem Mikroebenenresiduum R_{ij}:

$$Y_{ij} = \gamma_{00} + \gamma_{01}W_{1j} + \gamma_{02}W_{2j} + U_{0j} + (\gamma_{10} + \gamma_{11}W_{1j} + \gamma_{12}W_{2j} + U_{1j})(X_{ij} - \bar{X}_{.j}) + R_{ij}$$

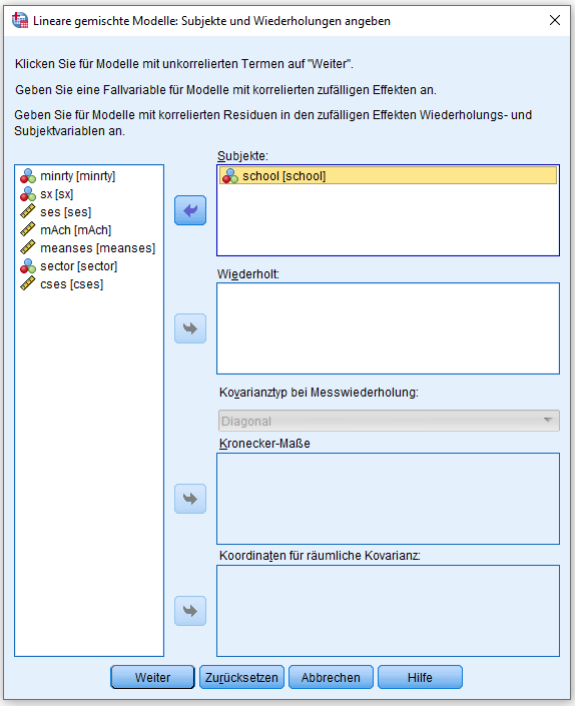
$$= \gamma_{00} + \gamma_{01}W_{1j} + \gamma_{02}W_{2j} + \gamma_{10}(X_{ij} - \bar{X}_{.j}) + \gamma_{11}W_{1j}(X_{ij} - \bar{X}_{.j}) + \gamma_{12}W_{2j}(X_{ij} - \bar{X}_{.j}) + U_{0j} + U_{1j}(X_{ij} - \bar{X}_{.j}) + R_{ij}$$

Das kombinierte Modell ist weniger übersichtlich als die nach Ebenen getrennte Formulierung.

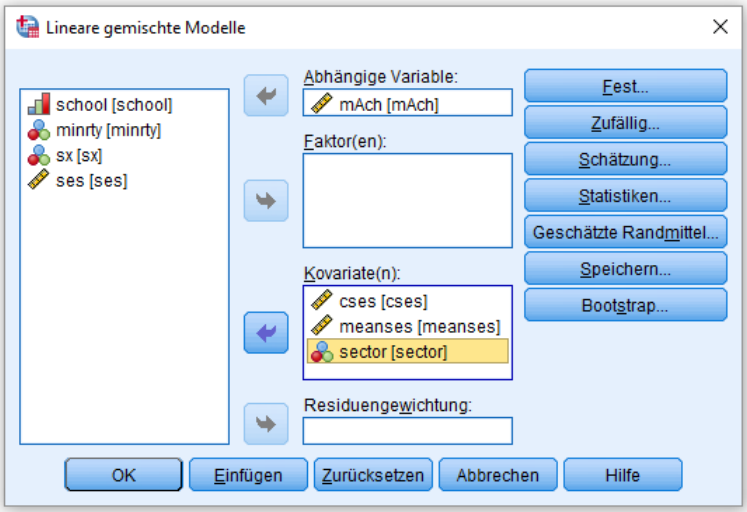
3.5.2 Anforderung der Analyse in SPSS

Gehen Sie folgendermaßen vor, um in SPSS Statistics das Modell mit Regressoren auf Makro- und Mikroebene auf die HS&B - Daten anzuwenden:

- Menübefehl **Analysieren > Gemischte Modelle > Linear**
- Wir legen fest, dass die Variable SCHOOL für die spätere Definition der Makroeinheiten in Frage kommt:

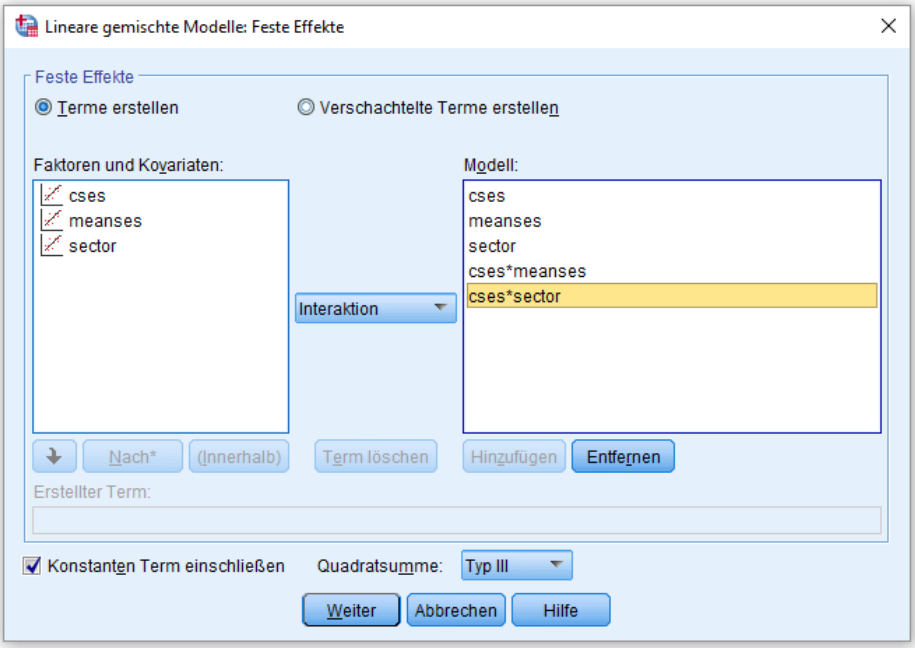


- Wir machen **weiter** und legen MACH als **abhängige Variable** sowie CSES, MEANSES und SECTOR als **Kovariaten** fest:



Weil SECTOR nur zwei Ausprägungen hat, ist die Behandlung als Kovariate zulässig und äquivalent zur Behandlung als **Faktor**.

- Wir spezifizieren in der Subdialogbox für die **festen Effekte** das oben formulierte kombinierte Modell:

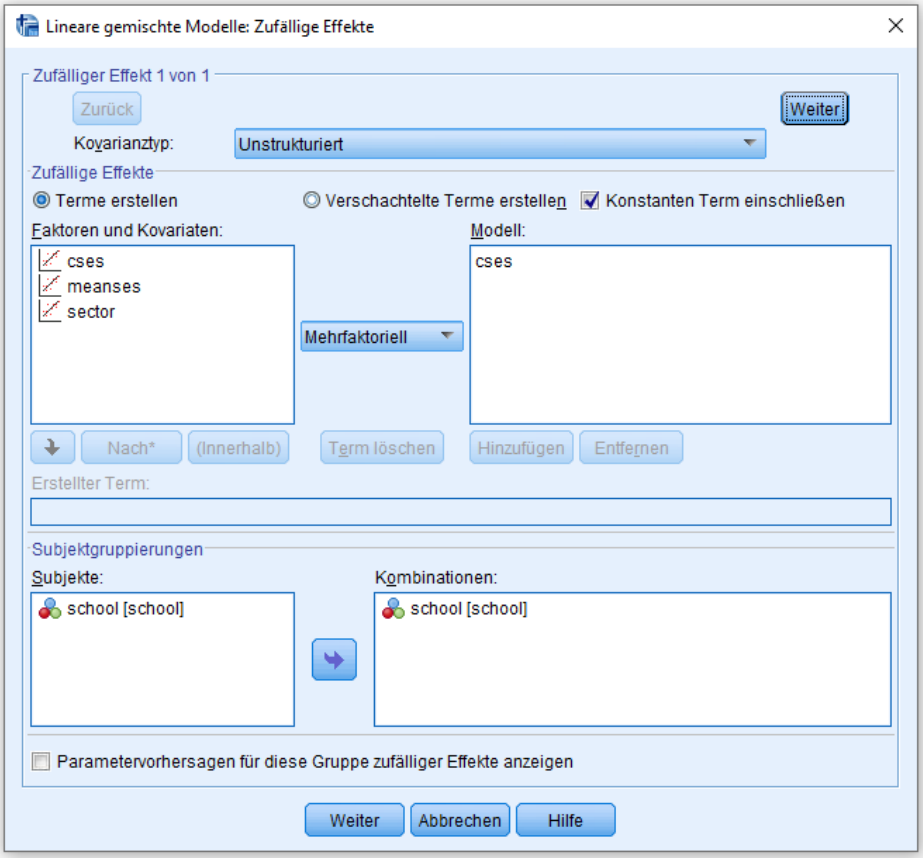


Zu den Spezifikationen:

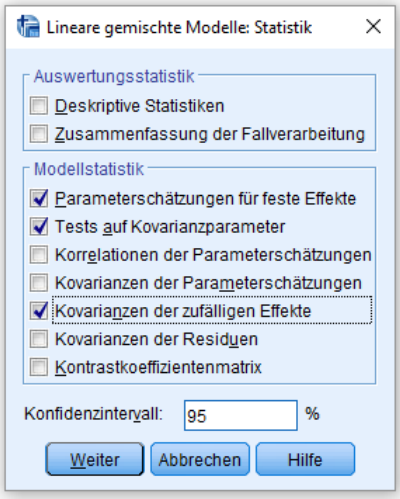
- Das Kontrollkästchen für den **konstanten Term** ist markiert (= Voreinstellung), weil der Parameter γ_{00} (Ordinatenabschnitt im Makromodell für β_{0j}) nicht auf null gesetzt, sondern frei schätzbar sein soll.
- Die **Modell**-Bestandteile CSES, MEANSES und SECTOR sollten schon deswegen vorhanden sein, weil diese Regressoren an einer Wechselwirkung beteiligt sind. Wer komplexe Überlegungen vermeiden will, sollte generell mit *hierarchisch wohlgeformten Modellen* arbeiten, die zu jedem Interaktionsterm auch alle Komponenten von niedrigerer Ordnung enthalten (Cohen et al. 2003, S. 284).

- Der **Modell**-Bestandteil CSES*MEANSES ist vorhanden, weil der Parameter γ_{11} (Steigungskoeffizient zu MEANSES im Makromodell für β_{1j}) nicht auf null gesetzt, sondern frei schätzbar sein soll. Wir erlauben also eine Moderation des gruppeninternen CSES-Effekts durch den Makroregressor MEANSES (eine Cross-Level - Interaktion).
- Der **Modell**-Bestandteil CSES*SECTOR ist vorhanden, weil der Parameter γ_{12} (Steigungskoeffizient zu SECTOR im Makromodell für β_{1j}) nicht auf null gesetzt, sondern frei schätzbar sein soll. Wir erlauben also eine Moderation des gruppeninternen CSES-Effekts durch den Makroregressor SECTOR (eine Cross-Level - Interaktion).

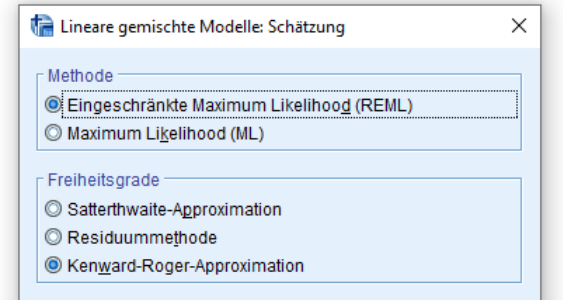
- In der Subdialogbox für **zufällige Effekte** und in der **Statistik**-Subdialogbox können wir die Einstellungen aus Abschnitt 3.4.3 übernehmen:



- Wir fordern in der **Statistik**-Subdialogbox **Parameterschätzer für feste Effekte**, **Tests auf Kovarianzparameter** und **Kovarianzen der zufälligen Effekte** an:



- Schließlich fordern wir im Subdialog **Schätzung** die **Kenward-Roger-Approximation** an, um die Präzision der Signifikanztests zu den festen Effekten zu verbessern, wobei in einer großen Stichprobe allerdings keine bedeutsamen Unterschiede im Vergleich zur voreingestellten Satterthwaite-Approximation zu erwarten sind:



Über den **Einfügen**-Schalter der Hauptdialogbox erhält man folgendes Syntax-Äquivalent zu der Modellspezifikation:[1]

```
MIXED mAch WITH cses meanses sector
  /CRITERIA=DFMETHOD(KENWARDROGER) CIN(95) MXITER(100) MXSTEP(10) SCORING(1)
    SINGULAR(0.000000000001) HCONVERGE(0, ABSOLUTE) LCONVERGE(0, ABSOLUTE)
    PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=cses meanses sector cses*meanses cses*sector | SSTYPE(3)
  /METHOD=REML
  /PRINT=G SOLUTION TESTCOV
  /RANDOM=INTERCEPT cses | SUBJECT(school) COVTYPE(UN).
```

Zur manuellen Erstellung einer äquivalenten Syntax ist durch Ausnutzung der passenden Voreinstellungen nur wenig Schreibarbeit erforderlich:

```
mixed mAch with meanses sector cses
 /criteria=dfmethod(kenwardroger)
 /fixed = cses meanses sector cses*meanses cses*sector
 /random = intercept cses | subject(school) covtype(un)
 /print = g solution testcov.
```

3.5.3 Diskussion der Ergebnisse

In der Modellspezifikation

[1] Von der SPSS-Version 26.0.0.0 (ohne FixPack 1 zu SPSS 26) erhält man per **Einfügen**-Schalter ein fehlerhaftes RANDOM-Subkommando mit ID statt UN als Kovarianztyp:

```
/RANDOM=INTERCEPT cses | SUBJECT(school) COVTYPE(ID).
```

Weil damit die Kovarianz der beiden Zufallseffekte aus dem Modell verschwindet, muss RANDOM-Subkommando unbedingt korrigiert werden. In SPSS 26.0.0.1 (*mit* FixPack 1) ist der Fehler behoben.

Modelldimension[a]

		Anzahl Ausprägungen	Kovarianzstruktur	Anzahl Parameter	Subjektvariablen
Feste Effekte	Konstanter Term	1		1	
	cses	1		1	
	meanses	1		1	
	sector	1		1	
	cses * meanses	1		1	
	cses * sector	1		1	
Zufallseffekte	Konstanter Term + cses [b]	2	Unstrukturiert	3	school
Residuum				1	
Gesamt		8		10	

a. Abhängige Variable: mAch.

b. Ab Version 11.5 gelten andere Syntaxregeln für den Unterbefehl RANDOM. Die eingegebene Befehlssyntax führt zu abweichenden Ergebnissen im Vergleich zu früheren Versionen. Wenn Sie Version-11-Syntax verwenden, ziehen Sie das aktuelle Syntax-Referenzhandbuch zu Rate.

tauchen zehn Parameter auf:

- Sechs Parameter für die festen Effekte ($\gamma_{00}, \gamma_{01}, \gamma_{02}, \gamma_{10}, \gamma_{11}, \gamma_{12}$)
- Drei Parameter in der Kovarianzmatrix mit den beiden Zufallseffekten (τ_{00}, τ_{11} und τ_{01})
- Die Residualvarianz σ^2

Die festen Effekte werden in der Tabelle mit den **Schätzungen fester Parameter** alle als signifikant beurteilt:

Schätzungen fester Parameter[a]

Parameter	Schätzung	Standard Fehler	Freiheitsgrade	T-Statistik	Sig.	Konfidenzintervall 95% Untergrenze	Obergrenze
Konstanter Term	12,127931	,199314	159,896	60,848	,000	11,734304	12,521558
cses	2,945041	,156029	139,619	18,875	,000	2,636555	3,253526
meanses	5,332874	,369204	150,986	14,444	,000	4,603402	6,062347
sector	1,226579	,306298	149,613	4,005	,000	,621349	1,831808
meanses * cses	1,039232	,299462	160,570	3,470	,001	,447840	1,630624
sector * cses	-1,642674	,240222	143,349	-6,838	,000	-2,117509	-1,167839

a. Abhängige Variable: mAch.

Beide Makroregressoren (MEANSES und SECTOR) moderieren den schulinternen CSES-Effekt auf MACH, so dass bei der Interpretation einige Sorgfalt angebracht ist.[1] Der berichtete „Haupteffekt" von CSES ($\hat{\gamma}_{10}$ = 2,945; $p < 0{,}001$) ist z. B. tatsächlich ein *bedingter* Effekt in öffentlichen Schulen (SECTOR = 0) mit einem mittleren sozioökonomischen Status (MEANSES = 0).

Gegeben den MEANSES-Mittelwert von null finden sich in Abhängigkeit vom Schultyp folgende Koeffizienten für die (bedingten) Regressionen von MACH auf CSES:

SECTOR	MACH-Mittelwert	CSES-Effekte auf MACH
0 (öffentliche Schulen)	12,128	2,945
1 (katholische Schulen)	13,355 = 12,128 + 1,227	1,302 = 2,945 - 1,643

In katholischen Schulen ist also das mittlere Mathematikniveau höher ($\hat{\gamma}_{02}$ = 1,227; $p < 0{,}001$) und der schulinterne Statuseffekt auf die Mathematikleistung geringer ($\hat{\gamma}_{12}$ = -1,643; $p < 0{,}001$).

[1] Bei der Interpretation und Illustration von Interaktionen kann man genauso vorgehen wie bei Moderatoranalysen in Einebenenmodellen (siehe z. B. Baltes-Götz 2018).

Mit dem mittleren sozioökonomischen Status einer Schule wächst ihr mittleres Mathematikniveau ($\hat{\gamma}_{01}$ = 5,333; $p < 0{,}001$) und der interne CSES-Effekt auf MACH ($\hat{\gamma}_{11}$ = 1,039; p = 0,001).

Aus der Tabelle mit den Varianzen und Kovarianzen der zufälligen Effekte

Schätzungen von Kovarianzparametern[a]

Parameter		Schätzung	Std.-Fehler	Wald Z	Sig.	Konfidenzintervall 95% Untergrenze	Obergrenze
Residuum		36,721155	,626134	58,647	,000	35,514236	37,969091
Konstanter Term + cses [Subjekt = school]	UN (1,1)	2,379579	,371456	6,406	,000	1,752372	3,231275
	UN (2,1)	,191826	,204357	,939	,348	-,208706	,592359
	UN (2,2)	,101216	,213792	,473	,636	,001612	6,356042

a. Abhängige Variable: mAch.

ergibt sich durch Vergleich mit dem *random coefficient* - Modell (siehe Abschnitt 3.4) für die Varianz τ_{00} (von SPSS mit `UN(1,1)` bezeichnet) ein beachtlicher erklärter Anteil:

$$\hat{R}^2_{\beta_0} = \frac{8{,}681 - 2{,}380}{8{,}681} = 0{,}726$$

Aber auch nach Kontrolle von MEANSES und SECTOR bestehen noch signifikante Unterschiede zwischen den Schulmittelwerten, so dass weitere Forschungsbemühungen angebracht sind.

Bei der Varianz τ_{11} der schulinternen Steigungen (von SPSS mit `UN(2,2)` bezeichnet) kann der Ausgangswert 0,694 aus dem *random coefficient* - Modell (siehe Abschnitt 3.4) so weit reduziert werden,

$$\hat{R}^2_{\beta_1} = \frac{0{,}694 - 0{,}101}{0{,}694} = 0{,}854$$

dass die verbleibende Varianz *nicht* mehr signifikant größer null ist.

Erwartungsgemäß besitzt der varianzarme Zufallseffekt für die schulinternen Steigungen auch keine bedeutsame Kovarianz mit dem Zufallseffekt für die Schulmittelwerte.

3.6 Modellannahmen prüfen

Für ein hierarchisches lineares Modell ist eine stattliche Liste von Voraussetzungen erforderlich:

- Die Makroresiduen bzw. Zufallseffekte (U_{0j}, U_{1j}, ..., U_{Qj}) der Makroeinheiten sind unabhängig voneinander.
- Alle Makroresiduen haben den Erwartungswert 0, d.h. vom Modell wird für alle Makroregressorkombinationen für den jeweils modellierten Mikrokoeffizienten der korrekte bedingte Erwartungswert behauptet.
- Die Vektoren (U_{0j}, U_{1j}, ..., U_{Qj}) mit den Makroresiduen sind gemeinsam multivariat normalverteilt mit einer für alle Makroeinheiten identischen Kovarianzmatrix (Normalität und Homoskedastizität auf der Makroebene).
- Die Mikroresiduen R_{ij} sind unabhängig voneinander.
- Alle Mikroresiduen haben den Erwartungswert von 0, d.h. vom Modell wird für alle Regressorwertekombinationen der korrekte bedingte Erwartungswert für das Kriterium behauptet.
- Die Mikroresiduen R_{ij} sind normalverteilt mit der konstanten Varianz σ^2 (Normalität und Homoskedastizität auf der Mikroebene).
- Die Makroresiduen U_{qj} sind unabhängig von den Mikroresiduen R_{ij}.

Dieser Abschnitt erläutert die Prüfung der HLM-Modellvoraussetzungen, wobei leider im Wesentlichen nur der Spezialfall eines hierarchischen Modells mit random intercepts als einzigem Zufallseffekt behandelt wird.

3.6.1 Vereinfachung des HS&B - Modells

Viele Mehrebenenmodelle enthalten nur zufällige Ordinatenabschnitte (random intercepts), aber keine zufälligen Steigungen (random slopes). Bei dieser Modellklasse ist die Untersuchung der Ergebnisse (insbesondere der Residuen) auf Modellspezifikationsfehler besonders einfach. Wir profitieren von dieser Bequemlichkeit, weil sich in Abschnitt 3.5.3 für das HS&B - Modell mit Mikro- und Makroregressoren sowie Cross-Level - Interaktionen gezeigt hat, dass die schulspezifischen CSES-Regressionskoeffizienten keine signifikante Varianz mehr besitzen (von SPSS in der folgenden Tabelle mit `UN(2,2)` bezeichnet):

Schätzungen von Kovarianzparametern[a]

Parameter		Schätzung	Std.-Fehler	Wald Z	Sig.	Konfidenzintervall 95% Untergrenze	Obergrenze
Residuum		36,721155	,626134	58,647	,000	35,514236	37,969091
Konstanter Term + cses [Subjekt = school]	UN (1,1)	2,379579	,371456	6,406	,000	1,752372	3,231275
	UN (2,1)	,191826	,204357	,939	,348	-,208706	,592359
	UN (2,2)	,101216	,213792	,473	,636	,001612	6,356042

a. Abhängige Variable: mAch.

Daher entfernen wir den random slope zu CSES aus dem Modell und fordern mit dem folgenden MIXED-Kommando eine neue Schätzung an:

```
mixed mAch with meanses sector cses
 /criteria=dfmethod(kenwardroger)
 /fixed = cses meanses sector cses*meanses cses*sector
 /random = intercept | subject(school) covtype(vc) solution
 /print = solution testcov
 /save=fixpred (fixpred) pred (pred) resid (resid).
```

Im Vergleich zum MIXED-Kommando in Abschnitt 3.5.2 ...

- wurde im Subkommando RANDOM
  - CSES aus der Effektliste entfernt,
  - der Kovarianztyp UN (unstrukturiert) ersetzt durch VC (Varianzkomponenten), weil nur noch *ein* Zufallseffekt vorhanden ist,
  - durch das Schlüsselwort SOLUTION die Ausgabe von geschätzte Makroresiduen im Ausgabefenster angefordert (siehe Abschnitt 3.6.2),
- wurde das Subkommando SAVE ergänzt, um Mikro- und Makroresiduen in die Arbeitsdatei zu schreiben (siehe Abschnitt 3.6.2).

Die Schätzwerte zu den festen Parametern ändern sich im Vergleich zum Modell mit random slopes nur geringfügig (vgl. Abschnitt 3.5.3):

Schätzungen fester Parameter[a]

Parameter	Schätzung	Standardfehler	Freiheitsgrade	T-Statistik	Sig.	Konfidenzintervall 95% Untergrenze	Obergrenze
Konstanter Term	12,128207	,199217	160,006	60,879	,000	11,734774	12,521640
cses	2,942146	,151209	7018,261	19,457	,000	2,645730	3,238561
meanses	5,336670	,369014	151,072	14,462	,000	4,607576	6,065764
sector	1,224529	,306140	149,695	4,000	,000	,619616	1,829442
meanses * cses	1,044438	,291048	7018,261	3,589	,000	,473896	1,614981
sector * cses	-1,642156	,233094	7018,261	-7,045	,000	-2,099090	-1,185222

a. Abhängige Variable: mAch.

3.6.2 Mikro- und Makroresiduen abspeichern

Um die Mikro- und Makroresiduen zu erhalten, lassen wir von der Prozedur MIXED über das Subkommando SAVE

```
/SAVE=FIXPRED (fixpred) PRED (pred) RESID (resid).
```

die folgenden neuen Variablen in die Arbeitsdatei schreiben:

- FIXPRED Modellprognose ohne random intercept
- PRED Modellprognose inklusive random intercept
- RESID Mikroresiduum (geschätzter R_{ij} - Wert)

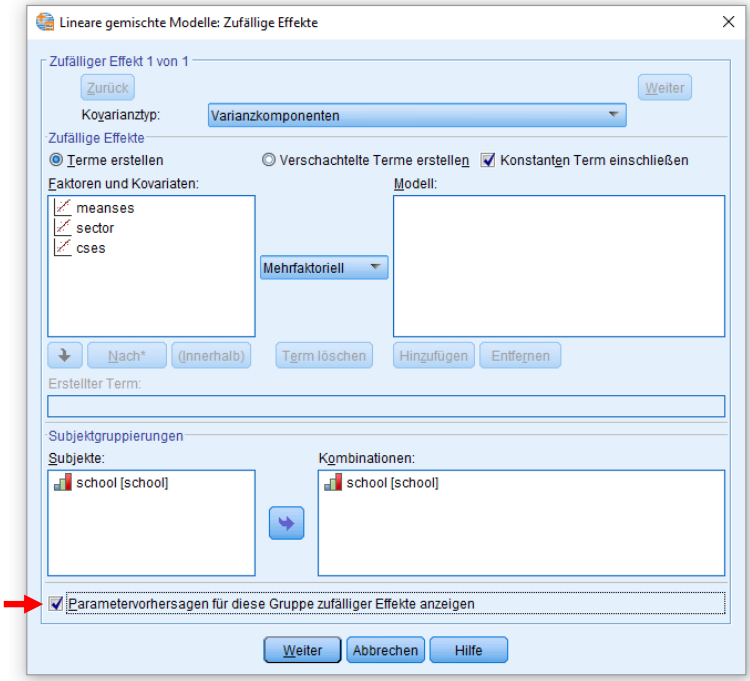
Optional kann man in runden Klammern die gewünschten Namen für die neuen Variablen festlegen.

Aus den Variablen FIXPRED und PRED kann man geschätzte Makroresiduen (die U_{0j} - Werte) ausrechnen, für die auch die Bezeichnung BLUP (*Best Linear Unbiased Prediction*) verwendet wird:

```
compute blup = pred - fixpred.
```

Diese Berechnungsmethode ist nur anwendbar unter der Voraussetzung, dass im Modell die zufälligen Ordinatenabschnitte (random intercepts) der einzige Zufallseffekt sind.

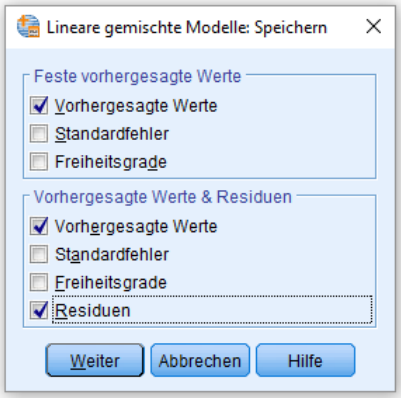
Wer die MIXED-Analyse per Dialogbox anfordert (vgl. Abschnitt 3.5.2), kann im Subdialog



für die zufälligen Effekte ...

- den **Kovarianztyp** ändern,
- die random slopes für CSES entfernen,
- über das Kontrollkästchen **Parametervorhersagen für diese Gruppe zufälliger Effekte anzeigen** dafür sorgen, dass die geschätzten U_{0j} - Werte für die 160 Schulen in der *Ausgabe* (im Viewer-Fenster) erscheinen.

Im **Speichern**-Subdialog lassen sich neue Variablen mit den oben beschriebenen Modellprognosen und Residuen anfordern, wobei aber im Unterschied zur Syntaxverwendung keine Möglichkeit besteht, die Variablennamen festzulegen:



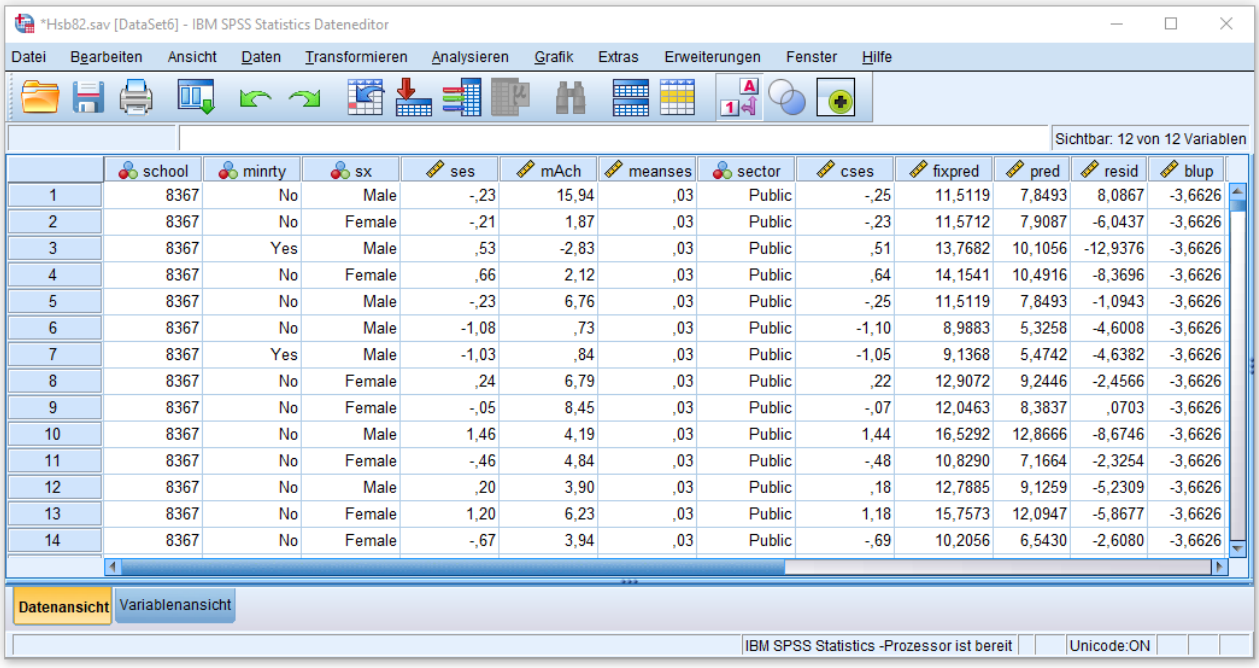
Ein Vergleich der Ausgabetabelle mit den BLUP-Werten (hier beschränkt auf die ersten 8 Schulen)

Empirische beste lineare unverzerrte Vorhersagen [a]

school	Parameter	Vorhersage	Std.-Fehler	df	Sig.	95%-Konfidenzintervall Untergrenze	95%-Konfidenzintervall Obergrenze
8367		-3,6626	1	287	0	-6	-1
8854		-2,5948	1	563	0	-4	-1
4458		-,5416	1	587	1	-2	1
5762		-1,0094	1	467	0	-3	1
6990		-2,7383	1	992	0	-4	-1
5815		-,7584	1	466	0	-3	1
7172		-2,7662	1	770	0	-4	-1
4868		-2,0407	1	669	0	-4	0

a. Abhängige Variable: mAch

mit der Arbeitsdatei



	school	minrty	sx	ses	mAch	meanses	sector	cses	fixpred	pred	resid	blup
1	8367	No	Male	-,23	15,94	,03	Public	-,25	11,5119	7,8493	8,0867	-3,6626
2	8367	No	Female	-,21	1,87	,03	Public	-,23	11,5712	7,9087	-6,0437	-3,6626
3	8367	Yes	Male	,53	-2,83	,03	Public	,51	13,7682	10,1056	-12,9376	-3,6626
4	8367	No	Female	,66	2,12	,03	Public	,64	14,1541	10,4916	-8,3696	-3,6626
5	8367	No	Male	-,23	6,76	,03	Public	-,25	11,5119	7,8493	-1,0943	-3,6626
6	8367	No	Male	-1,08	,73	,03	Public	-1,10	8,9883	5,3258	-4,6008	-3,6626
7	8367	Yes	Male	-1,03	,84	,03	Public	-1,05	9,1368	5,4742	-4,6382	-3,6626
8	8367	No	Female	,24	6,79	,03	Public	,22	12,9072	9,2446	-2,4566	-3,6626
9	8367	No	Female	-,05	8,45	,03	Public	-,07	12,0463	8,3837	,0703	-3,6626
10	8367	No	Male	1,46	4,19	,03	Public	1,44	16,5292	12,8666	-8,6746	-3,6626
11	8367	No	Female	-,46	4,84	,03	Public	-,48	10,8290	7,1664	-2,3254	-3,6626
12	8367	No	Male	,20	3,90	,03	Public	,18	12,7885	9,1259	-5,2309	-3,6626
13	8367	No	Female	1,20	6,23	,03	Public	1,18	15,7573	12,0947	-5,8677	-3,6626
14	8367	No	Female	-,67	3,94	,03	Public	-,69	10,2056	6,5430	-2,6080	-3,6626

bestätigt die Berechnungsmethode. Aus naheliegenden Gründen sind die Makroresiduen für alle Schüler in einer Makroeinheit identisch.

3.6.3 Schwächen der Modellspezifikation

In diesem Abschnitt werden Maßnahmen zur Modelldiagnose und -verbesserung beschrieben. Es besteht eine partielle Überlappung mit den in Abschnitt 4.4 präsentierten Überlegungen zur Modellierungsstrategie.

3.6.3.1 Schwächen der Modellspezifikation im Bereich der festen Effekte

Bei einem gültigen Modell sind die bedingten Erwartungswerte aller Residuen gleich 0. Folglich sind auf beiden Ebenen die Residuen zu analysieren, um Schwächen der Modellspezifikation im Bereich der festen Effekte aufzudecken.

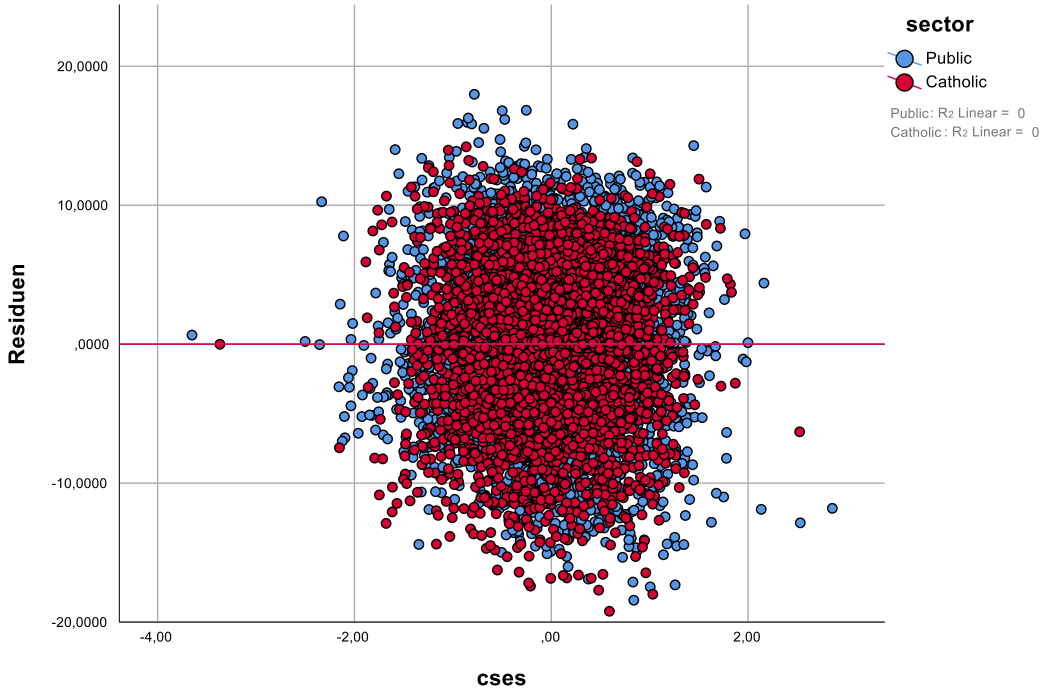
Sind in einem HLM nur die Ordinatenabschnitte zufällig, können auf folgende Weise Schwächen der Modellspezifikation auf der **Mikroebene** aufgedeckt werden:

- Mikroresiduen in eine Variable der Arbeitsdatei abspeichern lassen
- Zur Untersuchung dieser Variablen können die Techniken für lineare Modelle ohne Wechselwirkungen (siehe z. B. Baltes-Götz 2019) bzw. mit Wechselwirkungen (siehe z. B. Baltes-Götz 2018) angewendet werden.

In einem Modell ohne Wechselwirkungen zwischen Mikroregressoren kann man ...

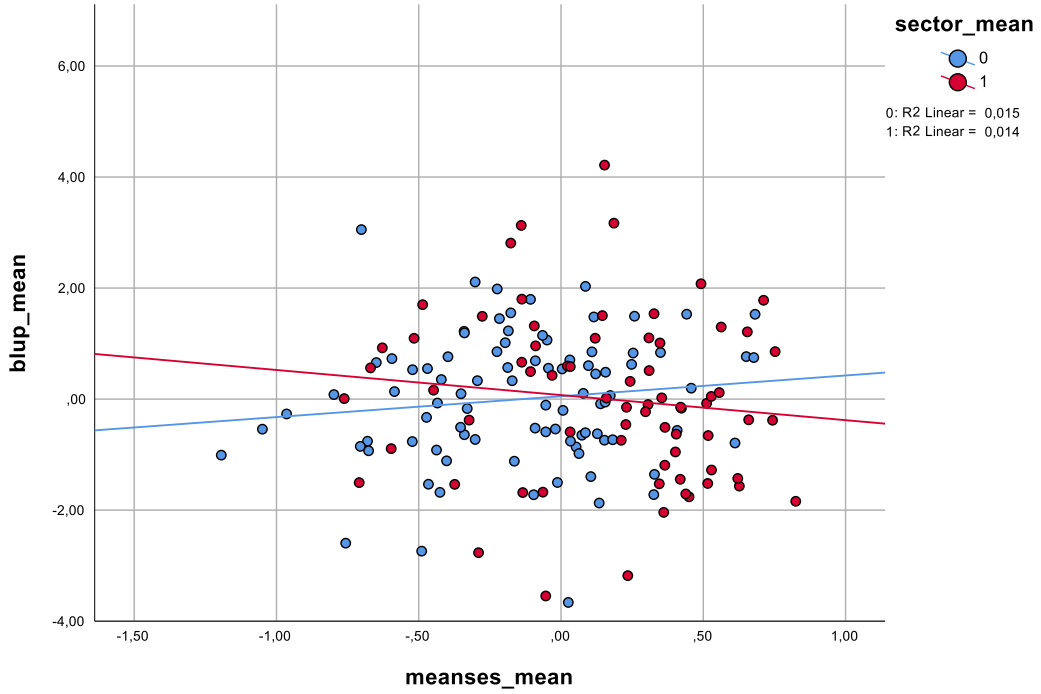
- die Mikroresiduen als Kriterium einer linearen Regressionsanalyse verwenden und dabei die Mikroregressoren einbeziehen
- partielle Regressionsdiagramme anfordern

Im aktuellen HS&B - Modell gibt es nur einen Mikroregressor (CSES). Weil der Makroregressor SECTOR nur zwei Ausprägungen hat, soll er in die Darstellung einbezogen werden, um eine differenziertere Beurteilung zu ermöglichen. Für beide Schultypen zeigt sich keine Abweichung von der Linearität:



Um Modellschwächen auf der **Makroebene** (im Modell für β_{0j}) aufzudecken, rechnet man mit den aggregierten Daten (mit den Schulen als Fällen, zum AGGREAGATE-Kommando siehe Abschnitt 3.4.1.1.2) eine lineare Regression mit der BLUP-Variablen als Kriterium und den Makroregressoren im Design. In einem Modell ohne Wechselwirkungen helfen die partiellen Regressionsdiagramme bei der Beurteilung der Linearität.

Im aktuellen HS&B - Modell gibt es einen metrischen (MEANSES) und einen dichotom-kategorialen Makroregressor (SECTOR). Ein gruppiertes Streudiagramm kann die Regression der Makroresiduen (U_{0j} - bzw. BLUP-Werte) auf die beiden Makroregressoren gut beschreiben:



Für beide Schultypen hängen die Makroresiduen modellgemäß nicht von MEANSES ab, so dass die Linearität bestätigt ist.

Insbesondere ist keine Wechselwirkung zwischen SECTOR und MEANSES zu erkennen, sodass es nicht erforderlich erscheint, einen entsprechenden Term in das Modell aufzunehmen. Tut man es trotzdem, wird diese Wechselwirkung nicht als signifikant beurteilt:[1]

Schätzungen fester Parameter[a]

Parameter	Schätzung	Standardfehler	Freiheitsgrade	T-Statistik	Sig.	Konfidenzintervall 95% Untergrenze	Konfidenzintervall 95% Obergrenze
Konstanter Term	12,201643	,203920	160,075	59,836	,000	11,798923	12,604363
meanses	5,868156	,506155	162,006	11,594	,000	4,868643	6,867669
sector	1,244028	,304733	147,556	4,082	,000	,641824	1,846232
meanses * sector	-1,121994	,734980	148,166	-1,527	,129	-2,574391	,330403
cses	2,942146	,151213	7017,640	19,457	,000	2,645722	3,238569
meanses * cses	1,044438	,291056	7017,640	3,588	,000	,473881	1,614996
sector * cses	-1,642156	,233100	7017,640	-7,045	,000	-2,099103	-1,185210

a. Abhängige Variable: mAch.

Zur Beseitigung von Schwächen der Modellspezifikation im Bereich der festen Effekte kommen folgende Maßnahmen in Frage:

[1] Verwendete Syntax für das Modell mit (MEANSES × SECTOR) - Wechselwirkung:

```
mixed mAch with meanses sector cses
 /criteria=dfmethod(kenwardroger)
 /fixed = meanses sector meanses*sector cses cses*meanses cses*sector
 /random intercept  | subject(school) covtype(vc)
 /print = g solution  testcov.
```

- Aufnahme von weiteren Regressoren
  Bei HLM-Modellen besteht eine wichtige Option darin, Within- und Between-Effekte von Regressoren mit Varianz auf der Mikro- und der Makroebene unterscheidbar zu machen (vgl. Abschnitt 2.2).
- Aufnahme von Wechselwirkungen (innerhalb einer Ebene und zwischen den Ebenen)
- Transformation des Kriteriums und/oder Transformation von Regressoren

Wenn die bedingten Erwartungswerte aller Residuen gleich null sind, ist in Beobachtungsstudien nur eine *notwendige* Voraussetzung für die *kausale* Interpretation der Effektschätzungen erfüllt. In dieser Hinsicht unterscheiden sich hierarchische lineare Modelle nicht von einfachen linearen Modellen (mit nur *einer* Ebene).

3.6.3.2 Schwächen der Modellspezifikation im Bereich der zufälligen Effekte

Um Schwächen der Modellspezifikation im Bereich der zufälligen Effekte aufzudecken, sollte nach einer Empfehlung von Snijders & Bosker (2012, S. 155f) für jeden Mikroregressor *X* von zentralem Interesse versuchsweise ein random slope in das Modell aufgenommen werden. Ein Signifikanztest zur random slope - Varianz ist geeignet, die Vernachlässigung der Steigungsheterogenität im Modell zu verhindern. Eine Beachtung dieser Empfehlung ist besonders wichtig, wenn der Mikroregressor *X* an einer Cross-Level-Interaktion mit dem Makroregressor *Z* beteiligt ist, weil ein fehlender random slope eine Unterschätzung des Standardfehlers zur Cross-Level-Interaktion bewirken kann.

Hinter der als statistisch bedeutsam erscheinenden random slope - Varianz eines Mikroregressors kann aber auch eine Fehlspezifikation im Bereich der festen Effekte stecken, z. B.:

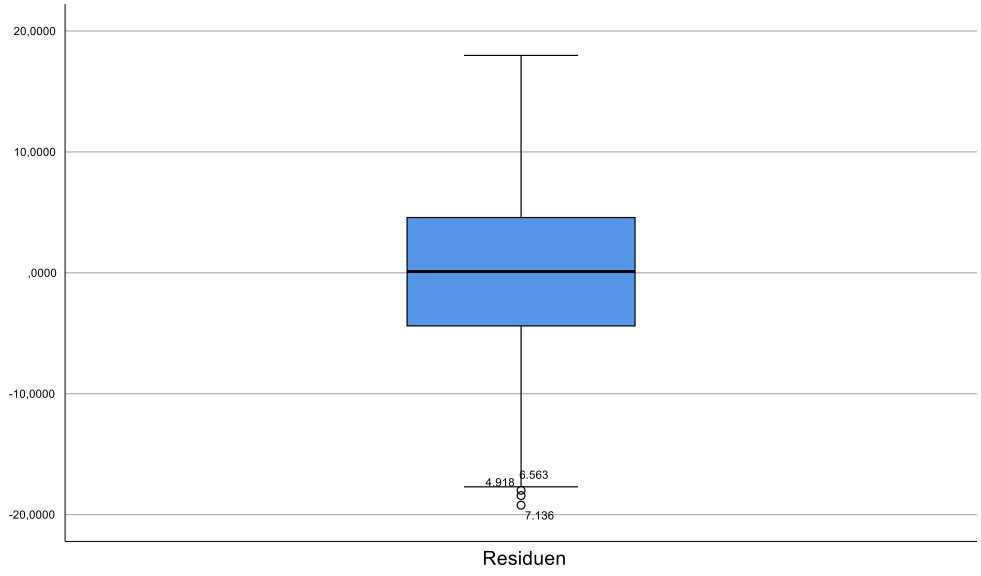
- Es wurde eine Wechselwirkung unter Beteiligung des Mikroregressors vergessen.
- Der Mikroregressor hat einen kurvilinearen Effekt, während im Modell ein linearer Effekt angenommen wird.

Wir haben für die HS&B - Daten nach der Erweiterung des Modells um die Cross-Level-Interaktionen CSES × MEANSES und CSES × SECTOR festgestellt, dass für den einzigen Mikroregressor CSES kein random slope mehr erforderlich ist (siehe Abschnitt 3.6.1).

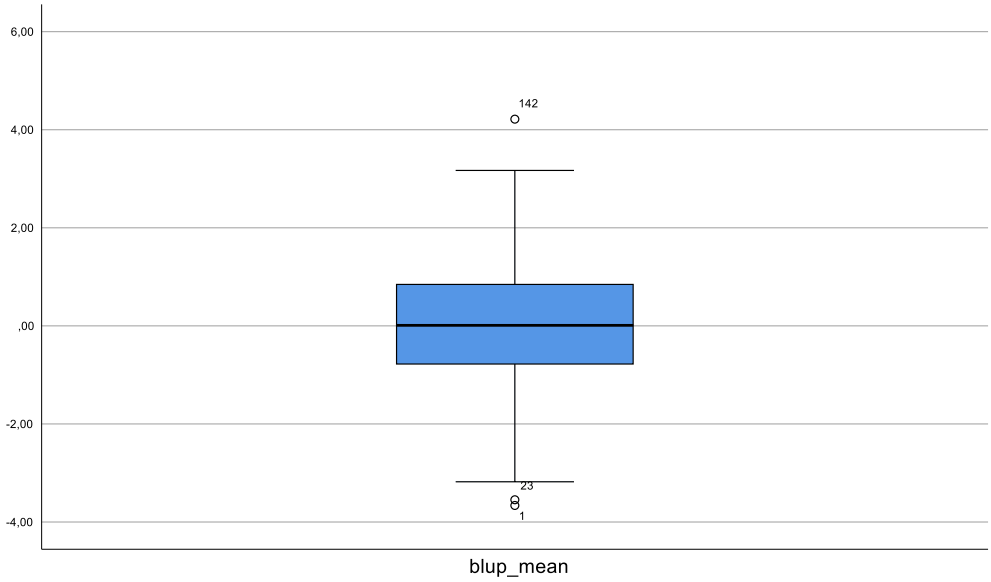
3.6.4 Ausreißer

Zur Identifikation von Ausreißern kann man z. B. die Mikro- und die Makroresiduen per Boxplot analysieren. Im HS&B - Beispiel zeigt sich dabei kein kritischer Befund.

Boxplot der Mikroresiduen:



Boxplot der Makroresiduen:



3.6.5 Homoskedastizität

Zur Prüfung der Homoskedastizität kann ein Plot der Mikroresiduen gegen die Modellprognosen beitragen.

Sind in einem HLM nur die Ordinatenabschnitte zufällig, dann können folgende Signifikanztests auf Heteroskedastizität durchgeführt werden:

- Heteroskedastizität der Mikroresiduen
  Wenn die Anzahl der Makroeinheiten nicht zu groß ist, kann mit der SPSS-Prozedur UNIAVOVA ein lineares Modell (Fixed-Effekts - Analyse) mit ...
  - den Mikroresiduen als Kriterium
  - der Clustervariablen als festem Faktor
  - den Mikroregressoren

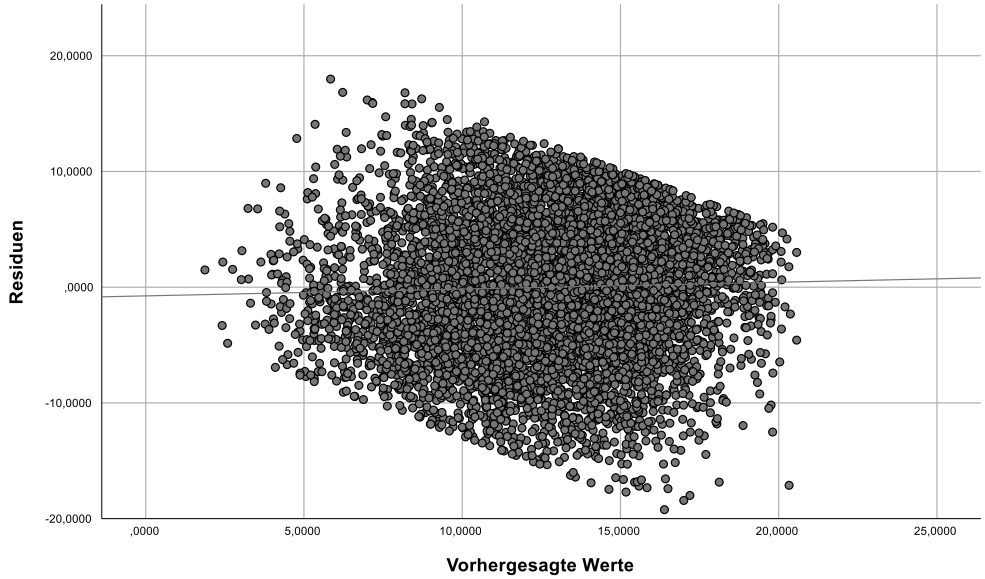
  geschätzt und der modifizierte Breusch-Pagan - Test auf Varianzheterogenität durchgeführt werden. Makroregressoren können und müssen in diesem Modell nicht berücksichtigt werden.
- Heteroskedastizität der Makroresiduen
  Nach einer Aggregation bzgl. der Cluster-Variablen kann mit der SPSS-Prozedur UNIAVOVA ein lineares Modell mit ...
  - den Makroresiduen als Kriterium
  - den Makroregressoren im Modell

  geschätzt und der modifizierte Breusch-Pagan - Test auf Varianzheterogenität durchgeführt werden.

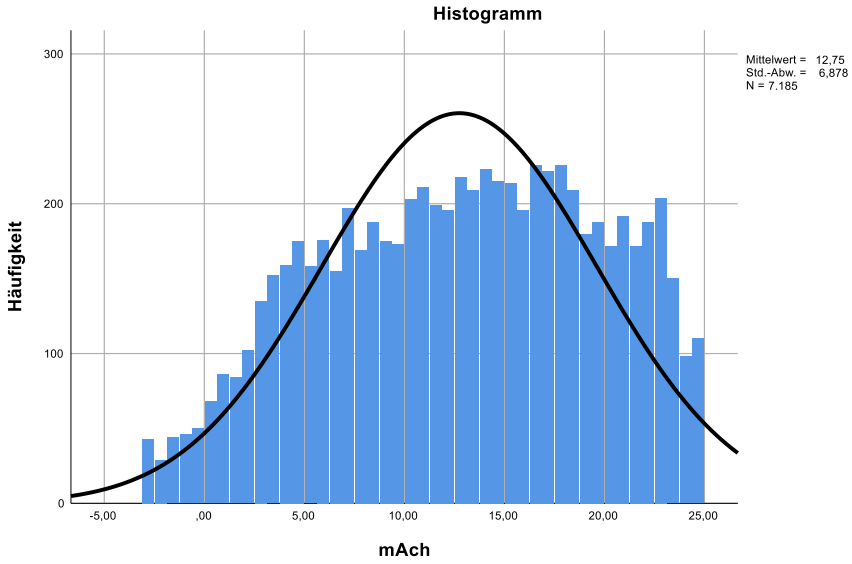
Eine Verletzung der Homoskedastizität kann ein Anlass zur Verbesserung des Modells sein:

- Erweiterung um zusätzliche Regressoren
- Ergänzung von Wechselwirkungen
- Ergänzung von Zufallseffekten

Im HSB&B - Beispiel zeigt der Plot der Mikroresiduen gegen die Modellprognosen ein für abhängige Variablen mit Deckeneffekt typisches Bild:



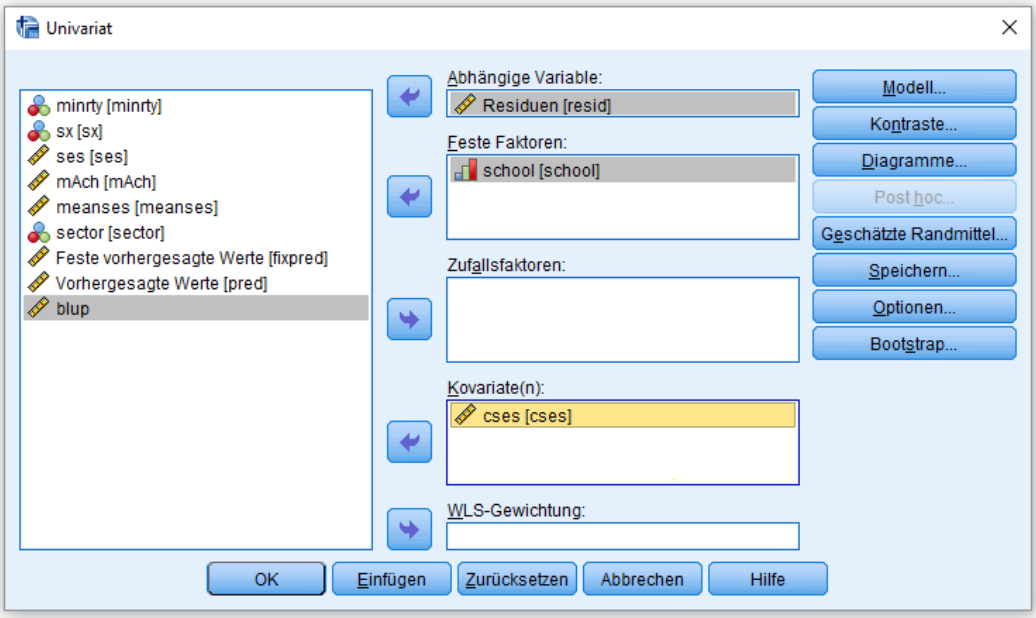
Hier ist die Randverteilung der Variablen MACH zu sehen:



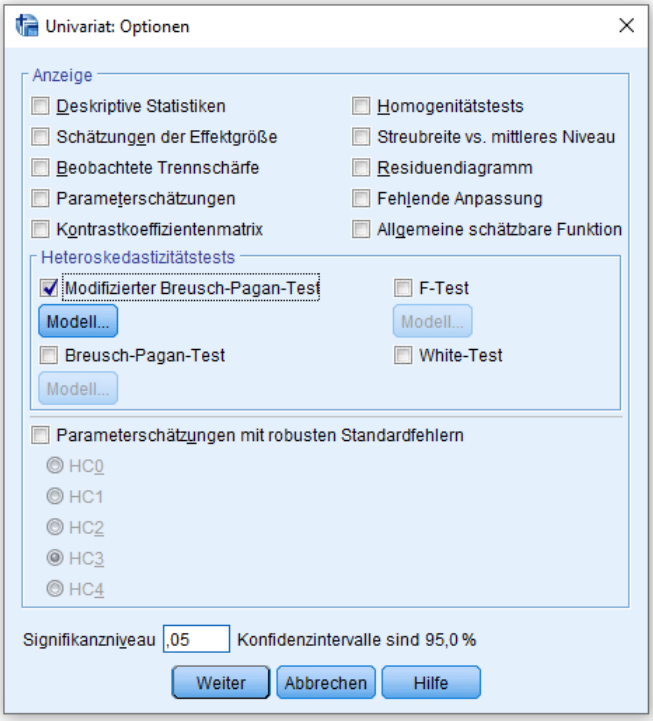
Die nach dem Menübefehl

Analysieren > Allgemeines lineares Modell > Univariat

im folgenden Dialog anzufordernde Fixed-Effekts - Analyse mit der SPSS-Prozedur UNIANOVA (vgl. Abschnitt 2.3)



bietet im **Optionen**-Dialog zur Prüfung der Homoskedastizität u.a. den modifizierten **Breusch-Pagan-Test** an:



Für die HS&B - Mikroresiduen wird die Homogenitäts-Nullhypothese akzeptiert:

Modifizierter Breusch-Pagan-Test auf Heteroskedastizität[a,b,c]

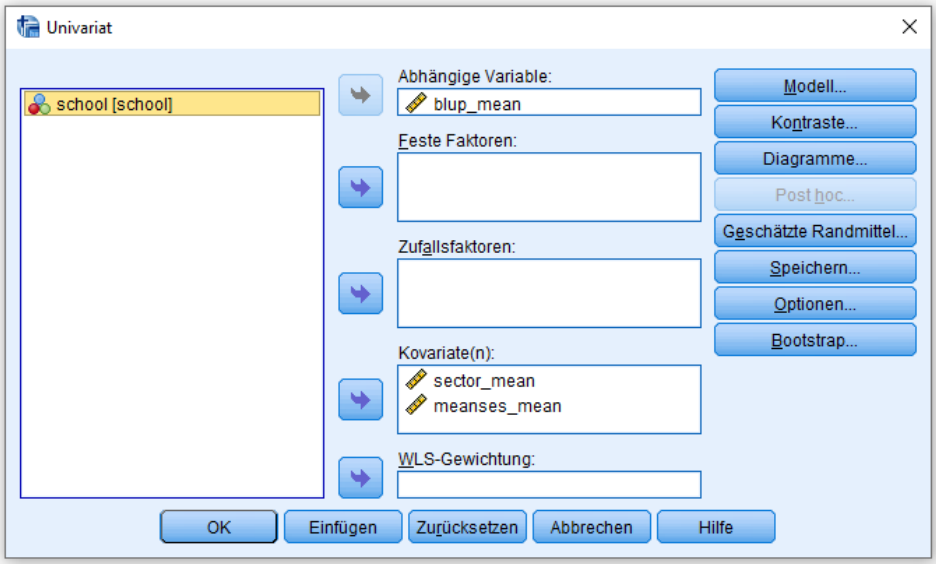
Chi-Quadrat	df	Sig.
2,061	1	,151

a. Abhängige Variable: Residuen

b. Testet die Nullhypothese, dass die Varianz der Fehler nicht von den Werte der unabhängigen Variablen abhängt.

c. Vorhergesagte Werte aus Design: Konstanter Term + cses + school

Lässt man für die aggregierten Daten (mit den Schulen als Fällen) von der SPSS-Prozedur UNIAVOVA eine Regression der Makroresiduen (BLUP-Werte) auf die Makroregressoren schätzen



und fordert **Optionen**-Dialog (siehe oben) den modifizierten Breusch-Pagan - Test an, dann wird die Nullhypothese der Homoskedastizität abgelehnt:

Modifizierter Breusch-Pagan-Test auf Heteroskedastizität[a,b,c]

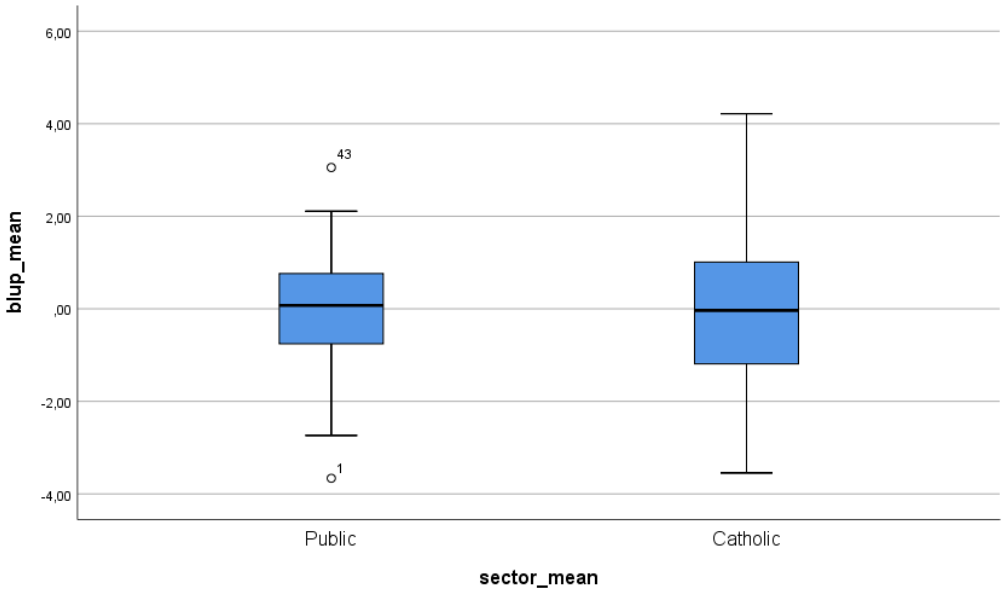
Chi-Quadrat	df	Sig.
4,704	1	,030

a. Abhängige Variable: blup_mean

b. Testet die Nullhypothese, dass die Varianz der Fehler nicht von den Werte der unabhängigen Variablen abhängt.

c. Vorhergesagte Werte aus Design: Konstanter Term + meanses_mean + sector_mean

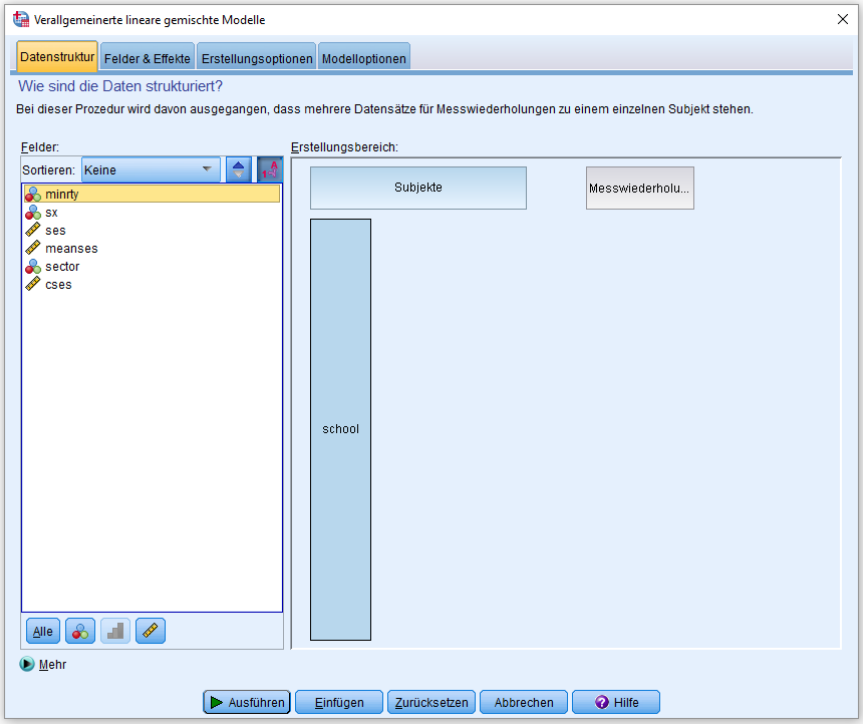
Für diese Heteroskedastizität ist vermutlich primär die höhere Variabilität der Makroresiduen im katholischen Sektor verantwortlich. In der folgenden Abbildung zeigen die BLUP-Boxplots zu den beiden Sektoren deutlich verschiedene Interquartilsbereiche:



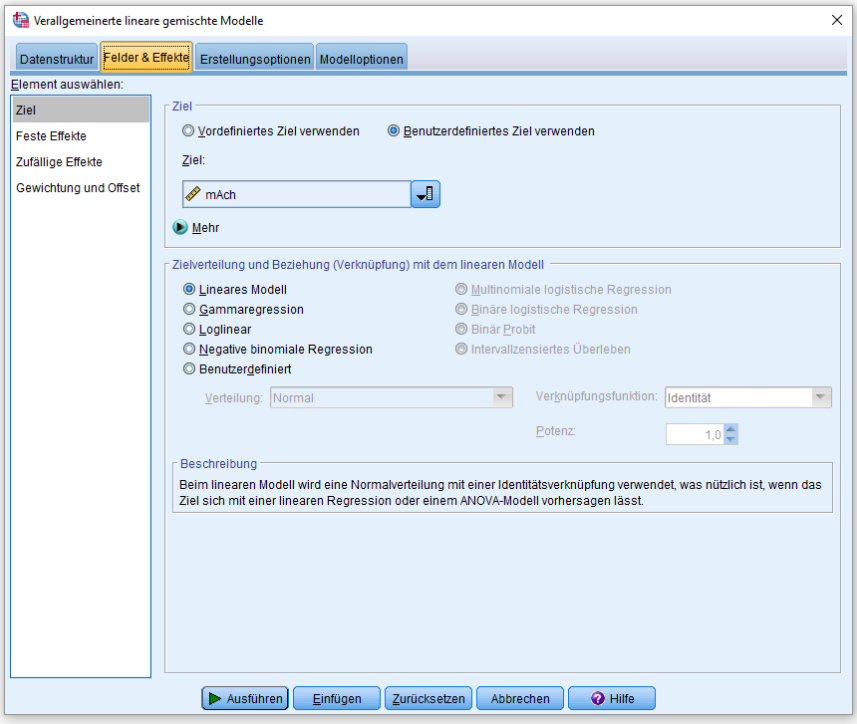
Wenn bei einer Mehrebenenanalyse die Beurteilung der Homoskedastizität ungünstig bzw. unsicher endet, können mit Hilfe der SPSS-Prozedur GENLINMIXED robuste Schätzer für die Standardfehler zu den Koeffizienten der festen Effekte berechnet werden, sodass die resultierenden Signifikanztests gegenüber einer verletzten Homoskedastizität unempfindlich werden. Dieses Verfahren soll nun auf die HS&B - Daten angewendete werden. Nach dem Menübefehl

Analysieren > Gemischte Modelle > Verallgemeinert linear

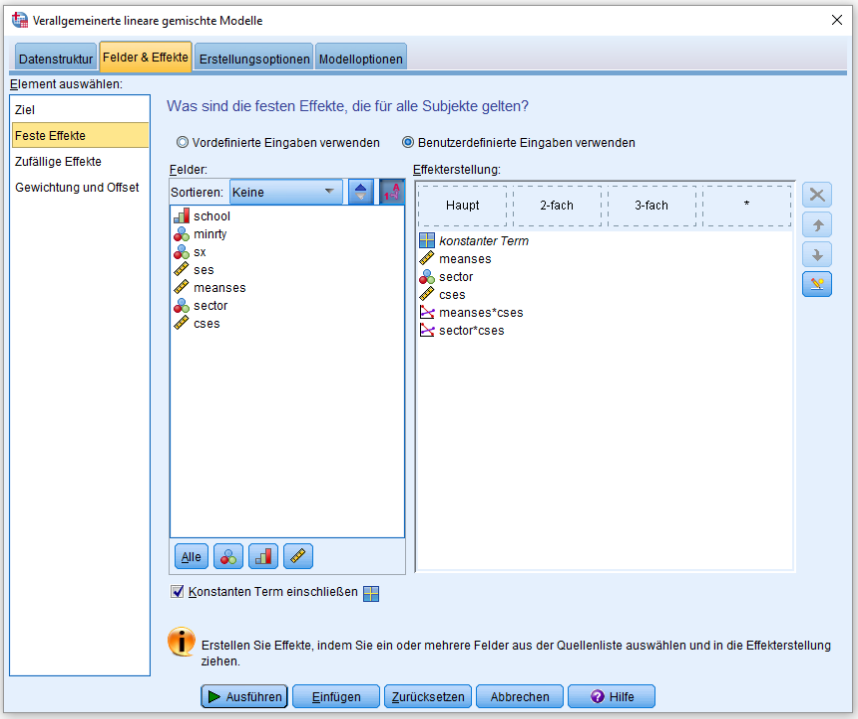
befördert man auf der Registerkarte **Datenstruktur** des folgenden Dialogs die Variable SCHOOL auf die Ablagezone **Subjekte**:



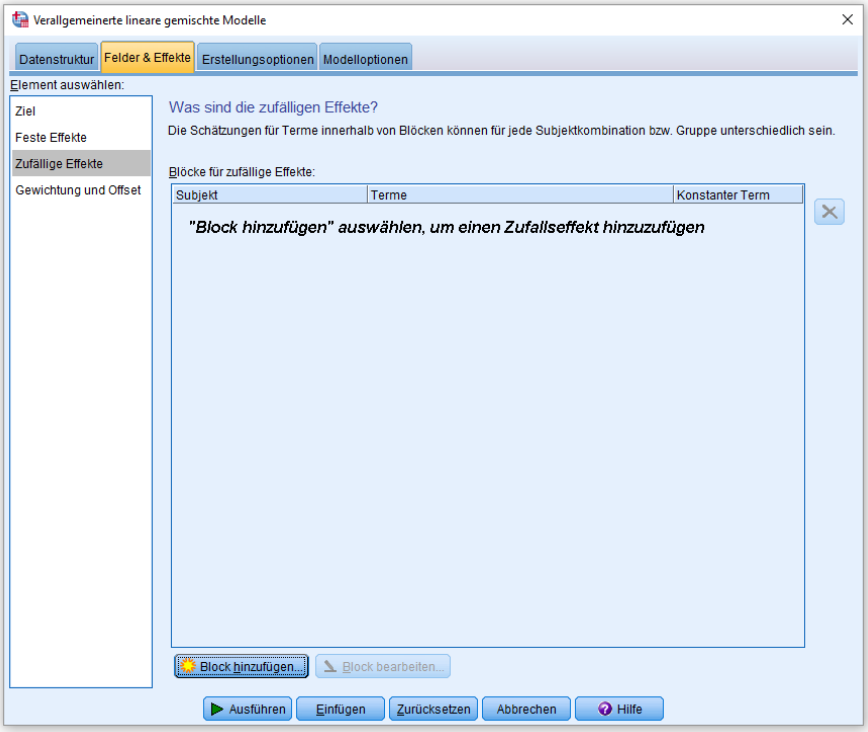
Auf der Registerkarte **Felder & Effekte** wählt man bei markiertem Element **Ziel** die Variable MACH als Kriterium (als **Ziel**-Variable) und akzeptiert das voreingestellte **lineare Modell**:



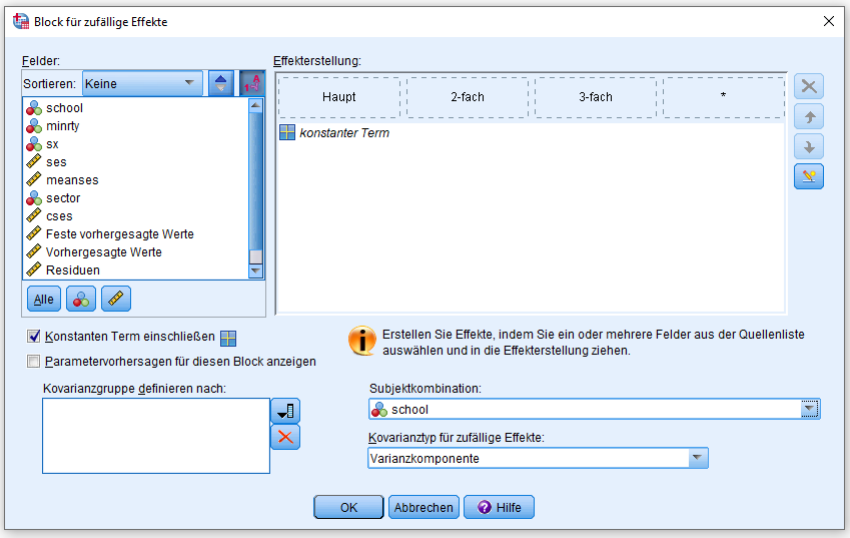
Auf derselben Registerkarte wählt man bei markiertem Element **Feste Effekte** die Haupt- und Wechselwirkungseffekte:



Auf derselben Registerkarte klickt man bei markiertem Element **Zufällige Effekte** auf den Schalter **Block hinzufügen**



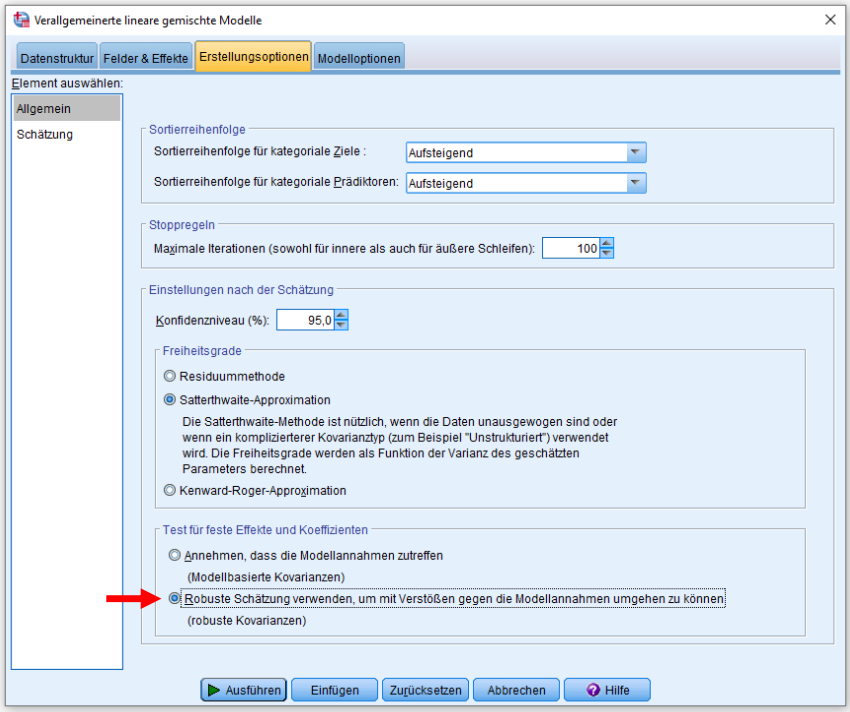
und kann dann im folgenden Dialog die zufälligen Ordinatenabschnitte in das Modell aufnehmen:



Dazu sind folgende Schritte erforderlich:

- Kontrollkästchen **Konstanten Term einschließen** markieren
- SCHOOL als **Subjektkombination** wählen
- Klick auf **OK**

Auf der Registerkarte **Erstellungsoptionen** wählt man bei markiertem Element **Allgemein** schließlich die **Robuste Schätzung**, die allerdings nicht zusammen mit der Kenward-Roger - Approximation für die Tests zu den festen Effekten zu haben ist:



Aus der beschriebenen Sequenz von Dialogen resultiert das folgende GENLINMIXED-Kommando:

```
GENLINMIXED
  /DATA_STRUCTURE SUBJECTS=school
  /FIELDS TARGET=mAch TRIALS=NONE OFFSET=NONE
  /TARGET_OPTIONS DISTRIBUTION=NORMAL LINK=IDENTITY
  /FIXED  EFFECTS=meanses sector cses meanses*cses sector*cses USE_INTERCEPT=TRUE
  /RANDOM USE_INTERCEPT=TRUE SUBJECTS=school COVARIANCE_TYPE=VARIANCE_COMPONENTS
   SOLUTION=FALSE
  /BUILD_OPTIONS TARGET_CATEGORY_ORDER=ASCENDING INPUTS_CATEGORY_ORDER=ASCENDING
   MAX_ITERATIONS=100 CONFIDENCE_LEVEL=95 DF_METHOD=SATTERTHWAITE COVB=ROBUST
   PCONVERGE=0.000001(ABSOLUTE) SCORING=0 SINGULAR=0.000000000001
  /EMMEANS_OPTIONS SCALE=ORIGINAL PADJUST=LSD.
```

Für die HS&B-Daten werden von GENLINMIXED in der Tabelle mit den **festen Koeffizienten** die MIXED-Ergebnisse weitgehend bestätigt:

Feste Koeffizienten [a]

Modellterm	Koeffizient	Standardfehler	T	Sig.	95% Konfidenzintervall	
					Unterer Wert	Oberer Wert
Konstanter Term	13,353	,2435	54,843	,000	12,873	13,833
meanses	5,337	,3345	15,952	,000	4,673	6,000
sector=0	-1,225	,3085	-3,969	,000	-1,834	-,615
sector=1	0 [b]	.	.	.	.	.
cses	1,300	,1732	7,506	,000	,960	1,640
meanses*cses	1,044	,3329	3,137	,002	,392	1,697
cses*[sector=0]	1,642	,2372	6,923	,000	1,177	2,107
cses*[sector=1]	0 [b]	.	.	.	.	.

Wahrscheinlichkeitsverteilung: Normal
Verknüpfungsfunktion: Identität

a. Ziel: mAch

b. Dieser Koeffizient wurde auf den Wert null gesetzt, da er redundant ist.

Für die scheinbar gravierenden Abweichungen der Koeffizienten zu SECTOR und CSES ist die von GENLINMIXED vorgenommene Kodierung des Faktors SECTOR verantwortlich. Die Wahl der robusten Schätzer hat einen deutlichen Einfluss auf die T-Prüfgrößen, wirkt sich aber nicht auf die Testentscheidungen aus.

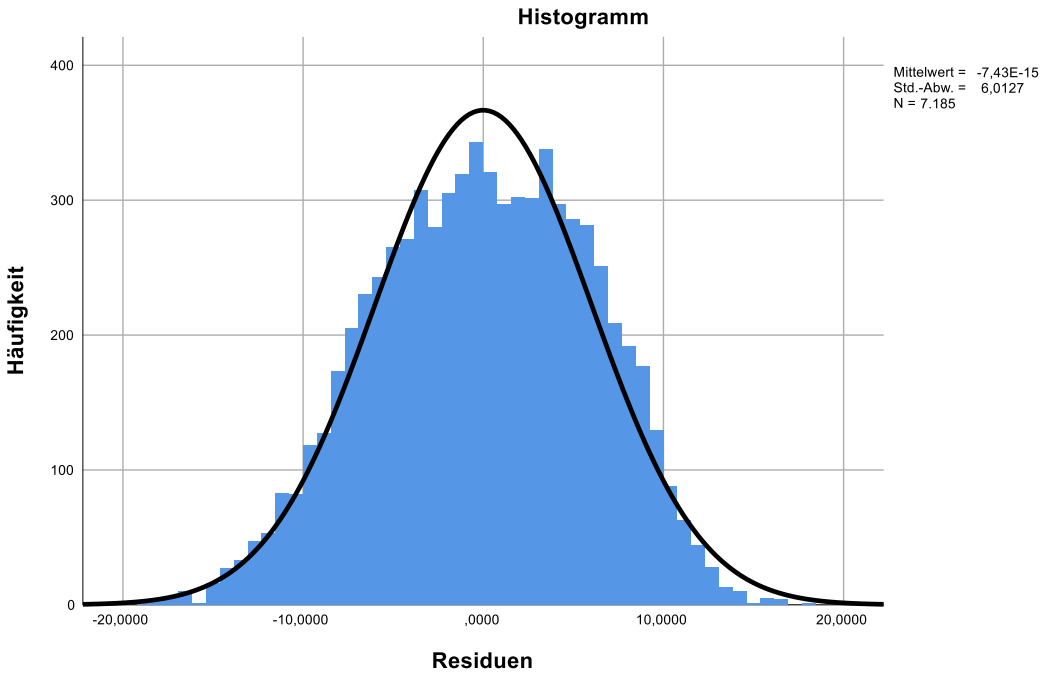
3.6.6 Normalität der Residuen

Zur Beurteilung der Normalität der Residuen auf der Mikro- und der Makroebene lässt man diese abspeichern (siehe Abschnitt 3.6.2) und untersucht sie z.B. mit Hilfe der Prozedur EXAMINE. Weil formale Signifikanztests (z. B. nach Shapiro & Wilk) stark von der Stichprobengröße abhängen, ist eine optische Prüfung oft angemessener.

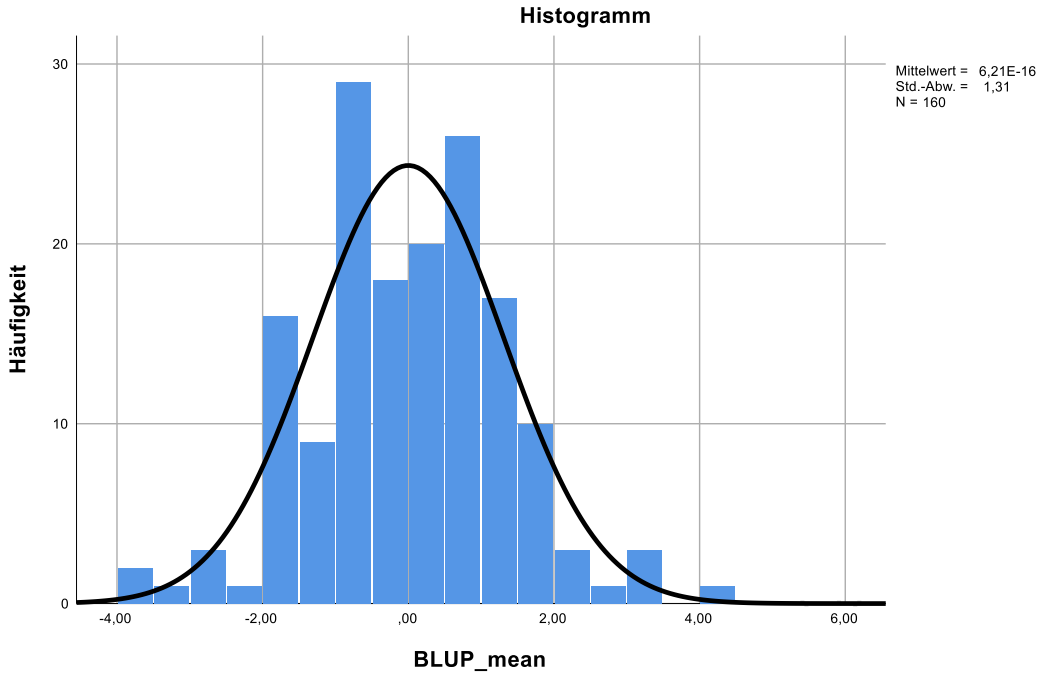
Bei einer groben Verletzung der Normalität der Residuen kommen als Maßnahmen in Frage:

- Transformation des Kriteriums
- Verwendung eines generalisierten linearen gemischten Modells, das keine normalverteilten Residuen annimmt (möglich über die SPSS-Prozedur GENLINMIXED)

Die Mikroresiduen zu den HS&B - Daten weichen zwar erkennbar von der Normalverteilung ab, doch hat diese Abweichung sicher keine Relevanz für die Inferenzstatistik im HLM:



Die Verteilung der U_{0j} - Werte gibt keinen Anlass zum Zweifel an der Normalitätsannahme:

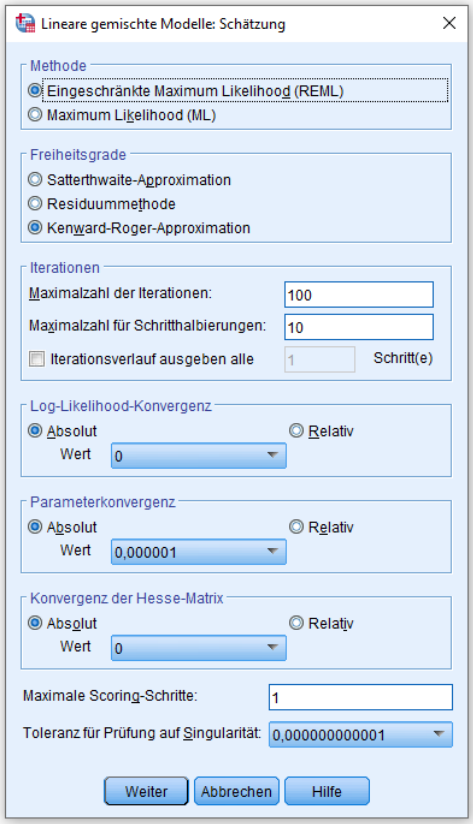


4 Methodische Spezialthemen

In diesem Abschnitt werden methodische Details behandelt, die bislang der Einfachheit halber weggelassen wurden. In der Praxis spielen sie eine erhebliche Rolle, entweder als attraktive Lösungen für typische Aufgaben (z. B. Likelihood-Quotienten - Tests), als unvermeidliche Pflichten (z. B. Power-Analyse) oder als Störungen der Routine (z. B. Schätzprobleme).

4.1 REML- vs. ML-Schätzung

Die SPSS-Prozedur MIXED unterstützt zwei Schätzmethoden, wählbar im folgenden, über den Schalter **Schätzung** erreichbaren Dialog:



Beide Methoden bestimmen nach dem Maximum Likelihood - Prinzip Parameterschätzungen so, dass die Wahrscheinlichkeit der beobachteten Daten unter dem parametrisch spezifizierten Modell maximal wird. Während die ML-Methode (oft auch als *FML* für *full maximum likelihood* bezeichnet) bei dieser Maximierung simultan die festen Parameter (Regressionskoeffizienten) *und* die Kovarianzparameter einbezieht, bestimmt die REML-Methode (*restricted maximum likelihood*) nur die Kovarianzparameter nach dem ML-Prinzip.[1] Mit Hilfe der geschätzten Kovarianzen lassen sich die festen Parameter und deren Standardfehler über die GLS-Methode (*generalized least squares*) schätzen (Weiss 2010).

Die REML-Methode berücksichtigt den Verlust von Freiheitsgraden aufgrund der Verwendung von Schätzungen für die festen Parameter und kann so eine Verzerrung beim Schätzen der Kovarianzparameter vermeiden (bei gleichen Gruppengrößen) oder zumindest reduzieren (bei ungleichen Gruppengrößen). Im Gegensatz dazu unterschätzt die FML-Methode Varianzen vor allem in kleinen Stichproben deutlich.

In einer Simulationsstudie mit dem folgenden wahren Modell:[2]

---

[1] Statt der Likelihood der Daten wird die Likelihood der Residuen maximiert. Zu Residuen gelangt man über OLS-Schätzer für die festen Parameter. Dass diese Entfernung der festen Parameter aus der Likelihood-Funktion zulässig ist, begründen z. B. de Leeuw & Meijer (2008).

[2] SPSS-Dateien mit Syntax, Daten und Ausgaben zur Simulation sind im Ordner **REML vs. FML** an der im Vorwort vereinbarten Stelle zu finden.

- keine Regressoren
- Zufälliger Ordinatenabschnitt mit Mittelwert $\gamma_{00} = 0$ und Varianz $\tau_{00} = 1$
- Residualvarianz $\sigma^2 = 1$

wurde eine Stichprobe mit 20 Makroeinheiten und jeweils 10 Individuen gezogen. Die REML-Methode geht im Test zum festen Effekt γ_{00} korrekt von 19 Freiheitsgraden aus

Schätzungen fester Parameter[a]

Parameter	Schätzung	Standard Fehler	Freiheitsgrade	T-Statistik	Sig.	Konfidenzintervall 95% Untergrenze	Konfidenzintervall 95% Obergrenze
Konstanter Term	-,041407	,230749	19	-,179	,859	-,524370	,441555

a. Abhängige Variable: y.

und liefert eine gute Schätzung für τ_{00}:

Schätzungen von Kovarianzparametern[a]

Parameter		Schätzung	Standard-fehler	Wald Z	Sig.	Konfidenzintervall 95% Untergrenze	Konfidenzintervall 95% Obergrenze
Residuum		,857137	,090350	9,487	,000	,697149	1,053841
Konstanter Term [Subjekt = id]	Varianz	,979185	,345617	2,833	,005	,490254	1,955727

a. Abhängige Variable: y.

Demgegenüber nimmt die FML-Methode im Test zum festen Effekt γ_{00} 20 Freiheitsgrade an

Schätzungen fester Parameter[a]

Parameter	Schätzung	Standard Fehler	Freiheitsgrade	T-Statistik	Sig.	Konfidenzintervall 95% Untergrenze	Konfidenzintervall 95% Obergrenze
Konstanter Term	-,041407	,224906	20	-,184	,856	-,510553	,427738

a. Abhängige Variable: y.

und ermittelt eine nach unten verzerrte Schätzung für τ_{00}:

Schätzungen von Kovarianzparametern[a]

Parameter		Schätzung	Standard-fehler	Wald Z	Sig.	Konfidenzintervall 95% Untergrenze	Konfidenzintervall 95% Obergrenze
Residuum		,857137	,090350	9,487	,000	,697149	1,053841
Konstanter Term [Subjekt = id]	Varianz	,925940	,320041	2,893	,004	,470300	1,823018

a. Abhängige Variable: y.

In der Praxis liefern die beiden Methoden ähnliche Schätzungen für die festen Parameter. Es bestehen jedoch Unterschiede bei den geschätzten Varianzen und Kovarianzen der Zufallseffekte, wenn die Anzahl der Gruppen abzüglich der Anzahl von Regressionskoeffizienten im Makromodell kleiner als 50 wird (Snijders & Bosker 2012, S. 60). In SPSS ist (wie in den meisten anderen Programmen) die insgesamt zu bevorzugende REML-Methode voreingestellt.

Ein Vorteil der ML-Methode besteht jedoch in der Möglichkeit, Likelihood-Quotienten - Tests (durch Vergleich von geschachtelten Modellen, vgl. Abschnitt 4.2) nicht nur für Kovarianzparameter (wie bei der REML-Methode), sondern auch für feste Parameter durchzuführen.

4.2 Likelihood-Quotienten - Tests

4.2.1 Eine generelle Teststrategie zum Vergleich von geschachtelten Modellen

Die in Abschnitt 3.5.3 zum HS&B - Beispiel berichteten Ergebnisse zeigen keine signifikante Varianz τ_{11} für den Zufallseffekt U_{1j} auf die Steigungen in den schulinternen Regressionen von MACH auf CSES nach Aufnahme der Makroregressoren MEANSES und SECTOR in das Modell. Daher haben wir in Abschnitt 3.6.1 den Zufallseffekt U_{1j} aus dem Makromodell entfernt:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_1 + \gamma_{02}W_2 + U_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_1 + \gamma_{12}W_2$$

Bei der Anpassung des MIXED-Kommandos sind aus dem RANDOM-Subkommando der Mikroregressor CSES und die Typangabe zur Kovarianzmatrix der Zufallseffekte zu entfernen:

```
mixed mAch with meanses sector cses
 /criteria=dfmethod(kenwardroger)
 /fixed = meanses cses sector cses*meanses cses*sector
 /random = intercept | subject(school)
 /print = solution testcov.
```

Beim Vergleich des eingeschränkten (um die *beiden* Parameter τ_{11} und τ_{01} reduzierten) Modells mit dem zuvor betrachteten (uneingeschränkten) Modell setzen wir einen **Likelihood-Quotienten - Test** ein. In der Literatur wird die mit (-2) multiplizierte logarithmierte Likelihood auch als *Devianz* bezeichnet und der Likelihood-Quotienten - Test dementsprechend als *Devianztest*. Er hat im Vergleich zum Wald-Test zwei Vorteile:

- Die berechneten Überschreitungswahrscheinlichkeiten besitzen insbesondere bei den Varianzen und Kovarianzen von Zufallseffekten eine höhere Präzision.
- Ein simultaner Likelihood-Quotienten - Test für *mehrere* Parameter macht weniger Aufwand als ein Multiparameter - Wald-Test (siehe Snijders & Bosker 2012. S. 96). Bedarf für einen solchen Test besteht z. B. auch bei der Beurteilung eines kategorialen Regressors mit mehr als zwei Ausprägungen.

(Ko)varianzen von Zufallseffekten können auf diese Weise bei Verwendung der ML- *oder* REML-Schätzmethode getestet werden. Ein analoger Test für *feste* Parameter ist nur bei Verwendung der ML-Schätzmethode möglich (siehe Abschnitt 4.1; vgl. Raudenbush & Bryk 2002, S. 64).

Wir erhalten für das vereinfachte Modell die folgende Tabelle mit den Informationskriterien, welche u.a. die logarithmierte und anschließend mit (-2) multiplizierte eingeschränkte Likelihood enthält:

Informationskriterien[a]

Eingeschränkte -2 Log Likelihood	46504,787
Akaike-Informationskriterium (AIC)	46508,787
Hurvich und Tsai (IC)	46508,789
Bozdogan-Kriterium (CAIC)	46524,545
Bayes-Kriterium von Schwarz (BIC)	46522,545

Die Informationskriterien werden in einem möglichst kleinen Format angezeigt.

a. Abhängige Variable: mAch.

Die Differenz:

$$46504{,}787 - 46503{,}663 = 1{,}124$$

ist bei Gültigkeit des eingeschränkten Modells χ^2-verteilt mit 2 Freiheitsgraden. Es ergibt sich eine Überschreitungswahrscheinlichkeit von 0,570, und folglich kann die H_0 (also das eingeschränkte Modell) akzeptiert werden.[1]

Erwartungsgemäß schneidet das eingeschränkte Modell bei den Informationskriterien AIC und BIC *besser* ab als die Variante mit zufälligen Steigungskoeffizienten.

4.2.2 Einseitige Tests für die Varianzen von Zufallseffekten

Weil die Varianz von random intercept - oder random slope - Zufallseffekten nicht negativ werden kann, sollten Likelihood-Quotienten - Tests mit einseitiger Testausrichtung und entsprechend optimierter Power durchgeführt werden.

Ist nur *ein* random intercept - Zufallseffekt im Modell, muss lediglich die gemäß Abschnitt 4.2.1 ermittelte Überschreitungswahrscheinlichkeit des Devianztests halbiert werden (Snijders & Bosker 2012, S. 98). Alternativ kann man das 0,90 - Fraktil der χ^2 - Verteilung mit einem Freiheitsgrad als kritischen Wert ermitteln, wobei 2,71 resultiert.

In einer Simulationsstudie wurde bei gültiger random intercept - Nullhypothese ($\tau_{00} = 0$) in 1000 Stichproben die Likelihood-Quotienten - Prüfgröße (Devianz-Differenz) ermittelt. Der eben begründete kritische Wert 2,71 wurde in 4,6% aller Stichproben überschritten, so dass der Test bei gültiger H_0 sein α-Niveau eingehalten hat:

Devianz-Differenz signifikant

	Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Nein	954	95,4	95,4	95,4
Ja	46	4,6	4,6	100,0
Gesamt	1000	100,0	100,0	

Sind *mehrere* Zufallseffekte im Modell, wird es komplizierter mit der Begründung des Power-optimierten einseitigen Tests, weil eine laut H_0 auf null gesetzte Varianz impliziert, dass die beteiligten Kovarianzen ebenfalls gleich null sind. In Snijders & Bosker (2012, S. 99) werden für einen Test von random slope - Varianzen die Verfahrensschritte beschrieben.

4.3 Erforderlicher Stichprobenumfang und Power

4.3.1 Minimal erforderlicher Stichprobenumfang

In einer Simulationsstudie haben Maas & Hox (2005) für das folgende Zweiebenenmodell mit random intercept, random slope und cross-level - Interaktion

$$Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} W_j + \gamma_{11} X_{ij} W_j \quad + \quad U_{0j} + U_{1j} X_{ij} + R_{ij}$$

für

- Stichprobenumfänge auf der Makroebene (Gruppenzahlen) von 30 bis 100
- Stichprobenumfänge auf der Mikroebene (Individuen pro Gruppe) von 5 bis 50
- Intraklassenkorrelationen von 0,1 bis 0,3

die Unverzerrtheit und Standardfehler der Schätzungen untersucht. Dabei resultierten überwiegend beruhigende Ergebnisse (S. 90):

---

[1] Zur Berechnung der Überschreitungswahrscheinlichkeit mit SPSS taugt z. B. das folgende Kommando:
compute p = 1 - CDF.CHISQ(1.124, 2).

> Summing up, both the regression coefficients and the variance components are all estimated without bias, in all of the simulated conditions. The standard errors of the regression coefficients are also estimated accurately, in all of the simulated conditions. The standard errors of the second-level variances are estimated too small when the number of groups is substantially lower than 100.

In der folgenden Tabelle mit Ergebnisdetails aus Maas & Hox (2005, S. 91) sind für eher kritische Konstellationen die relativen Häufigkeiten von fehlerhaften 95% - Vertrauensintervallen (ohne Überdeckung des wahren Parameters) angegeben. Insbesondere für die Koeffizienten zum Mikro- bzw. Makroregressor und zur Interaktion liegen die Fehlerraten nahe beim korrekten Wert 0,05:

Anzahl Gruppen	**Gruppen-größe**	**ICC**	τ_{00}	τ_{11}	σ^2	γ_{00}	γ_{10} (zu *X*)	γ_{01} (zu *W*)	γ_{11} (zu *XW*)
30	5	0,2	,086	,093	,067	,073	,063	,052	,062
30	30	0,2	,079	,075	,053	,064	,049	,041	,051
50	5	0,2	,070	,066	,058	,059	,057	,051	,047

Damit sind wohl in vielen Zweiebenenstudien keine irregulären Ergebnisse aufgrund einer unzureichenden Stichprobengröße zu befürchten. Insbesondere zeigen die Tests zu den festen Koeffizienten ab 30 Gruppen zuverlässige Fehlerraten.

Es ist auch kein Problem, wenn manche Gruppen lediglich *ein* Mitglied enthalten, solange andere Gruppen größer sind (Snijders & Bosker 2012, S. 56).

Für Studien mit einer kleinen Zahl von Gruppen ($J \leq 50$) empfehlen Snijders & Bosker (2012, S. 94), wegen der besser geschätzten Standardfehler die REML-Methode gegenüber der ML-Methode zu bevorzugen.

Für durch Gruppenmittelwerte geschätzte Makroregressoren haben Lüdtke et al. (2008) eine Verfälschung (Minderung) der geschätzten Regressionskoeffizienten nachgewiesen, die durch Reliabilitätsmängel bei der Schätzung von Gruppenausprägungen durch eine zu geringe Zahl von Gruppenmitgliedern entsteht. Diese Minderung war besonders ausgeprägt bei Gruppengrößen kleiner 30 und bei einer niedrigen Intraklassenkorrelation (ICC = 0,05). Wenn durch Gruppenmittelwerte geschätzte Makroregressoren zum Einsatz kommen, was in Abschnitt 3.4.1.1.2 im Zusammenhang mit der gruppeninternen Zentrierung von Mikroregressoren empfohlen wird, sind offenbar Gruppengrößen ab 30 wünschenswert, um eine verzerrte (geminderte) Schätzung von Between-Effekten zu vermeiden. Alternativ kommt an Stelle der in diesem Manuskript beschriebenen, mit der SPSS-Prozedur MIXED realisierbaren, Mehrebenentechnik eine mit Strukturgleichungsanalyseprogrammen wie Mplus realisierbare Multilevel-SEM - Technik in Frage, welche die Minderung von Between-Effekten durch Reliabilitätsmängel verhindert.

4.3.2 Beta-Fehler-basierte Stichprobenumfangsplanung

Eine **zulässige** Stichprobengröße (ohne ernsthafte Verzerrungen bei Parameterschätzungen und Standardfehlern) garantiert noch nicht, dass die **Power** der Hypothesentests und die **Präzision** der Vertrauensintervalle zufriedenstellend sind. Leider ist eine Stichprobenumfangsplanung bzw. eine post hoc - Poweranalyse bei Mehrebenenmodellen erheblich komplizierter als z. B. bei der linearen Regression. Eine gelungene Darstellung der Thematik findet sich in Snijders (2005). Zu Beginn macht der Autor zwei wichtige qualitative Aussagen (S. 2):

- Der Stichprobenumfang auf der höchsten Ebene ist entscheidend (z. B. die Anzahl der Gruppen in einer Zweiebenenstudie). Wenn die Anzahl der Gruppe zu einer ausreichenden Power für die Prüfung der Effekte der Makroregressoren führt, ist in aller Regel die Power zur Prüfung der Effekte der Mikroregressoren mehr als ausreichend. Vom β-Fehlerrisiko sind bei einer Mehrebenenanalyse also vor allem die Makroregressoren betroffen.
- Geringe Fallzahlen auf der niedrigsten Ebene (z. B. durchschnittlich 3 Gruppenmitglieder in einer Zweiebenenstudie) stellen für die Power der Signifikanztests zu den Regressionskoeffizienten kein Problem dar. Darunter leider nur die Signifikanztests bzw. Vertrauensintervalle zu den intercept- bzw. slope-Varianzen.

Die Power des Signifikanztests zu einem Regressionsparameter hängt ab ...

- von der betragsmäßigen Größe des Parameters
- vom Standardfehler der Parameterschätzung
- von der (ein- oder zweiseitigen) Testausrichtung
- von der gewählten Fehlerrate erster Art (α-Niveau)

Um in einem einseitigen Test zum Niveau α über den Parameter γ mit approximativ normalverteiltem Stichprobenschätzer $\hat{\gamma}$ die Power (1- β) zu erzielen, muss der Quotient aus dem Parameter und seinem Standardfehler mindestens den folgenden Wert annehmen Snijders (2005, S. 2):

$$\frac{\gamma}{\sigma(\hat{\gamma})} \approx z_{1-\alpha} + z_{1-\beta}$$

Dabei sind $z_{1-\alpha}$ bzw. $z_{1-\beta}$ das (1 - α) bzw. (1 - β) - Fraktil der Standardnormalverteilung. Ist bei $\alpha = 0{,}05$ eine Power von (1 - β) = 0,80 gewünscht, resultiert z. B. der Wert 2,4:

$$z_{0,95} + z_{0,8} = 2{,}4$$

Für den zweiseitigen Test lautet die Forderung für den Quotienten aus dem Parameter und seinem Standardfehler:

$$\frac{\gamma}{\sigma(\hat{\gamma})} \approx z_{1-\alpha/2} + z_{1-\beta}$$

Ist bei $\alpha = 0{,}05$ eine Power von (1 - β) = 0,80 gewünscht, resultiert z. B. der Wert 2,8:

$$z_{0,975} + z_{0,8} = 2{,}8$$

Für die Parameter in einigen sehr einfachen Mehrebenenmodellen sind Formeln zur Berechnung des Standardfehlers bekannt (siehe Snijders 2005), so dass der erforderliche Stichprobenumfang für eine angenommene γ-Ausprägung über die obigen approximativen Gleichungen ermittelt werden kann.

Eine flexible, wenngleich aufwändige Prozedur zur Stichprobenumfangsplanung für Hypothesentests zu Parametern in *beliebigen* Mehrebenenmodellen besteht darin, ...

- eine Population mit den angenommenen Parameterausprägungen zu simulieren,
- aus dieser Population eine *große* Stichprobe ziehen, um den interessierenden Standardfehler präzise schätzen zu können,
- den benötigten Stichprobenumfang über die oben angegebene Bedingung für den Quotienten aus einem Parameter und seinem Standardfehler zu berechnen.

Zu Demonstrationszwecken soll ein relativ einfaches Beispiel mit dem folgenden Modell betrachtet werden:

$$Y_{ij} = \gamma_{00} + \gamma_W W_j + \gamma_X X_{ij} + U_{0j} + R_{ij}$$

Die Regressionskoeffizienten γ_X und γ_W wurden abweichend von der bisherigen Notation vereinfacht bezeichnet.

Durch die folgende SPSS-Syntax

```
set seed = 1.
input program.
loop #j = 1 to 10000.
  compute id = #j.
  leave id.
  compute w = normal(1).
  leave w.
  compute u0j = normal(1).
  leave u0j.
  loop #k = 1 to 5.
     compute x =  normal(1).
     compute y = 1 + 0.3*x + 0.3*w + u0j  + normal(1).
     end case.
  end loop.
end loop.
end file.
end input program.
execute.
```

werden 10.000 Gruppen mit jeweils 5 Mitgliedern aus einer Population mit den folgenden Parameterwerten gezogen:

- $\gamma_X = \gamma_W = 0{,}3$
- $\sigma_R^2 = \tau_{00} = 1$

Es soll eine Zweiebenenstudie mit der identischen Gruppengröße $n = 5$ durchgeführt werden. Die Anzahl der Gruppen ist so zu bestimmen, dass für die angenommenen Parameterwerte in einem zweiseitigen Test zum Niveau $\alpha = 0{,}05$ eine Power von $(1 - \beta) = 0{,}8$ resultiert.

Aus den mit obiger Syntax simulierten Daten werden über das folgende MIXED-Kommando

```
mixed y with x w
 /criteria=dfmethod(kenwardroger)
 /fixed = x w
 /random = intercept | subject(id)
 /print = solution testcov.
```

die Standardfehler zu den Regressionskoeffizienten geschätzt:

Schätzungen fester Parameter[a]

Parameter	Schätzung	Standardfehler	Freiheitsgrade	T-Statistik	Sig.	Konfidenzintervall 95% Untergrenze	Konfidenzintervall 95% Obergrenze
Konstanter Term	1,002668	,010894	10002,088	92,036	,000	,981313	1,024023
x	,306595	,004878	43055,117	62,853	,000	,297034	,316156
w	,298807	,010831	10002,091	27,589	,000	,277576	,320037

a. Abhängige Variable: y.

In jedem geschätzten Standardfehler (z. B. $\hat{\sigma}(\hat{\gamma}_X) = 0{,}004878$ steckt der Faktor $\frac{1}{\sqrt{10000}}$, der zunächst herausgerechnet wird, z.B.:

$$0{,}004878 \cdot \sqrt{10000} = 0{,}4878$$

Die gesuchte Gruppenzahl J_X muss folgende Bedingung erfüllen:

$$\frac{0{,}3}{\sqrt{\frac{1}{J_X}}0{,}4878} = 2{,}8$$

Durch Umstellen resultiert

$$J_X = \left(\frac{2{,}8 \cdot 0{,}4878}{0{,}3} \right)^2 \approx 20{,}728$$

Analog erhalten wir als benötigte Gruppenzahl J_W für den Makroregressor W (mit $\hat{\sigma}(\hat{\gamma}_W) = 0{,}010831$):

$$0{,}010831 \cdot \sqrt{10000} = 1{,}0831$$

$$J_W = \left(\frac{2{,}8 \cdot 1{,}0831}{0{,}3} \right)^2 \approx 102{,}191$$

Bei der Stichprobenplanung ist die größere Zahl relevant. Weil in die Formel

$$\frac{\gamma}{\sigma(\hat{\gamma})} \approx z_{1-\alpha/2} + z_{1-\beta}$$

eine unsichere Normalverteilungsannahme eingeflossen ist, sollte man die benötigte Gruppenzahl leicht „aufrunden“.

Zur Kontrolle wurden aus der simulierten Population 500 Zufallsstichproben mit jeweils 106 Gruppen und der identischen Gruppengröße 5 gezogen und mit der Prozedur MIXED analysiert. Der Mikroregressor wurde *immer* signifikant, während für den Makroregressor annähernd die geplante Power erreicht wurde (79 % signifikante Tests).

4.3.3 Designeffekt

Im Zusammenhang mit der Poweranalyse für Mehrebenenmodelle ist der sogenannte **Designeffekt** von Interesse, der als Faktor die Veränderung des Stichprobenbedarfs im Vergleich zu einem Regressionsmodell mit unabhängigen Beobachtungen angibt. Soll der Koeffizient zu einem Makroregressor getestet werden, dann steigt der Stichprobenbedarf bei positiver Intraklassenkorrelation ρ um den Faktor (Snijders 2005, S. 5)

$$1 + (n-1)\rho$$

Dabei sind:

- n die Anzahl der Fälle in einer Makroeinheit (Gruppengröße)
- ρ die Intraklassenkorrelation (vgl. Abschnitt 3.2.2)

Wer hofft, mit seinem Wissen um die Poweranalyse bei der linearen Regression (siehe z. B. Baltes-Götz 2019) und einer Formel zum Designeffekt die Poweranalyse für Mehrebenenmodelle elegant erledigen zu können, sieht sich getäuscht. Besagte Formel hängt nämlich vom konkreten Modell ab und ist nicht für alle Mehrebenenmodelle bekannt. Ist z. B. für einen Mikroregressor der Steigungskoeffizient fixiert (für alle Gruppen gleich), dann *sinkt* der Stichprobenbedarf bei positiver Intraklassenkorrelation ρ um den folgenden Faktor (Snijders 2005, S. 4):

$$1 - \rho$$

Durch die im Modell enthaltenen Gruppeneffekte U_{0j} sinkt die Fehlervarianz und damit auch der Standardfehler.

4.4 Modellierungsstrategie

Mehrebenenmodelle sind meist komplex, und für die Suche nach dem korrekten Modell kann es keinen Algorithmus mit Erfolgsgarantie geben. Als Modellierungsstrategien kommen generell in Frage:

- **Bottom-Up**
  Man beginnt mit einem einfachen Modell und erweitert es sukzessive um Regressoren bzw. Parameter.
- **Top-Down**
  Am Anfang steht ein komplexes Modell mit allen potentiell relevanten festen und zufälligen Effekten, das aufgrund von Signifikanztests sukzessive vereinfacht wird.

Gegen die Top-Down - Strategie spricht, dass der Start mit einem komplexen, mit vielen Schwächen belasteten Modell durch Konvergenzprobleme beim Schätzen und hohem Rechenaufwand die Forschung behindern kann. Ein Risiko der Bottom-Up - Strategie besteht darin, einen Regressor früh auszusortieren, der bei Anwesenheit eines anderen Regressors (z. B. durch Kompensation) einen signifikanten Beitrag leisten kann.

Viele Autoren bevorzugen eine Bottom-Up - Strategie, und Hox (2010, S. 56ff) gibt detaillierte Empfehlungen zur Schrittfolge:

Schritt 1 (random intercept - Modell ohne Regressoren)
Dieses Modell enthält keine Regressoren, aber Residuen auf der Mikro- und der Makroebene,

$$Y_{ij} = \gamma_{00} + U_{0j} + R_{ij}$$

sodass die Intraklassenkorrelation geschätzt und getestet werden kann (vgl. Abschnitte 3.2 und 4.2). Raudenbush & Bryk (2002, S. 24) beschreiben dieses Modell als *fully unconditional*, und wir haben es in Abschnitt 3.2 als *Basismodell* bezeichnet.

Schritt 2 (Aufnahme von Mikroregressoren mit fixierten Parametern)
In das Modell aus Schritt 1 werden Mikroregressoren aufgenommen und auf Signifikanz beurteilt, wobei ein festes (für alle Gruppen identisches) Regressionsgewicht angenommen wird. Bei einem einzigen Mikroregressor resultiert das folgende Modell (in Eingleichungsnotation):

$$Y_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + U_{0j} + R_{ij}$$

Schritt 3 (Aufnahme von Makroregressoren)
In das Modell aus Schritt 2 werden Makroregressoren mit Einfluss auf die Ordinatenabschnitte β_{0j} der Mikroebene aufgenommen und auf Signifikanz beurteilt. Das folgende Modell (in Eingleichungsnotation) besitzt auf beiden Ebenen einen einzigen Regressor:

$$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{10} X_{ij} + U_{0j} + R_{ij}$$

Schritt 4 (random slope - Modell)
Im Vergleich zum Modell aus Schritt 3 werden gruppenspezifische Steigungskoeffizienten der Mikroregressoren zugelassen. Das folgende Modell (in Eingleichungsnotation) besitzt auf beiden Ebenen einen einzigen Regressor:

$$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{10} X_{ij} + U_{1j} X_{ij} + U_{0j} + R_{ij}$$

Schritt 5 (Aufnahme von Cross-Level - Interaktionen)
Im Vergleich zum Modell aus Schritt 4 werden Effekte von Makroregressoren auf die Steigungskoeffizienten der Mikroregressoren zugelassen. Das folgende Modell (in Eingleichungsnotation) besitzt einen Mikro- und einen Makroregressor, die miteinander interagieren:

$$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{10} X_{ij} + \gamma_{11} W_j X_{ij} + U_{1j} X_{ij} + U_{0j} + R_{ij}$$

Snijders & Bosker (2012, Abschnitt 6.4) beschreiben zunächst eine Bottom-Up - Strategie mit strikter (Mikro → Makro) - Sequenz. Dabei sollte zunächst das Mikromodell entwickelt werden, in dem Unterschiede zwischen den Makroeinheiten durch random intercept - und random slope - Zufallseffekte vertreten sind. Zur Erklärung der (nicht zuletzt aufgrund von Signifikanztests) im Modell verbleibenden Zufallseffekte wird anschließend das Makromodell entwickelt, d.h. es werden Makroregressoren mit Haupt- und Wechselwirkungseffekten (inkl. Cross-Level - Interaktionen) ins Modell aufgenommen.

Außerdem geben Snijders & Bosker die folgenden Hinweise zur Modellspezifikation:

- Während die festen Effekte eher aus theoretischen Gründen ins Modell gelangen, ist bei den Zufallseffekten oft eine Daten-orientierte Modellspezifikation anzutreffen und nach Meinung der Autoren auch zulässig.
- Abweichend von der strikten (Mikro → Makro) - Sequenz bei der Modellbildung sollten theoretisch plausible Cross-Level - Interaktionen auch bei nicht-signifikanter random-slope - Varianz einbezogen werden, weil der Test für den Koeffizienten zur Wechselwirkung eine bessere Power besitzt als der Test zur random-slope - Varianz. Mit analoger Begründung sollte nach Snijders & Bosker ein Makroregressor mit theoretisch relevantem Haupteffekt auch dann ins Modell aufgenommen werden, wenn auf der Mikroebene keine signifikante random-intercept - Varianz nachgewiesen werden konnte. Damit verflüchtigen sich allmählich die kleinen Unterschiede zwischen den von Snijders & Bosker (2012) sowie von Hox (2010) empfohlenen Analysestrategien.
- Vollständigkeitsempfehlungen:
    - Mit einer Interaktion sollten auch die zugehörigen Terme von niedrigerer Ordnung (unabhängig von Signifikanzbeurteilungen) ins Modell aufgenommen werden. Diese generelle Modellierungsempfehlung gilt auch für Mehrebenenmodelle.
    - Wird für einen Mikroregressor eine random slope - Varianz angenommen, dann sollte sich auch der zugehörige feste Effekt im Modell befinden.
    - Für Zufallseffekte im Modell sollten auch deren Kovarianzen einbezogen werden.
- Für Interaktionen von Mikroregressoren sollten *keine* zufälligen Koeffizienten einbezogen werden, weil diese oft schwer zu interpretieren sind.
- Bei manchen Mikroregressoren muss damit gerechnet werden, dass sich Within- und Between-Effekte unterscheiden (siehe Abschnitt 3.4.1). Bei einem betroffenen Mikroregressor sollte der zugehörige Makroregressor mit den Gruppenmittelwerten von Beginn an in das Modell aufgenommen werden.

Einige Hinweise von Snijders & Bosker sprechen gegen eine strikte (Mikro → Makro) - Sequenz bei der Modellbildung. Die Autoren erwägen daher auch eine Top-Down - Strategie mit den oben genannten Vor- und Nachteilen.

Ein universell anwendbares „Schema F“ mit maximaler Chance, das korrekte Modell zu finden, existiert wohl nicht. Die in Abschnitt 3 in Anlehnung an Raudenbush & Bryk (2002, Kap. 4) gewählte Sequenz von Modellen zu den HS&B - Daten war überwiegend didaktisch motiviert. Dabei sollten ohne vordringliche Orientierung an einer Modellierungsstrategie methodische Themen in angemessener Dosierung präsentiert werden.

Alle empfohlenen Modellierungsstrategien enthalten neben einer konfirmatorischen auch eine explorative (induktive) Komponente. Nach Snijders & Bosker (2012, S. 107) muss dabei der konfirmatorische Anspruch einer Studie *nicht* verloren gehen. Um einen vor der Datenerhebung geplanten Test über einen Hauptparameter gültig durchführen zu können, ist eventuell im Bereich von Nebenparametern ein induktives (Daten-orientiertes) Vorgehen erforderlich.

4.5 Kosten für Zufallseffekte

Auf dem letzten Entwicklungsstand des HS&B - Modells haben wir den Zufallseffekt U_{1j} aus dem Makromodell für β_{1j} entfernt, weil seine Varianz nach der Aufnahme von Makroregressoren nicht mehr signifikant von null verschieden war. Dies wirkt sich günstig auf die Power der Hypothesentests zu CSES,

CSES × MEANSES und CSES × SECTOR aus: Die Standardfehler schrumpfen, und die Zahl der Freiheitsgrade steigt sprunghaft, weil nicht mehr die Zahl der Makroeinheiten, sondern die Zahl der Mikroeinheiten den Ausgangspunkt für die Kenward-Roger - bzw. Satterthwaite - Korrektur bildet. Dies zeigt sich beim Vergleich der in Abschnitt 3.5.3 berichteten Schätzergebnisse für die festen Parameter

Schätzungen fester Parameter[a]

						Konfidenzintervall 95%	
Parameter	Schätzung	Standardfehler	Freiheitsgrade	T-Statistik	Sig.	Untergrenze	Obergrenze
Konstanter Term	12,127931	,199314	159,896	60,848	,000	11,734304	12,521558
cses	2,945041	,156029	139,619	18,875	,000	2,636555	3,253526
meanses	5,332874	,369204	150,986	14,444	,000	4,603402	6,062347
sector	1,226579	,306298	149,613	4,005	,000	,621349	1,831808
meanses * cses	1,039232	,299462	160,570	3,470	,001	,447840	1,630624
sector * cses	-1,642674	,240222	143,349	-6,838	,000	-2,117509	-1,167839

a. Abhängige Variable: mAch.

mit dem Ergebnis *ohne* den Zufallseffekt U_{1j}:

Schätzungen fester Parameter[a]

						Konfidenzintervall 95%	
Parameter	Schätzung	Standardfehler	Freiheitsgrade	T-Statistik	Sig.	Untergrenze	Obergrenze
Konstanter Term	12,128207	,199217	160,006	60,879	,000	11,734774	12,521640
cses	2,942146	,151209	7018,261	19,457	,000	2,645730	3,238561
meanses	5,336670	,369014	151,072	14,462	,000	4,607576	6,065764
sector	1,224529	,306140	149,695	4,000	,000	,619616	1,829442
meanses * cses	1,044438	,291048	7018,261	3,589	,000	,473896	1,614981
sector * cses	-1,642156	,233094	7018,261	-7,045	,000	-2,099090	-1,185222

a. Abhängige Variable: mAch.

Man wird in der Regel zu Steigungskoeffizienten initial Zufallseffekte zulassen, muss aber den Preis einer reduzierten Power bei den Hypothesentests für die zugehörigen festen Parameter in Kauf nehmen (Tabachnick & Fidell 2007, S. 829). Besitzt ein Zufallseffekt (nach Aufnahme von Makroregressoren) keine signifikante Varianz, sollte man ihn entfernen.

4.6 Schätzprobleme

Die Parameter werden in einem iterativen Verfahren geschätzt. Für eine misslungene Konvergenz

Warnungen

Die Iteration wurde beendet, ohne Konvergenz zu erreichen. Die Prozedur MIXED wird trotz dieser Warnung fortgesetzt. Nachfolgend erzeugte Ergebnisse basieren auf der letzten Iteration. Die Gültigkeit der Anpassungsgüte des Modells ist nicht gesichert.

der Schätzungen kommen folgenden Ursachen in Frage (vgl. Hox 2010, S. 42; Snijders & Bosker 2012, S. 90; Tabachnick & Fidell 2007, S. 834):

- Gravierende Fehler im Modell
- Anzahl der Iterationen zu klein
  Maßnahme: Anzahl der Iterationen über den Subdialog **Schätzung** erhöhen
- Stichprobe zu klein (vgl. Abschnitt 4.3)
- Zufallseffekte mit sehr kleiner Varianz vorhanden
  Maßnahme: Betroffene Zufallseffekte entfernen
- Zu viele random slopes
  Das Risiko für Konvergenzprobleme steigt, wenn mehr als ca. drei Regressoren einen zufälligen Steigungskoeffizienten haben. Manchmal hilft es, solche Variablen zu zentrieren.

4.7 Notation für gemischte Modelle

Die von der SPSS-Prozedur MIXED in der Ausgabe verwendeten Bezeichnungen **G** für die Kovarianzmatrix der zufälligen Effekte und **R** für die Kovarianzmatrix der Mikroresiduen stammen aus einer verbreiteten Notation für lineare gemischte Modelle mit festen und zufälligen Effekten (LMM, *Linear Mixed Models*). Hierarchische lineare Modelle (für Cluster- oder Messwiederholungsdaten) können als Spezialfälle von linearen gemischten Modellen aufgefasst werden, und es lohnt sich, einen Blick auf die LMM-Notation zu werfen, die von der im Manuskript verwendeten Notation bei grundsätzlicher Kompatibilität leicht abweicht.

In der üblichen Matrixschreibweise wird ein gemischtes lineares Modell folgendermaßen notiert (vgl. AL-Marshadi 2007):

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{ZU} + \mathbf{E}$$

Darin bedeuten:

- $\mathbf{Y}$ ($N \times 1$) - Vektor der Kriteriumswerte
- $\mathbf{X}$ ($N \times p$) - Designmatrix für die festen Effekte
- $\boldsymbol{\beta}$ ($p \times 1$) - Vektor mit den festen Parametern
- $\mathbf{Z}$ ($N \times q$) - Designmatrix für die zufälligen Effekte
- $\mathbf{U}$ ($q \times 1$) - Vektor der zufälligen Effekte
  Es wird angenommen, dass die zufälligen Effekte den Erwartungswert 0 besitzen und gemeinsam normalverteilt sind mit der Kovarianzmatrix **G**, kurz: $\mathbf{U} \sim N(0, \mathbf{G})$
- $\mathbf{E}$ ($N \times 1$) - Vektor der Individualresiduen
  Es wird angenommen, dass die Individualresiduen den Erwartungswert 0 besitzen und gemeinsam normalverteilt sind mit der Kovarianzmatrix **R**, kurz: $\mathbf{E} \sim N(0, \mathbf{R})$

Für die vektorielle Zufallsgröße **Y** resultiert eine Normalverteilung mit dem Erwartungswertevektor $\mathbf{X}\boldsymbol{\beta}$ und einer Kovarianzmatrix **V**, die sich folgendermaßen aus den Kovarianzmatrizen **G** und **R** berechnen lässt:

$$\mathbf{V} = \mathbf{ZGZ'} + \mathbf{R}$$

5 Gekreuzte Zufallseffekte

Eine Clusterstichprobe ist nicht immer hierarchisch organisiert. In einem oft herangezogenen Beispiel werden ...

- *J* Gymnasien (Sekundärschulen) einbezogen,
- wobei die Schüler aus *K* Stadtteilen stammen.

Bei der Untersuchung eines Indikators für den Schulerfolg kann man sich für die Beiträge der Schule *und* des Wohnumfelds interessieren. Ein Schüler (eine Mikroeinheit) gehört zu zwei, *nicht* geschachtelten Makroeinheiten (zu einer Sekundärschule und zu einem Stadtteil). Man kann von gekreuzten Zufallseffekten (engl.: *crossed random effects*) sprechen, und ein Schüler gehört in eine Zelle der *zweifaktoriellen* Makroebene. Wir werden in diesem Abschnitt fiktive Daten betrachten mit 4248 Schülern, die aus 30 verschiedenen Stadtteilen (Wohngegenden) stammen und 30 verschiedene Gymnasien besuchen.

Sicherlich spielt bei der Schulwahl die Entfernung eine Rolle, doch sind in vielen Gymnasien zahlreiche Stadtteile vertreten, weil sich in vielen Stadtteilen keine Sekundärschule befindet, und die Beförderung mit dem öffentlichen Personennahverkehr vielen Familien eine Schulwahl ermöglicht. In einer konkreten Studie mit den Verhältnissen aus dem Beispiel besteht eine methodische Option darin, den Stadtteil als kategoriale Mikrovariable mit fixierten Effekten (vgl. Abschnitt 2.3) zu behandeln, sofern keine Makrovariable bezogen auf die Stadtteile zu untersuchen ist. Generell sollte man aber darauf vorbereitet sein, Daten mit gekreuzten Zufallsfaktoren zu analysieren.

Im folgenden Beispiel mit den ersten 10 Schülern (dargestellt durch die eindeutige Fallnummer) soll die Verteilungslogik veranschaulicht werden:

Stadtteil	**Gymnasium**			
	A	**B**	**C**	**...**
I	1, 2		3	
II	4	5, 6	7	
III		8	9, 10	
...				

Über diesen Abschnitt hinausgehende Details zu gekreuzten Zufallsfaktoren finden sich bei z. B. bei Hox (2010), Judd et al. (2017), Snijders & Bosker (2012) und West et al. (2015).

5.1 Modell

Im fiktiven Beispiel soll noch analog zum realen HS&B - Beispiel (siehe Abschnitt 3) die Variable SECTOR mit den folgenden Werten einbezogen werden:

- 1: Konfessionsgebundene Schule
- 0: Öffentliche Schule

Das Mikroebenenmodell lautet:

$$Y_{i(j,k)} = \beta_{0(j,k)} + R_{i(j,k)} \quad \text{für } i = 1, .., n_{(j,k)}, j = 1, .., J \text{ und } k = 1, .., K$$

Durch die Klammernotation kommt zum Ausdruck, dass mit *i* eine Mikroeinheit und mit (*j*, *k*) eine Zelle der zweifaktoriellen Makroebene bezeichnet wird.

Im Makromodell wird neben den gekreuzten Zufallseffekten

- U_{0j} für die Schulen
- V_{0k} für die Stadtteile

noch der Regressor SECTOR (Abkürzung *W*) eingesetzt, um den zufälligen Ordinatenabschnitt der Mikroebene $\beta_{0(j,k)}$ (Mittelwert der Makrozelle (*j*, *k*)) zu erklären:

$$\beta_{0(j,k)} = \gamma_{00} + \gamma_{01} W_j + U_{0j} + V_{0k} \quad \text{für } j = 1, .., J, k = 1, .., K$$

Für die Zufallseffekte U_{0j} und V_{0k} wird angenommen:

- Für zwei Wertekombinationen (j, k) und (j', k') der Makroebene (mit $j \neq j'$ und $k \neq k'$) sind die Zufallsvektoren (U_{0j}, V_{0k}) und ($U_{0j'}$, $V_{0k'}$) unabhängig voneinander.
- Für eine Wertekombination (j, k) der Makroebene sind die Zufallseffekte U_{0j} und V_{0k} unabhängig von den Mikroresiduen $R_{i(j,k)}$.
- Für eine Wertekombination (j, k) der Makroebene dürfen die Zufallseffekte U_{0j} und V_{0k} korrelieren.
- Die Zufallseffekte sind normalverteilt:
  - $U_{0j} \sim N(0, Var(U_{0j}))$
  - $V_{0k} \sim N(0, Var(V_{0k}))$

Damit lautet das kombinierte Modell:

$$Y_{i(j,k)} = \gamma_{00} + \gamma_{01} W_j + U_{0j} + V_{0k} + R_{i(j,k)}$$

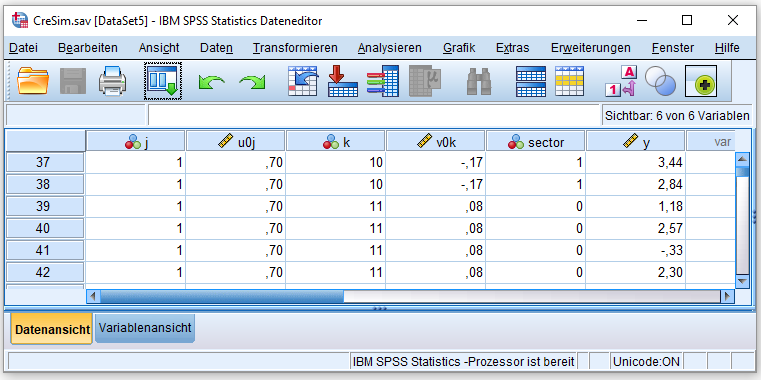
Die wahren Parameterausprägungen in der simulierten Population sind:

- γ_{00}: 1
- γ_{01}: 0,6
- $Var(U_{0j}) = Var(V_{0k})$: 0,25
- $Var(R_{i(j,k)})$ (σ^2) 1

Im Datensatz befinden sich 30 Gymnasien und 30 Stadtteile. Jedes Gymnasium wird von durchschnittlich 5 Schülern aus jedem Stadtteil besucht.[1] Um das wahre Leben zumindest partiell zu simulieren, nehmen Gymnasien mit einem positivem Zufallseffekt aus Stadtteilen mit einem negativen Zufallseffekt nur 2 Schüler, aus Stadtteilen mit positivem Zufallseffekt hingegen 8 Schüler auf. Damit korrelieren über alle Schüler die schulbezogenen und die ortsbezogenen Zufallseffekte positiv. Das verstößt nicht gegen die Modellannahmen und ist für die Analyse nicht problematischer als die Korrelation zwischen zwei Regressoren in einem linearen Modell. Solange beide Zufallseffekte einbezogen werden, sind keine verzerrten Schätzer zu befürchten. Ignoriert aber man einen von den beiden Zufallseffekte, ...

- sind verzerrte Effektschätzungen für die Regressoren im Modell zu befürchten (omitted variable error),
- wird die Varianz der anderen Zufallskomponenten (verbleibender Zufallseffekt, Residuum) überschätzt.

In der Datei **CRESim.sav** mit den simulierten Daten



befinden sich die Variablen:

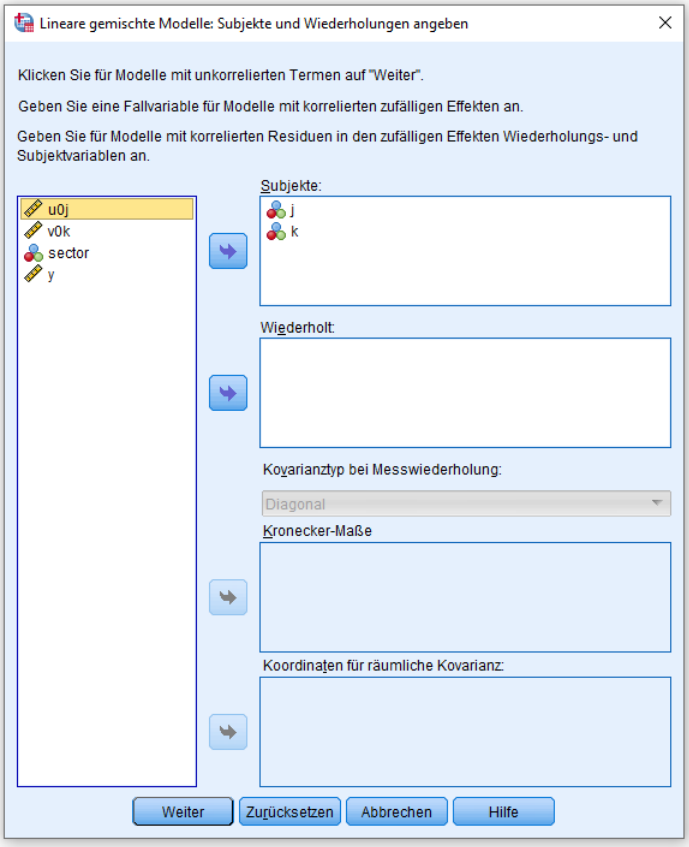
[1] SPSS-Dateien mit Syntax, Daten und Ausgaben zur Simulation sind im Ordner **Gekreuzte Zufallseffekte** an der im Vorwort vereinbarten Stelle zu finden.

- J, U0J
  Nummer und Zufallseffekt der Schule
- K, V0K
  Nummer und Zufallseffekt des Stadtteils
- SECTOR
  Typ der Schule:
  - 1: Konfessionsgebundene Schule
  - 0: Öffentliche Schule
- Y
  Abhängige Variable auf der Mikroebene: Indikator für den Schulerfolg

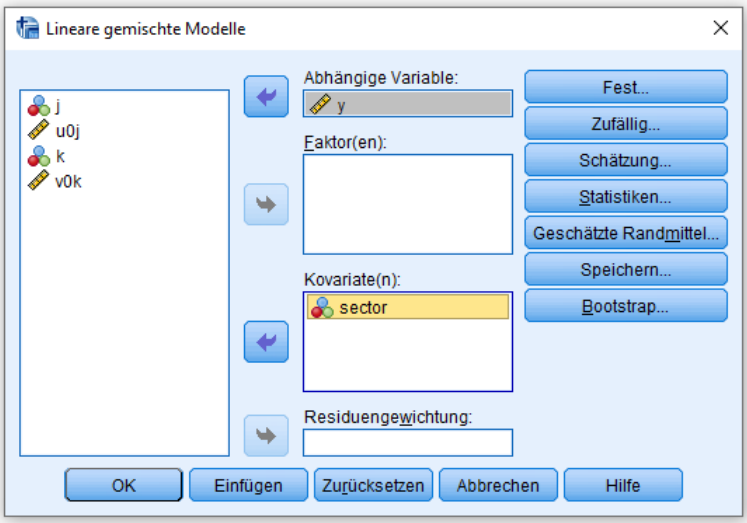
5.2 Anforderung der Analyse in SPSS

Gehen Sie folgendermaßen vor, um in SPSS ein Modell mit gekreuzten Zufallseffekten auf die simulierten Daten anzuwenden:

- Menübefehl **Analysieren > Gemischte Modelle > Linear**
- Wir legen fest, dass die Variablen J und K für die spätere Definition der Makroeinheiten in Frage kommen:

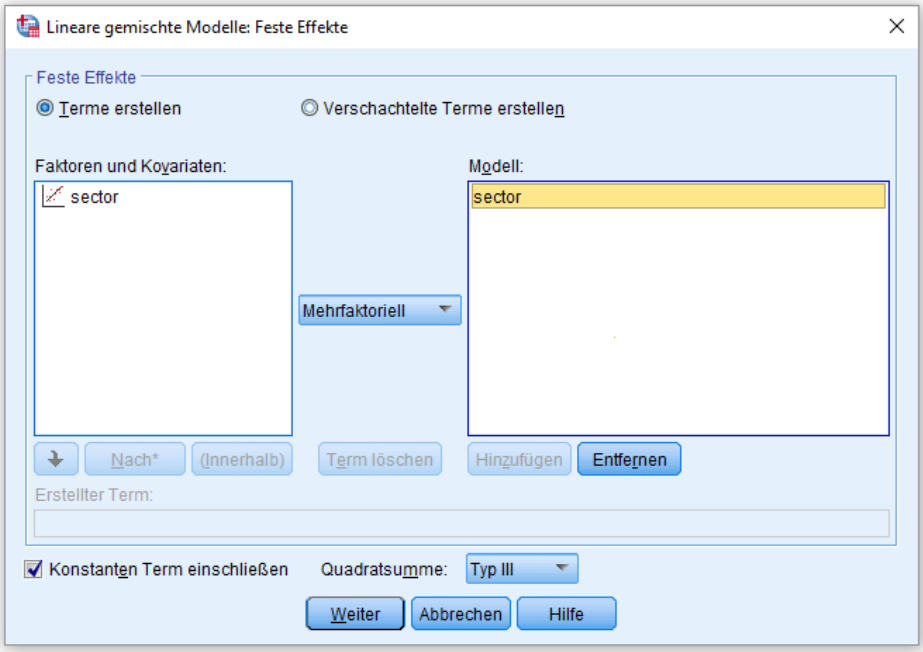


- Wir machen **weiter** und legen Y als **abhängige Variable** sowie SECTOR als **Kovariate** fest:



Weil SECTOR nur zwei Ausprägungen hat, ist gegen die Behandlung als Kovariate nichts einzuwenden.

- Wir fügen in der Subdialogbox für die **festen Effekte** SECTOR zum **Modell** hinzu:

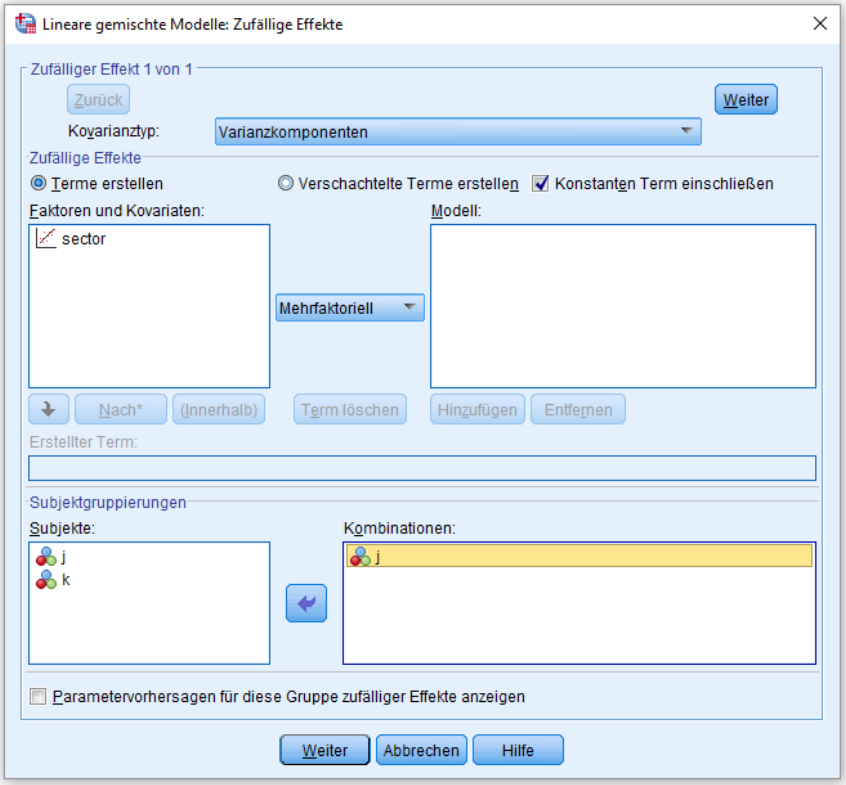


Das Kontrollkästchen für den **konstanten Term** sollte hier auf jeden Fall markiert sein (= Voreinstellung). Somit werden im Dialog mit den **festen Effekten** die folgenden Parameter deklariert:

- Konstanter Term γ_{00}
- SECTOR γ_{01}

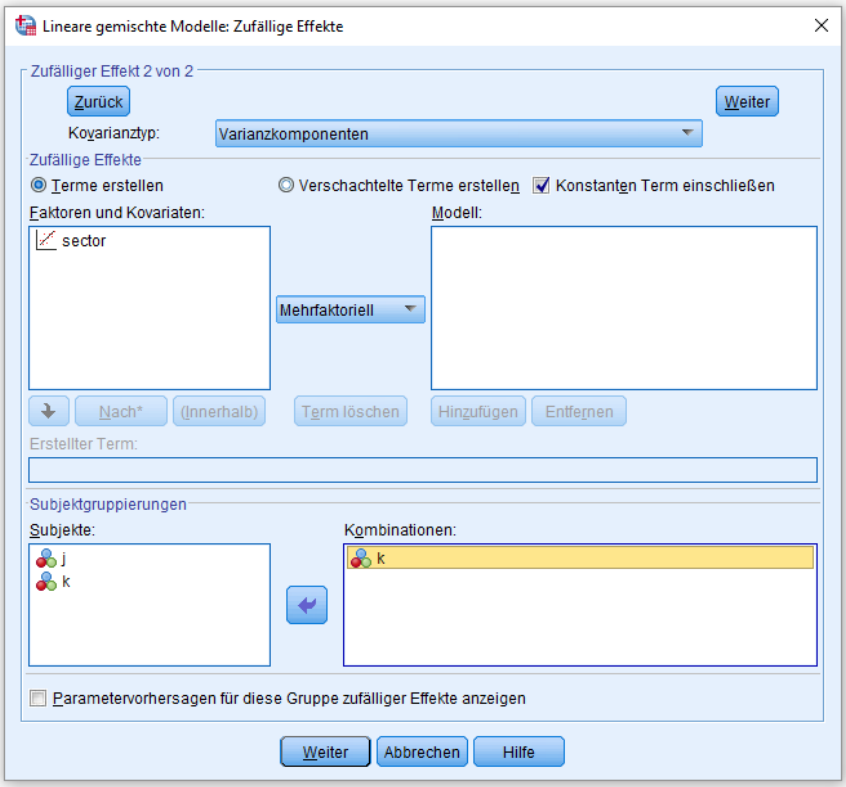
- Wir quittieren die Subdialogbox für die festen Effekte mit **Weiter** und begeben uns in das Gegenstück für die **zufälligen Effekte**. Hier sind folgende Einstellungen vorzunehmen:

- Der **konstante Term** wird **eingeschlossen**, und die Variable J wird in die Liste der **Kombinationen** befördert:



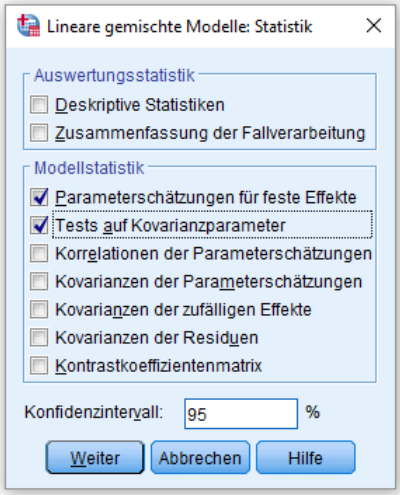
Damit sind die schulbezogenen Zufallseffekte U_{0j} mit dem zugehörigen Parameter $\mathrm{Var}(U_{0j})$ ins Modell einbezogen.

- Nach einem Klick auf den Schalter **Weiter** im Rahmen mit den **zufälligen Effekten** (!) verfahren wir analog, um die ortsbezogenen Zufallseffekte V_{0j} mit dem zugehörigen Parameter $\mathrm{Var}(V_{0k})$ ins Modell aufzunehmen:

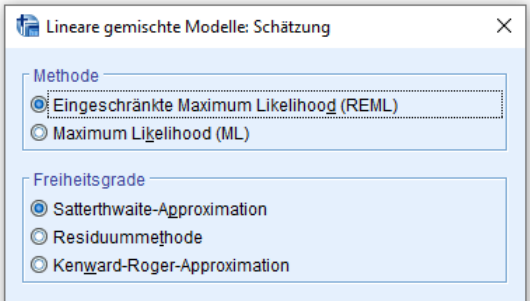


- Als **Kovarianztyp** kann für die beiden gekreuzten Zufallseffekte die Voreinstellung **Varianzkomponenten** beibehalten werden, weil im Modell kein potentieller Kovarianzpartner vorhanden ist.

- Wir fordern in der **Statistik**-Subdialogbox **Parameterschätzer** und **Tests für die Kovarianzparameter** an:



- Im Subdialog **Schätzung** ersetzen wir die **Satterthwaite-Approximation** *nicht* durch die bisher im Manuskript bevorzugte **Kenward-Roger-Approximation**, weil diese präzisere Methode im Modell mit gekreuzten Zufallseffekten den Rechenzeitbedarf enorm ansteigen lässt und dabei im Beispiel praktisch keine Änderung der Ergebnisse bewirkt:[1]



Über den **Einfügen**-Schalter der Hauptdialogbox erhält man folgendes Syntax-Äquivalent zur Modellspezifikation:

```
MIXED y WITH sector
  /CRITERIA=DFMETHOD(SATTERTHWAITE) CIN(95) MXITER(100) MXSTEP(10) SCORING(1)
   SINGULAR(0.000000000001) HCONVERGE(0, ABSOLUTE) LCONVERGE(0, ABSOLUTE)
   PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=sector | SSTYPE(3)
  /METHOD=REML
  /PRINT=SOLUTION TESTCOV
  /RANDOM=INTERCEPT | SUBJECT(j) COVTYPE(VC)
  /RANDOM=INTERCEPT | SUBJECT(k) COVTYPE(VC).
```

Zur manuellen Erstellung einer äquivalenten Syntax ist durch Ausnutzung der passenden Voreinstellungen nur wenig Schreibarbeit erforderlich:

[1] Die Modellschätzung mit **Kenward-Roger-Approximation** dauerte auf einem Rechner mit der Intel-CPU Core i3 ca. 7 Minuten.

```
mixed y with sector
  /fixed = sector
  /random = intercept | subject(j)
  /random = intercept | subject(k)
  /print = solution testcov.
```

5.3 Diskussion der Ergebnisse

In der Modellspezifikation

Modelldimension[a]

		Anzahl Ausprägungen	Kovarianzstruktur	Anzahl Parameter	Subjektvariablen
Feste Effekte	Konstanter Term	1		1	
	sector	1		1	
Zufallseffekte	Konstanter Term	1	Varianz-Komponenten	1	j
	Konstanter Term	1	Varianz-Komponenten	1	k
Residuum				1	
Gesamt		4		5	

a. Abhängige Variable: y.

tauchen fünf Parameter auf:

- Zwei Parameter für festen Effekte (γ_{00}, γ_{01})
- Zwei Parameter für Varianzen von Zufallseffekten (Var(U_{0j}), Var(V_{0k}))
- Die Residualvarianz σ^2

Die festen Effekte werden relativ präzise geschätzt (wahre Werte: $\gamma_{00} = 1$, $\gamma_{10} = 0{,}6$) und als signifikant beurteilt:

Schätzungen fester Parameter[a]

						Konfidenzintervall 95%	
Parameter	Schätzung	Std.-Fehler	Freiheitsgrade	T-Statistik	Sig.	Untergrenze	Obergrenze
Konstanter Term	1,103623	,125868	56,123	8,768	,000	,851492	1,355754
sector	,403928	,156978	28,158	2,573	,016	,082454	,725401

a. Abhängige Variable: y.

Die Varianzen der drei Zufallskomponenten (Residuum und zwei Zufallseffekte werden ebenfalls relativ präzise geschätzt (wahre Werte: $\sigma^2 = 1$, Var(U_{0j}) = Var(V_{0k}) = 0,25) und als signifikant beurteilt:

Schätzungen von Kovarianzparametern[a]

						Konfidenzintervall 95%	
Parameter		Schätzung	Std.-Fehler	Wald Z	Sig.	Untergrenze	Obergrenze
Residuum		,988617	,021599	45,771	,000	,947177	1,031871
Konstanter Term [Subjekt = j]	Varianz	,229380	,062242	3,685	,000	,134767	,390416
Konstanter Term [Subjekt = k]	Varianz	,156534	,043777	3,576	,000	,090481	,270806

a. Abhängige Variable: y.

Die Varianz des Residuums wird präziser geschätzt als die Varianzen der Zufallseffekte, weil im Datensatz über die Verteilung der Residuen (bei 4500 Schülern) sehr viel mehr Information vorhanden ist als über die Verteilung der Zufallseffekte (bei 30 Gymnasien bzw. Stadtteilen).

Weil die Wald-Tests zu den Zufallseffekten sehr deutlich ausfallen, kann eigentlich auf Likelihood-Quotienten - Tests verzichtet werden (vgl. Abschnitt 4.2). Allerdings interessiert das Verhalten der Schät-

zer, wenn z. B. der ortsbezogene Zufallseffekt entfernt wird. Der Devianzwert des bisher betrachteten korrekten Modells

Informationskriterien[a]

Eingeschränkte -2 Log Likelihood	12205,373
Akaike-Informationskriterium (AIC)	12211,373
Hurvich und Tsai (IC)	12211,379
Bozdogan-Kriterium (CAIC)	12233,435
Bayes-Kriterium von Schwarz (BIC)	12230,435

Die Informationskriterien werden in einem möglichst kleinen Format angezeigt.

a. Abhängige Variable: y.

steigt nach der Entfernung des ortsbezogenen Zufallseffekts

```
mixed y with sector
  /fixed = sector
  /random = intercept | subject(j)
  /print = solution testcov.
```

gewaltig an, sodass sich eine Beurteilung der Differenz durch Vergleich mit der χ_1^2-Verteilung (Erwartungswert 1) erübrigt:

Informationskriterien[a]

Eingeschränkte -2 Log Likelihood	12697,966
Akaike-Informationskriterium (AIC)	12701,966
Hurvich und Tsai (IC)	12701,969
Bozdogan-Kriterium (CAIC)	12716,674
Bayes-Kriterium von Schwarz (BIC)	12714,674

Die Informationskriterien werden in einem möglichst kleinen Format angezeigt.

a. Abhängige Variable: y.

Außerdem werden die Varianzen der beiden verbliebenen Zufallskomponenten überschätzt:

Schätzungen von Kovarianzparametern[a]

Parameter		Schätzung	Std.-Fehler	Wald Z	Sig.	Konfidenzintervall 95% Untergrenze	Konfidenzintervall 95% Obergrenze
Residuum		1,131726	,024646	45,918	,000	1,084437	1,181078
Konstanter Term [Subjekt = j]	Varianz	,321014	,086424	3,714	,000	,189392	,544109

a. Abhängige Variable: y.

5.4 Intraklassenkorrelationen

Das durch folgende Syntax

```
mixed y
  /random = intercept | subject(j)
  /random = intercept | subject(k)
  /print = solution testcov
```

geschätzte Basismodell mit zwei gekreuzten Zufallseffekten liefert folgende Varianzschätzer

Schätzungen von Kovarianzparametern[a]

Parameter		Schätzung	Std.-Fehler	Wald Z	Sig.	Konfidenzintervall 95% Untergrenze	Obergrenze
Residuum		,988636	,021600	45,770	,000	,947195	1,031892
Konstanter Term [Subjekt = j]	Varianz	,228448	,061994	3,685	,000	,134214	,388846
Konstanter Term [Subjekt = k]	Varianz	,188507	,051344	3,671	,000	,110531	,321491

a. Abhängige Variable: y.

zur Berechnung von *drei* Intraklassenkorrelationen:

- Intraklassenkorrelation zu den Schulen

$$\frac{0{,}228448}{0{,}988636 + 0{,}228448 + 0{,}188507} = 0{,}16252797$$

- Intraklassenkorrelation zu den Stadtteilen

$$\frac{0{,}188507}{0{,}988636 + 0{,}228448 + 0{,}188507} = 0{,}13411202$$

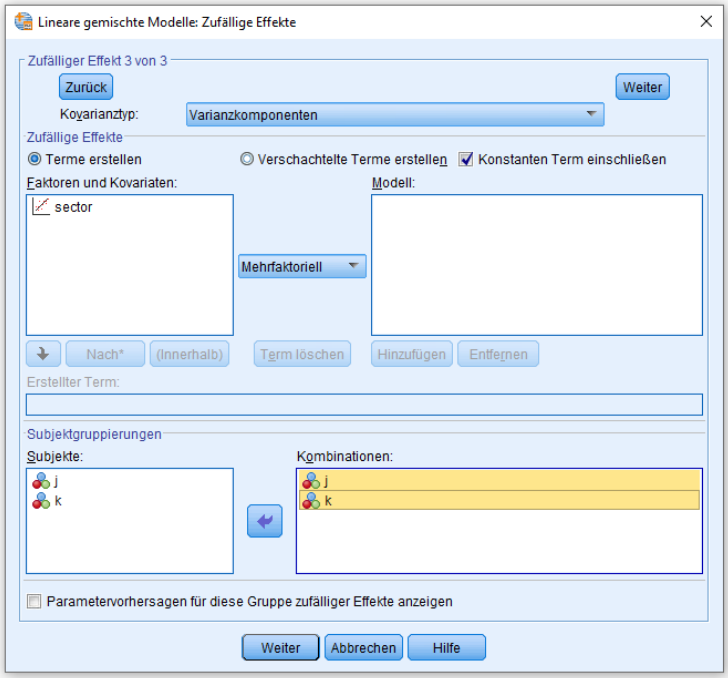
- Intraklassenkorrelation zur Makroebene mit den beiden gekreuzten Zufallseffekten

$$\frac{0{,}228448 + 0{,}188507}{0{,}988636 + 0{,}228448 + 0{,}188507} = 0{,}29663999$$

5.5 Interaktion zwischen gekreuzten Zufallseffekten

Eine mögliche Korrelation zwischen gekreuzten Zufallsfaktoren muss (analog zur Korrelation zwischen zwei Regressoren) *nicht* ins Modell einbezogen werden, eine Interaktion aber schon. In der simulierten Welt des Beispiels ist keine Interaktion zwischen den Zufallsfaktoren Schule und Stadtteil vorhanden. In der realen Welt wäre es aber durchaus denkbar, dass der Effekt bestimmter Schulen vom Wohnort der Schüler abhängt.

Um bei einer Analyseanforderung über die grafische Bedienoberfläche die Interaktion zwischen den Zufallseffekten Schule und Stadtteil ins Modell aufzunehmen, wählt man im Subdialog mit den zufälligen Effekten nach einem Klick auf den Schalter **Weiter** im Rahmen mit den **zufälligen Effekten** (!) die Subjekt-Variablen J und K gemeinsam aus und markiert das Kontrollkästchen für den **konstanten Term**:



Zur manuellen Erstellung einer äquivalenten Syntax ist durch Ausnutzung der passenden Voreinstellungen nur wenig Schreibarbeit erforderlich:

```
mixed y with sector
  /fixed = sector
  /random = intercept | subject(j)
  /random = intercept | subject(k)
  /random = intercept | subject(j*k)
  /print = solution testcov.
```

Zur Modellspezifikation bei gekreuzten Zufallseffekten siehe auch Judd et al. (2017).

Die Schätzung dauert im Beispiel mehrere Minuten und endet mit einer Warnung:

Warnungen

Die endgültige Hesse-Matrix ist nicht positiv definit, obwohl sämtliche Konvergenzkriterien erfüllt sind. Die Prozedur MIXED wird trotz dieser Warnung fortgesetzt. Die Gültigkeit nachfolgender Ergebnisse kann nicht gewährleistet werden.

Verantwortlich für dieses Problem ist wohl die sehr geringe Varianz des aufgenommenen Zufallseffekts (siehe Abschnitt 4.6):

Schätzungen von Kovarianzparametern[a]

Parameter		Schätzung	Standardfehler	Wald Z	Sig.	Konfidenzintervall 95% Untergrenze	Konfidenzintervall 95% Obergrenze
Residuum		,988617	,021599	45,771	,000	,947177	1,031871
Konstanter Term [Subjekt = j]	Varianz	,229380	,062242	3,685	,000	,134767	,390416
Konstanter Term [Subjekt = k]	Varianz	,156534	,043777	3,576	,000	,090481	,270806
Konstanter Term [Subjekt = j * k]	Varianz	,000000 [b]	,000000	.	.	.	.

a. Abhängige Variable: y.

b. Dieser Kovarianzparameter ist überflüssig. Die Teststatistik und das Konfidenzintervall können nicht berechnet werden.

6 Analyse von Messwiederholungsdaten

Wir verwenden nun die HLM-Technik für die Modellierung von Messwiederholungs- bzw. Paneldatensätzen, die ...

- für mehrere Personen oder sonstige Fälle (Makroebene)
- Daten zu mehreren Zeitpunkten bzw. unter mehreren Bedingungen enthalten (Mikroebene).

Ein hierarchisches lineares Modell für Längsschnitt- bzw. Messwiederholungsdaten enthält auf der **Mikroebene** ein individuelles, zufällige Parameter enthaltendes Modell zur Erklärung der zu jedem Zeitpunkt bzw. unter jeder Bedingung erhobenen abhängigen Variablen ...

- durch die Zeit bzw. durch die Stufen des Messwiederholungsfaktors,
- durch weitere Regressoren der Mikroebene, die oft als zeit- bzw. bedingungsabhängige Kovariaten bezeichnet werden (bei Personen z. B. Stimmung oder Müdigkeit zu den einzelnen Zeitpunkten bzw. unter den einzelnen Bedingungen).

In einer pädagogischen Studie mit mehrmaliger Beobachtung von Grundschulkindern könnte z. B. ein lineares Modell für die Wortschatzerweiterung in Abhängigkeit von der Zeit angesetzt werden. Für das Ausgangsniveau und für die Lernrate sind zufällige Parameter (individuelle Werte) anzunehmen.

Die zufälligen Parameter des Mikroebenenmodells werden nach Möglichkeit durch Regressoren auf der **Makroebene** erklärt. Im Beispiel könnte die Abhängigkeit des individuellen Ausgangsniveaus und der individuellen Lernrate von der intellektuellen Anregung und der emotionalen Zuwendung im Elternhaus untersucht werden.

Wir orientieren uns anschließend bei Darstellung und Terminologie an Norušis (2008, S. 227ff), Raudenbush & Bryk (2002, S. 160ff) und Singer (1998, S. 340ff).

6.1 Beispiel 1: Veränderung der Wortbildungsleistung

Der zur Demonstration verwendete Datensatz stammt von Singer & Willett (2003) und beschreibt die Leistungsentwicklung von 35 Personen bei einer Wortbildungsaufgabe (*opposite naming*). Auf der Mikroebene werden folgende Variablen betrachtet:

- TIME
  Die Teilnehmer wurden viermal im Abstand von jeweils einer Woche untersucht. Damit im gleich vorzustellenden Miroebenenmodell dem Ordinatenabschnitt gerade das Ausgangsniveau entspricht, wurde für TIME die Kodierung 0, 1, 2, 3 gewählt. Wie schon im Schulbeispiel von Abschnitt 3 sorgen wir also durch geeignete Rekodierung der Mikroebenenvariablen für relevante Koeffizienten, deren Modellierung auf der Makroebene (mit Personmerkmalen als Regressoren) gut interpretierbare Ergebnisse liefert.
- OPP
  Als Kriterium wird die Leistung im Wortbildungstest betrachtet, der zu jedem Messzeitpunkt durchgeführt wurde. In Modellgleichungen wird das Kriterium wie bisher als *Y* notiert.

Von der Makroebene stammt folgender Regressor:

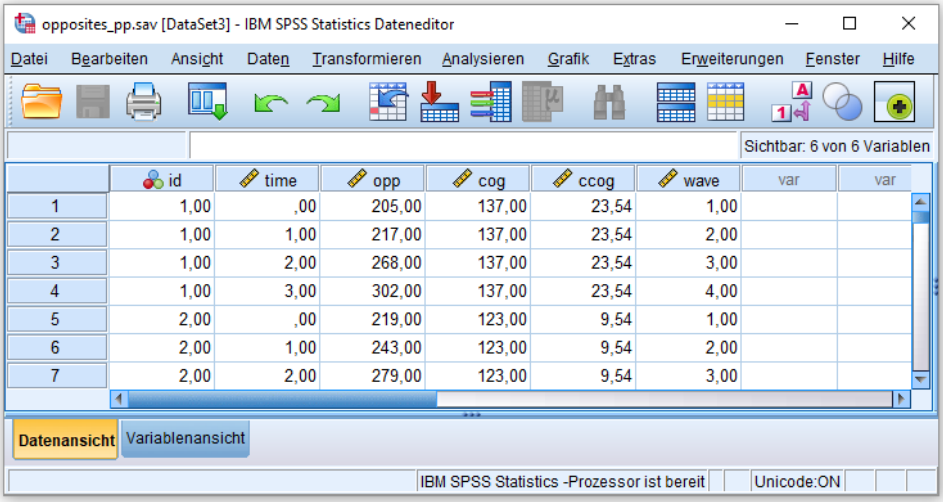
- CCOG
  Zum ersten Messzeitpunkt wurde die allgemeine kognitive Kompetenz erhoben. Analog zum Vorgehen in Abschnitt 3.3.1 verwenden wir die (am Globalmittel) zentrierte Variante des Makroregressors, um die Parameter im Makromodell leichter interpretieren zu können..

Die Daten werden in der Datei **opposites_pp.sav** als Bestandteil des Archives **aldaspss.zip** über den folgenden Link vom *Institute for Digital Research and Education* an der *University of California* (Los Angeles) angeboten:

https://stats.idre.ucla.edu/wp-content/uploads/2016/02/aldaspss.zip

6.2 Benötigte Struktur der SPSS-Arbeitsdatei

SPSS verlangt für die geplanten Analysen eine Arbeitsdatei mit den Person-Zeit - Kombinationen als Zeilen, z. B.:[1]



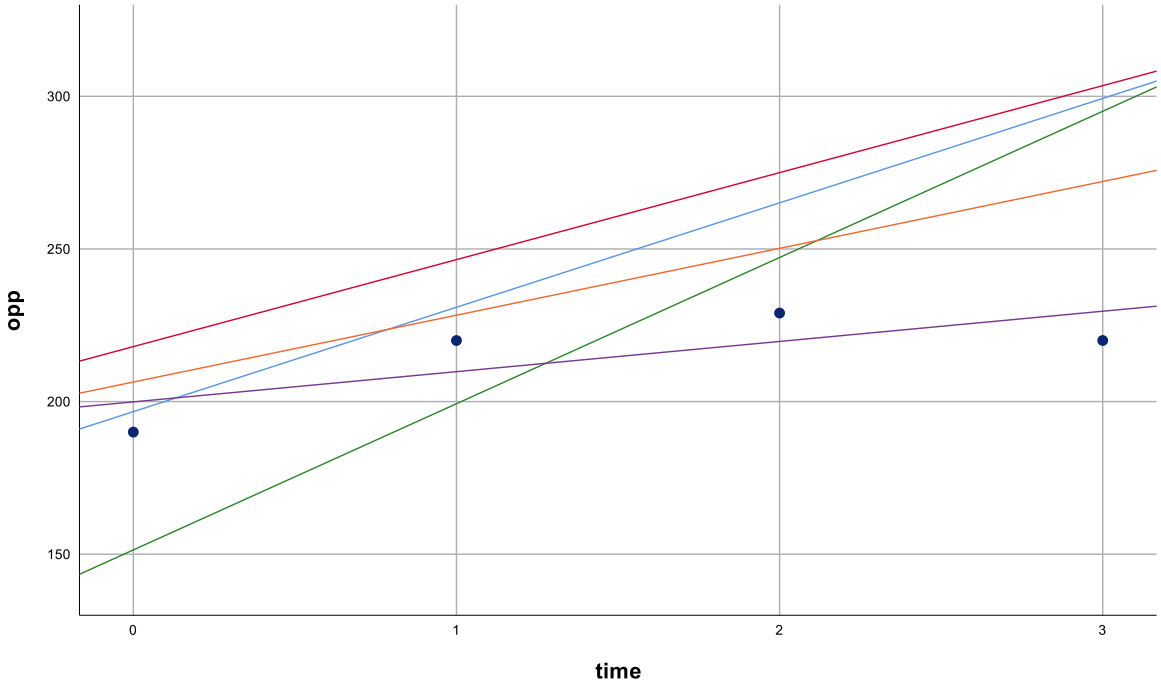
Ggf. müssen Sie also die Struktur Ihrer eigenen Daten anpassen mit dem Kommando VARSTOCASES, das nach

Daten > Umstrukturieren

mit Assistentenhilfe ausgeführt bzw. erstellt werden kann.

6.3 Basismodell mit individuellen Veränderungsparametern

Wir verwenden der Einfachheit halber auf der Ebene 1 (intraindividuell) zur Erklärung des Kriteriums durch die Zeit ein Polynom ersten Grades. In der folgenden Abbildung sind die geschätzten Regressionsgeraden bzw. Modelle nur für die ersten fünf Personen zu sehen, um eine unübersichtliche „Spaghetti- oder Mikado-Abbildung“ zu vermeiden:



[1] In der Datei **opposites_pp.sav** wurden die Messniveaus der Variablen (in der **Variablenansicht** des Dateneditors) manuell korrigiert, was keinen Einfluss auf die beschriebenen Analysen hat.

Es handelt sich schlicht um OLS-Regressionsmodelle (*Ordinary Least Squares*) für die vier Beobachtungen jeder Person. Wie die exemplarisch für den Fall 5 eingetragenen Beobachtungswerte zeigen, kommt für manche intraindividuelle Modelle eventuell ein Polynom *zweiten* Grades in Frage.

Technisch wurde die Abbildung mit SPSS Statistics 26 folgendermaßen realisiert:

- Fallauswahl (**Daten > Fälle auswählen**)
- GRAPH-Kommando

```
GRAPH
  /SCATTERPLOT(BIVAR)=time WITH opp BY id
  /MISSING=LISTWISE.
```

- Diverse Nachbearbeitungen im Grafikeditor:
  - **Elemente > Anpassungslinie bei Untergruppen**
  - Markierung des Kontrollkästchen **Beschriftungen zu Linie** entfernen
  - Rahmen- und Füllungsfarben der Markierungen anpassen
  - **Optionen > Legende ausblenden**
  - Dezimalstellen der Achsenteilstrichbeschriftungen entfernen

6.3.1 Modell

Im Hinblick auf 3-Ebenen-Modelle zur Analyse von *Veränderungen* bei *Individuen* in *Gruppen* haben sich bei hierarchischen linearen Modellen für Längsschnittdaten folgende Bezeichnungskonventionen eingebürgert (siehe Raudenbush & Bryk 2002, S. 162):

- Der Koeffizient zur *p*-ten Potenz der Zeitvariablen beim Fall *i* wird mit π_{pi} bezeichnet.
- Das Mikroresiduum beim Fall *i* zur Zeit *t* wird mit E_{ti} bezeichnet.
- Die Parameter im Makromodell werden durch β's bezeichnet.
  β_{10} ist z. B. der Ordinatenabschnitt im Modell zu π_{1i}.
- Das Makroresiduum zum Fall *i* in der Gleichung zur *p*-ten Potenz der Zeitvariablen wird mit R_{pi} bezeichnet.

Das allgemeine (polynomische) intraindividuelle Modell der Ebene 1 für die Zeitvariable *T* mit dem Zeitindex *t* lautet:

$$Y_{ti} = \pi_{0i} + \pi_{1i}T_{ti} + \pi_{2i}T_{ti}^2 + \ldots + \pi_{Pi}T_{ti}^P + E_{ti} \quad \text{mit } E_{ti} \sim \mathrm{N}(0, \sigma^2)$$

Wir beschränken uns auf den Polynomgrad eins:

$$Y_{ti} = \pi_{0i} + \pi_{1i}T_{ti} + E_{ti}$$

Ein Unterschied zu unserer Analyse von Cluster-Daten (vgl. Abschnitt 3.2) besteht darin, dass hier bereits das Basismodell einen Regressor auf Ebene 1 besitzt, nämlich die Zeit (bzw. das Alter der Personen). Wir erlauben im Modell der Ebene 1 für jede Person individuelle Veränderungskoeffizienten (Ordinatenabschnitt und Steigung, random coefficient model).

Im Modell der Ebene 2 verwenden wir zunächst eine sehr einfache Darstellung für die Erstebenenkoeffizienten:

$$\pi_{0i} = \beta_{00} + R_{0i}$$
$$\pi_{1i} = \beta_{10} + R_{1i}$$

Für die Makroresiduen wird eine bivariate Normalverteilung postuliert:

$$\begin{pmatrix} R_{0i} \\ R_{1i} \end{pmatrix} \sim \mathrm{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}\right]$$

Wir beobachten also die Kovarianzmatrix der Erstebenenparameter und verzichten vorläufig auf Erklärungsversuche durch Makroregressoren.

Für die bivariate Zufallsvariable (R_{0i} , R_{1i}) zu Person *i* wird angenommen, dass sie von den Makroresiduen zu anderen Personen sowie von den Mikroresiduen E_{ti} unabhängig ist.

Durch Einsetzen des Makromodells in das Mikromodell resultiert das kombinierte Modell:

$$\begin{aligned} Y_{ti} &= \beta_{00} + R_{0i} + (\beta_{10} + R_{1i})T_{ti} + E_{ti} \\ &= \beta_{00} + \beta_{10}T_{ti} \quad + \quad R_{0i} + R_{1i}T_{ti} + E_{ti} \end{aligned}$$

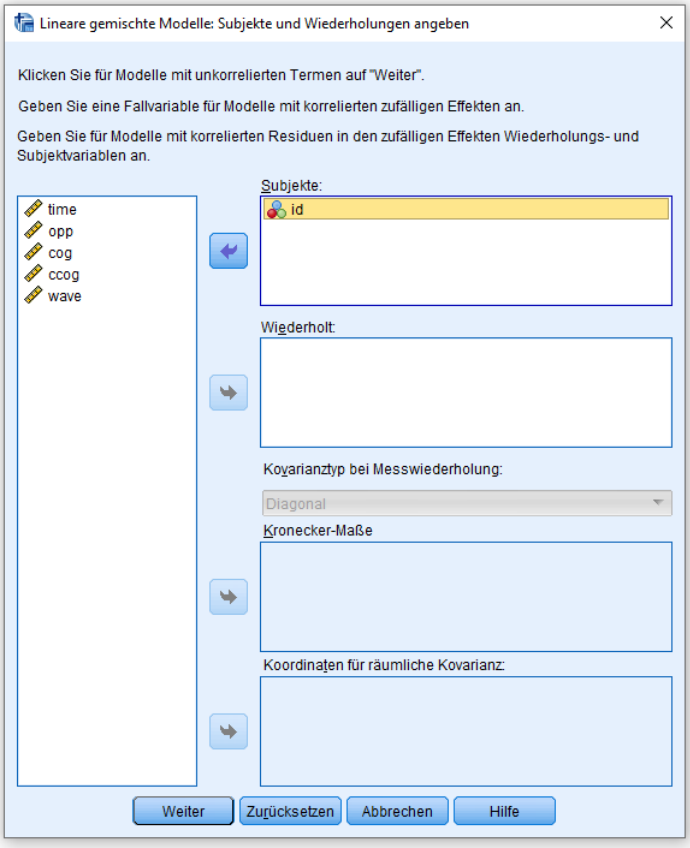
Als Parameter enthält das Modell:

- zwei feste Effekte (β_{00} , β_{01}),
- drei (Ko)varianzen (τ_{00}, τ_{11} , τ_{01}) für die beiden zufälligen Effekte (R_{0i} , R_{1i})
- und die Varianz σ^2 des Mikroresiduums E_{ti} .

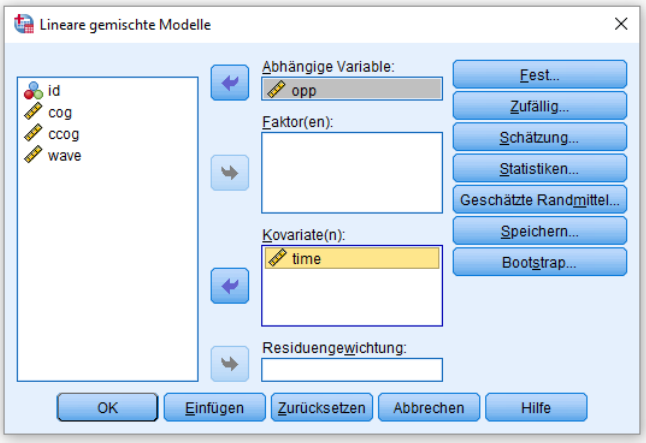
6.3.2 Anforderung der Analyse in SPSS

Gehen Sie folgendermaßen vor, um in SPSS für Windows das Basismodell auf die Daten von Singer & Willett (2003) anzuwenden:

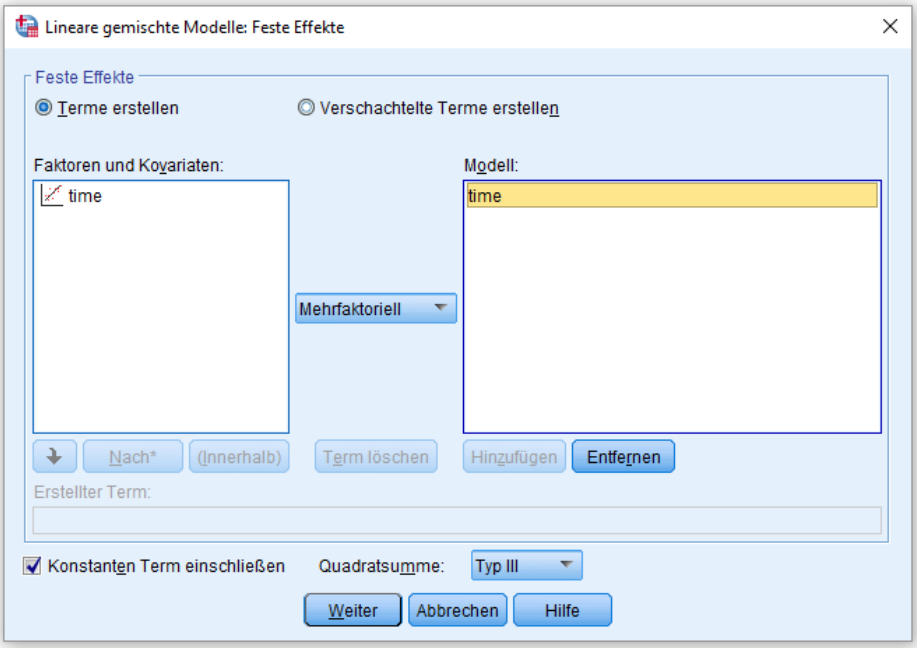
- Menübefehl **Analysieren > Gemischte Modelle > Linear**
- Wir legen fest, dass die Variable ID für die spätere Definition der Makroeinheiten in Frage kommt:



- Wir machen **weiter** und legen OPP als **abhängige Variable** sowie TIME als **Kovariate** fest:

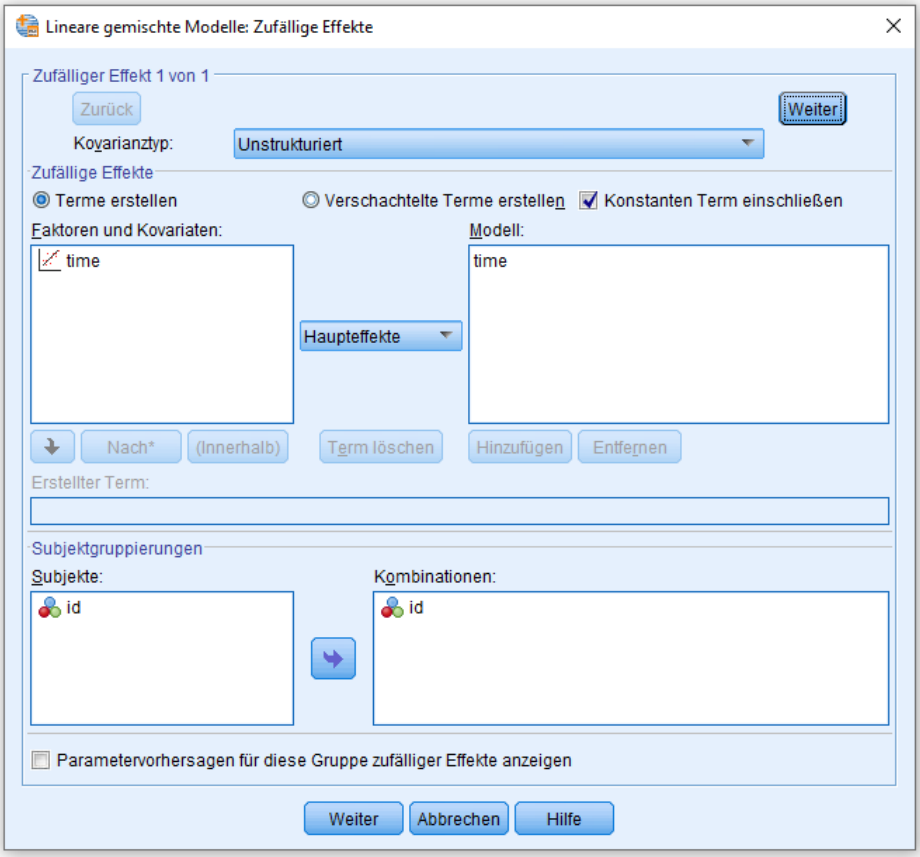


- Wir fügen in folgender Subdialogbox für die **festen Effekte** TIME zum **Modell** hinzu:



Ansonsten wird der Parameter β_{10} für den mittleren Anstieg auf null gesetzt.
Das Kontrollkästchen für den **konstanten Term** muss auf jeden Fall markiert sein (= Voreinstellung). Ansonsten wird der Parameter β_{00} für das mittlere Ausgangsniveau auf null gesetzt.

- Wir quittieren die Subdialogbox für die festen Effekte mit **Weiter** und begeben uns in das Gegenstück für die **zufälligen Effekte**:



Hier sind folgende Einstellungen vorzunehmen:

- Als **Kovarianztyp** wählen wir **Unstrukturiert (UN)**.
  Damit kann die Kovarianzmatrix der zufälligen Effekten (R_{0i} , R_{1i}) frei geschätzt werden.
- Der **konstante Term** wird **eingeschlossen**.
  Ansonsten wird die Varianz τ_{00} der individuellen Ausgangswerte gleich null gesetzt:

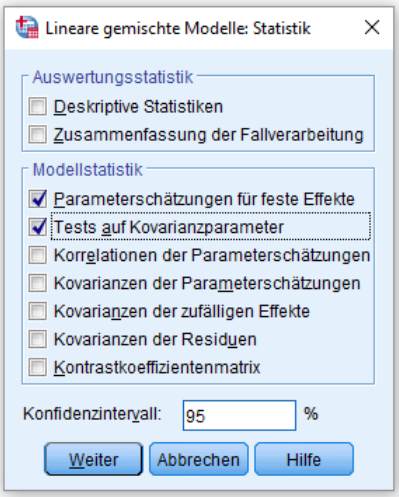
$$\tau_{00} = 0$$
$$\pi_{0i} = \beta_{00}$$

- Die Variable TIME wird ins Modell befördert.
  Ansonsten wird die Varianz τ_{11} der individuellen Steigungen gleich 0 gesetzt:

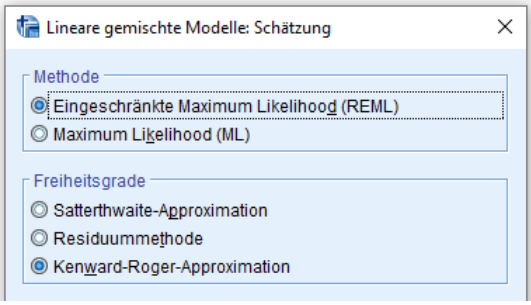
$$\tau_{11} = 0$$
$$\pi_{1i} = \beta_{10}$$

- Die Variable ID wird in die Liste der **Kombinationen** befördert, womit die oben im Dialog **Subjekte und Wiederholungen angeben** vorbereitete Definition der Makroeinheiten abgeschlossen ist.

- Wir fordern im **Statistik**-Subdialog **Parameterschätzer für feste Effekte** und **Tests auf Kovarianzparameter** an:



- Schließlich wählen wir im Subdialog **Schätzung** die **Kenward-Roger-Approximation**, um die Präzision der Signifikanztests zu den festen Effekten zu verbessern:



Über den **Einfügen**-Schalter der Hauptdialogbox erhält man folgendes Syntax-Äquivalent zu der Modellspezifikation:[1]

```
MIXED opp WITH time
  /CRITERIA=DFMETHOD(KENWARDROGER) CIN(95) MXITER(100) MXSTEP(10) SCORING(1)
    SINGULAR(0.000000000001) HCONVERGE(0, ABSOLUTE) LCONVERGE(0, ABSOLUTE)
    PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=time | SSTYPE(3)
  /METHOD=REML
  /PRINT=SOLUTION TESTCOV
  /RANDOM=INTERCEPT time | SUBJECT(id) COVTYPE(UN).
```

Zur manuellen Erstellung einer äquivalenten Syntax ist durch Ausnutzung der passenden Voreinstellungen nur wenig Schreibarbeit erforderlich:

```
mixed  opp with time
  /criteria=dfmethod(kenwardroger)
  /fixed = time
  /random = intercept time  | subject(id) covtype(un)
  /print = solution testcov.
```

Diese Syntax bzw. Modellspezifikation entspricht der in Abschnitt 3.4 behandelten Modellvariante zur Analyse der HS&B - Cluster-Daten (ein Regressor auf Ebene 1, keine Regressoren auf Ebene 2).

---

[1] Von der SPSS-Version 26.0.0.0 (ohne FixPack 1 zu SPSS 26) erhält man per **Einfügen**-Schalter ein fehlerhaftes RANDOM-Subkommando mit ID statt UN als Kovarianztyp:

```
/RANDOM=INTERCEPT time | SUBJECT(id) COVTYPE(ID).
```

Weil damit die Kovarianz der beiden Zufallseffekte aus dem Modell verschwindet, muss das RANDOM-Subkommando unbedingt korrigiert werden. In SPSS 26.0.0.1 (*mit* FixPack 1) ist der Fehler behoben.

6.3.3 Diskussion der Ergebnisse

In der Modellspezifikation

Modelldimension[a]

		Anzahl Ausprägungen	Kovarianzstruktur	Anzahl Parameter	Subjektvariablen
Feste Effekte	Konstanter Term	1		1	
	time	1		1	
Zufallseffekte	Konstanter Term + time [b]	2	Unstrukturiert	3	id
Residuum				1	
Gesamt		4		6	

a. Abhängige Variable: opp.

b. Ab Version 11.5 gelten andere Syntaxregeln für den Unterbefehl RANDOM. Die eingegebene Befehlssyntax führt zu abweichenden Ergebnissen im Vergleich zu früheren Versionen. Wenn Sie Version-11-Syntax verwenden, ziehen Sie das aktuelle Syntax-Referenzhandbuch zu Rate.

protokolliert SPSS sechs Parameter:

- Zwei feste Effekte (β_{00} und β_{10})
- Zwei Varianzen (π_{00} und π_{11}) und eine Kovarianz (π_{01}) für die zufälligen Effekte
- Die Residualvarianz σ^2

In der Tabelle mit den festen Effekten zeigt sich, dass sowohl das mittlere Ausgangsniveau (= 164,37) als auch die mittlere Wachstumsrate (= 26,96) signifikant von null verschieden sind:

Schätzungen fester Parameter[a]

						Konfidenzintervall 95%	
Parameter	Schätzung	Std.-Fehler	Freiheitsgrade	T-Statistik	Sig.	Untergrenze	Obergrenze
Konstanter Term	164,374286	6,118849	34	26,864	,000	151,939289	176,809282
time	26,960000	2,166604	34	12,443	,000	22,556932	31,363068

a. Abhängige Variable: opp.

Der Tabelle mit den Kovarianzparametern ist zu entnehmen, dass die individuellen Ausgangsniveaus bzw. Wachstumsraten jeweils über eine signifikante Varianz verfügen (Bezeichnung `UN(1,1)` bzw. `UN(2,2)`), die durch erweiterte Modelle der Ebene 2 (mit Personmerkmalen als Regressoren) erklärt werden sollte:

Schätzungen von Kovarianzparametern[a]

						Konfidenzintervall 95%	
Parameter		Schätzung	Std.-Fehler	Wald Z	Sig.	Untergrenze	Obergrenze
Residuum		159,477143	26,956557	5,916	,000	114,503596	222,114938
Konstanter Term + time [Subjekt = id]	UN (1,1)	1198,776790	318,380967	3,765	,000	712,310329	2017,471505
	UN (2,1)	-179,255563	88,963416	-2,015	,044	-353,620655	-4,890471
	UN (2,2)	132,400571	40,210696	3,293	,001	73,008902	240,106491

a. Abhängige Variable: opp.

Außerdem ist eine signifikante Kovarianz der Zufallseffekte mit negativem Vorzeichen festzustellen (Bezeichnung: `UN(2,1)`): Je höher der Ausgangswert, desto flacher der Anstieg.

6.4 Regressoren auf der Makroebene

6.4.1 Modell

Wir verwenden dasselbe Mikroebenenmodell

$$Y_{ti} = \pi_{0i} + \pi_{1i} T_{ti} + E_{ti} \quad \text{mit} \quad E_{ti} \sim \mathrm{N}(0, \sigma^2)$$

wie in Abschnitt 6.3.1, versuchen aber nun auf der Makroebene eine Erklärung der Erstebenenkoeffizienten mit Hilfe des Personmerkmals CCOG (zentrierte allgemeine kognitive Kompetenz zum ersten Beobachtungszeitpunkt, Abkürzung: *C*):

$$\pi_{0i} = \beta_{00} + \beta_{01} C + R_{0i}$$
$$\pi_{1i} = \beta_{10} + \beta_{11} C + R_{1i}$$

Weil CCOG zentriert wurde, stehen β_{00} bzw. β_{10} für das mittlere Ausgangsniveau bzw. für die mittlere Steigung in der Regression von OPP auf die Zeit.

Für die Residuen (R_{0i} , R_{1i}) des Makroebenenmodells resultiert eine neue Bedeutung, doch bleibt es bei den Annahmen:

- bivariate Normalverteilung innerhalb einer Person
- Unabhängigkeit von den Makroresiduen zu anderen Personen sowie von den Mikroresiduen E_{ti}

Durch Einsetzen des Makromodells in das Mikromodell ergibt sich das kombinierte Modell:

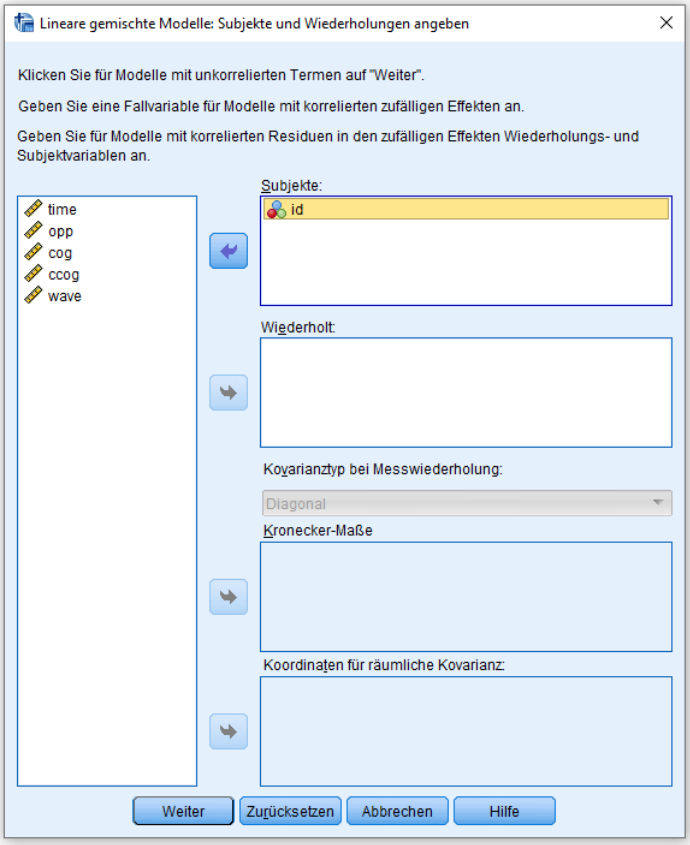
$$\begin{aligned} Y_{ti} &= \beta_{00} + \beta_{01} C + R_{0i} + (\beta_{10} + \beta_{11} C + R_{1i}) T_{ti} + E_{ti} \\ &= \beta_{00} + \beta_{01} C + \beta_{10} T_{ti} + \beta_{11} C T_{ti} \quad + \quad R_{0i} + R_{1i} T_{ti} + E_{ti} \end{aligned}$$

Es enthält vier feste Parameter (β_{00} , β_{01} , β_{10} , β_{11}), drei Kovarianzparameter (τ_{00}, τ_{11} , τ_{01}) für die beiden zufälligen Effekte (R_{0i} , R_{1i}) und den Varianzparameter σ^2 für das Mikroresiduum E_{ti} .

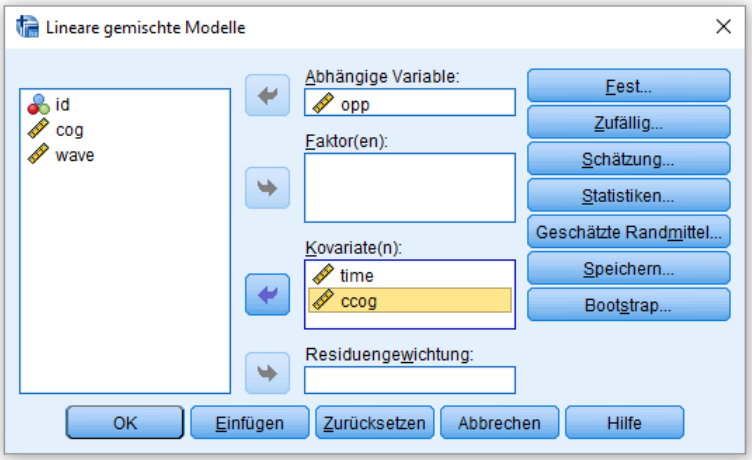
6.4.2 Anforderung der Analyse in SPSS

Gehen Sie folgendermaßen vor, um in SPSS das Modell mit einem Makroregressor auf die Daten von Singer & Willett (2003) anzuwenden:

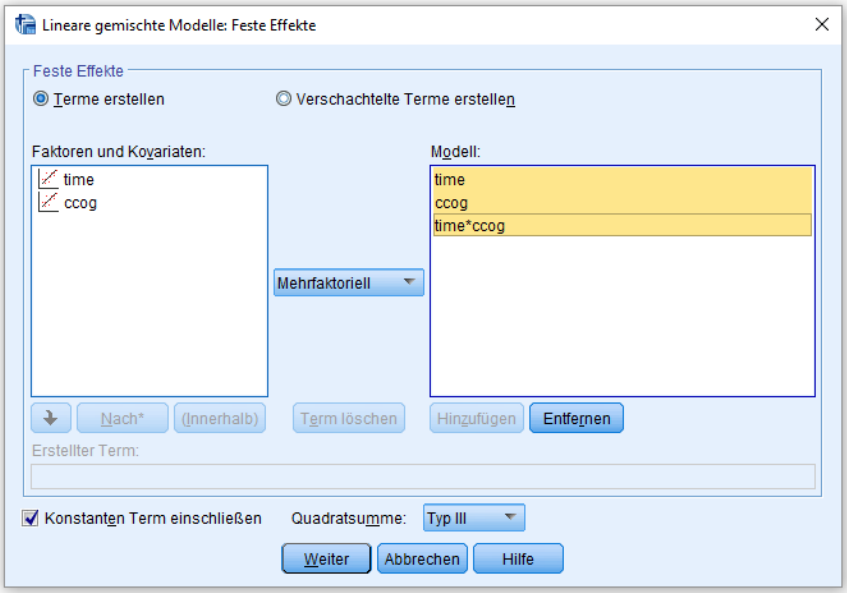
- Menübefehl **Analysieren > Gemischte Modelle > Linear**
- Wir legen erneut fest, dass die Variable ID für die spätere Definition der Makroeinheiten in Frage kommt:



- Wir machen **weiter** und legen OPP als **abhängige Variable** sowie TIME und CCOG als **Kovariaten** fest:



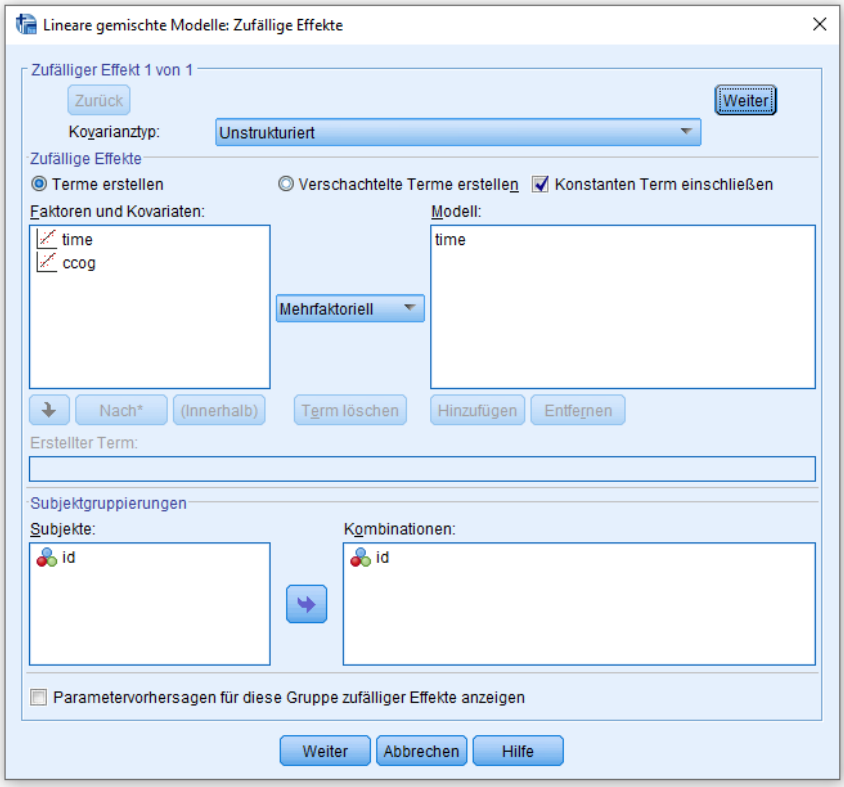
- Wir fügen in der Subdialogbox für die **festen Effekte** TIME, CCOG sowie den Wechselwirkungsterm TIME * CCOG zum **Modell** hinzu:



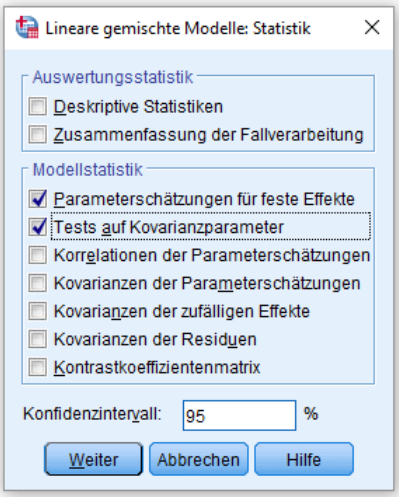
Das Kontrollkästchen für den **konstanten Term** muss auf jeden Fall markiert sein (= Voreinstellung). Somit sind die Parameter zu den vier festen Effekten des Modells deklariert:

- Konstanter Term β_{00}
- CCOG β_{01}
- TIME β_{10}
- CCOG * TIME β_{11}

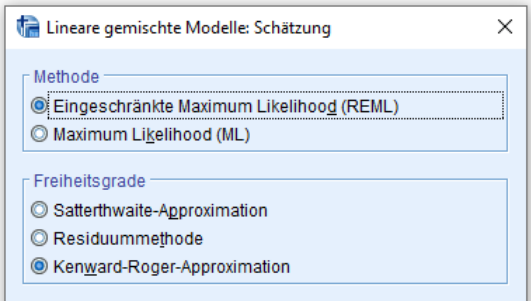
- In der Subdialogbox für die **zufälligen Effekte** sind keine Änderungen gegenüber dem Basismodell erforderlich (vgl. Abschnitt 6.3.2):



- Wir fordern im **Statistik**-Subdialog **Parameterschätzer für feste Effekte** und **Tests auf Kovarianzparameter** an:



- Schließlich wählen wir im Subdialog **Schätzung** die **Kenward-Roger-Approximation**, um die Präzision der Signifikanztests zu den festen Effekten zu verbessern:



Über den **Einfügen**-Schalter der Hauptdialogbox erhält man folgendes Syntax-Äquivalent zu der Modellspezifikation:[1]

```
MIXED opp WITH time ccog
  /CRITERIA=DFMETHOD(KENWARDROGER) CIN(95) MXITER(100) MXSTEP(10) SCORING(1)
    SINGULAR(0.000000000001) HCONVERGE(0, ABSOLUTE) LCONVERGE(0, ABSOLUTE)
    PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=time ccog time*ccog | SSTYPE(3)
  /METHOD=REML
  /PRINT=SOLUTION TESTCOV
  /RANDOM=INTERCEPT time | SUBJECT(id) COVTYPE(UN).
```

Zur manuellen Erstellung einer äquivalenten Syntax ist durch Ausnutzung der passenden Voreinstellungen nur wenig Schreibarbeit erforderlich:

```
mixed  opp with time ccog
  /criteria=dfmethod(kenwardroger)
  /fixed = time ccog time*ccog
  /random = intercept time | subject(id) covtype(un)
  /print = solution testcov.
```

Diese Syntax bzw. Modellspezifikation entspricht der in Abschnitt 3.5 behandelten Modellvariante zur Analyse der HS&B - Cluster-Daten mit Regressoren auf der Mikro- und der Makroebene.

---

[1] Von der SPSS-Version 26.0.0.0 (ohne FixPack 1 zu SPSS 26) erhält man per **Einfügen**-Schalter ein fehlerhaftes RANDOM-Subkommando mit ID statt UN als Kovarianztyp:

```
  /RANDOM=INTERCEPT time | SUBJECT(id) COVTYPE(ID).
```

Weil damit die Kovarianz der beiden Zufallseffekte aus dem Modell verschwindet, muss das RANDOM-Subkommando unbedingt korrigiert werden. In SPSS 26.0.0.1 (*mit* FixPack 1) ist der Fehler behoben.

6.4.3 Diskussion der Ergebnisse

In der Modellspezifikation

Modelldimension[a]

		Anzahl Ausprägungen	Kovarianzstruktur	Anzahl Parameter	Subjektvariablen
Feste Effekte	Konstanter Term	1		1	
	time	1		1	
	ccog	1		1	
	time * ccog	1		1	
Zufallseffekte	Konstanter Term + time [b]	2	Unstrukturiert	3	id
Residuum				1	
Gesamt		6		8	

a. Abhängige Variable: opp.

b. Ab Version 11.5 gelten andere Syntaxregeln für den Unterbefehl RANDOM. Die eingegebene Befehlssyntax führt zu abweichenden Ergebnissen im Vergleich zu früheren Versionen. Wenn Sie Version-11-Syntax verwenden, ziehen Sie das aktuelle Syntax-Referenzhandbuch zu Rate.

protokolliert SPSS die erwarteten acht Parameter:

- Vier Parameter für feste Effekte (β_{00}, β_{01}, β_{10}, β_{11})
- Zwei Varianzen (π_{00}, π_{11}) und eine Kovarianz (π_{01}) für die Zufallseffekte
- Die Residualvarianz σ^2

In der Tabelle mit den festen Effekten

Schätzungen fester Parameter[a]

						Konfidenzintervall 95%	
Parameter	Schätzung	Std.-Fehler	Freiheitsgrade	T-Statistik	Sig.	Untergrenze	Obergrenze
Konstanter Term	164,374291	6,206095	33,000	26,486	,000	151,747895	177,000687
time	26,959981	1,993881	33	13,521	,000	22,903401	31,016562
ccog	-,113553	,504012	33	-,225	,823	-1,138973	,911867
time * ccog	,432858	,161928	33	2,673	,012	,103413	,762302

a. Abhängige Variable: opp.

zeigt sich:

- An den Schätzungen für β_{00} (= 164,37) und β_{01} (= 26,96) hat sich nichts geändert, weil diese Parameter auch im erweiterten Modell für das mittlere Ausgangsniveau bzw. die mittlere Wachstumsrate stehen:
  - In den Regressionsmodellen der Ebene 2
    $$\pi_{0i} = \beta_{00} + \beta_{01}C + r_{0i}$$
    $$\pi_{1i} = \beta_{10} + \beta_{11}C + r_{1i}$$
    stehen β_{00} bzw. β_{10} für den erwarteten Kriteriumswert bei $C = 0$.
  - Weil wir den Makroregressor CCOG zentriert haben, ist 0 gerade sein Mittelwert. Folglich stehen β_{00} bzw. β_{10} für den erwarteten Kriteriumswert bei mittlerer Regressorausprägung.
  - Für den Mittelwert des Regressors wird der Mittelwert des Kriteriums geschätzt, also hier das mittlere Ausgangsniveau bzw. die mittlere Wachstumsrate.
- Die allgemeine kognitive Kompetenz hat (CCOG) keinen signifikanten Effekt auf das *Ausgangsniveau* der abhängigen Variablen ($\hat{\beta}_{01} = -0{,}11$, $p = 0{,}82$). Dies ist nicht der „Haupteffekt“ von CCOG, sondern der bedingte Effekte für TIME = 0.
- CCOG hat einen signifikanten positiven Effekt auf die Lernrate ($\hat{\beta}_{11} = 0{,}43$, $p = 0{,}01$), moderiert also den Zeiteffekt. Je höher die allgemeine kognitive Kompetenz, desto größer ist der Lerngewinn pro Zeiteinheit.

In der Tabelle mit den Kovarianzparametern

Schätzungen von Kovarianzparametern[a]

Parameter		Schätzung	Std.-Fehler	Wald Z	Sig.	Konfidenzintervall 95% Untergrenze	Konfidenzintervall 95% Obergrenze
Residuum		159,477143	26,956557	5,916	,000	114,503596	222,114938
Konstanter Term + time [Subjekt = id]	UN (1,1)	1236,412706	332,402178	3,720	,000	730,000241	2094,131334
	UN (2,1)	-178,233247	85,429778	-2,086	,037	-345,672535	-10,793960
	UN (2,2)	107,249191	34,676704	3,093	,002	56,908407	202,121084

a. Abhängige Variable: opp.

zeigt sich:

- Im Vergleich zum Basismodell hat sich die geschätzte Mikroebenen-Residualvarianz ($\hat{\sigma}^2 = 159,48$) *nicht* geändert, weil beide Modelle für jede Person individuelle Lernparameter (Ausgangsniveau und Steigung) erlauben.
- Wie es aufgrund des insignifikanten $\hat{\beta}_{01}$ - Schätzwertes zu erwarten war, hat sich die Residualvarianz in den Ausgangsniveaus ($\hat{\tau}_{00} = 1236,41$) kaum geändert (Bezeichnung: `UN(1,1)`).
- Demgegenüber ist die Residualvarianz $\hat{\tau}_{11}$ in den individuellen Lernraten (Bezeichnung: `UN(2,2)`) deutlich geschrumpft von 132,40 auf 107,25. Damit kann CCOG 19,0% der Lernratenvarianz erklären.
- Für den nicht erklärbaren Anteil der individuellen Lernrate (R_{1i}) wird eine signifikante negative Kovarianz mit dem Residuum des Ausgangsniveaus (R_{0i}) geschätzt ($\hat{\tau}_{01} = -178,23$, p = 0,04, Bezeichnung: `UN(2,1)`). Weil sich das Mikroebenenwachstumsmodell auf mehr als zwei Zeitpunkten stützt, kommt die Regression zum Mittelwert als Erklärung für die negative Kovarianz *nicht* in Frage (vgl. Raudenbush & Bryk 2003, S. 161). Vielleicht zeigen sich hier die Effekte einer bedarfsorientierten pädagogischen Förderung.

6.5 Kovarianzmatrix der Residuen als Modellierungsoption

Ist (wie im Singer/Willett - Beispiel) für alle Subjekte derselbe Erhebungsplan realisiert worden, dann kommt ein direktes Modellieren der Kovarianzstruktur der Residuen in Frage, wobei die bisher vorgestellte Modellierung (durch zufällige Effekte) ersetzt oder ergänzt werden kann. Die in Abschnitt 4.7 vorgestellte Notation für gemischte Modelle hilft, Verwechslungen zwischen der aktuell diskutierten Kovarianzmatrix der Residuen und der im bisherigen Verlauf des Manuskripts oft behandelten Kovarianzmatrix der Zufallseffekte zu vermeiden. Für die vektorielle Zufallsgröße **Y** behauptet das Modell in Matrixschreibweise:

$$\mathbf{Y} = \mathbf{X\beta} + \mathbf{ZU} + \mathbf{E}$$

Darin sind u.a. enthalten:

U ($q \times 1$) - Vektor der zufälligen Effekte
Es wird angenommen, dass die zufälligen Effekte den Erwartungswert 0 besitzen und gemeinsam normalverteilt sind mit der Kovarianzmatrix **G**, kurz: **U** ~ N(0, **G**)

E ($N \times 1$) - Vektor der Individualresiduen
Es wird angenommen, dass die Individualresiduen den Erwartungswert 0 besitzen und gemeinsam normalverteilt sind mit der Kovarianzmatrix **R**, kurz: **E** ~ N(0, **R**)

Für **Y** resultiert eine Normalverteilung mit dem Erwartungswertevektor **Xβ** und einer Kovarianzmatrix **V**, die sich folgendermaßen aus den Kovarianzmatrizen **G** und **R** berechnen lässt:

$$\mathbf{V} = \mathbf{ZGZ'} + \mathbf{R}$$

In unseren bisherigen Modellen war **R** die mit σ^2 vormultiplizierte N-dimensionale Einheitsmatrix:

$$\mathbf{R} = \sigma^2 \mathbf{I}_N$$

Folglich leistete **R** keinen Beitrag zu den Kovarianzen der Beobachtungen. Nun werden wir die Matrix **R** stärker in die Modellformulierung einbeziehen:

- In Abschnitt 6.5.1 wird gezeigt, dass man bestimmte Zufallseffekte durch die Spezifikation einer speziellen Struktur für die Matrix **R** ersetzen und dabei zu einem äquivalenten Modell gelangen kann.
- In Abschnitt 6.5.2 wird gezeigt, wie man Zufallseffekte (mit einer Kovarianzmatrix **G**) durch eine Residual-Kovarianzmatrix **R** mit bestimmter Struktur ergänzen und so die Flexibilität gemischter Modelle ausnutzen kann.

Bei der Modellierung von Längsschnittdaten mit einem für alle Fälle identischen Erhebungsplan ist in der Regel eine direkte Spezifikation der Residual-Kovarianzmatrix **R** sehr zu empfehlen. In dieser Situation hängt die Korrelation der von einem Fall stammenden Beobachtungen von der Zeitdistanz ab, und mit Zufallseffekten ist diese Zeitabhängigkeit nicht zu modellieren.

Individuelle Erhebungspläne haben individuelle Kovarianzstrukturen zur Folge, und die Spezifikation einer bestimmten (notwendigerweise für alle Fälle *gemeinsamen*) Kovarianzstruktur ist weniger sinnvoll. Fehlende Werte stellen für das Modellieren der Kovarianzstruktur jedoch *kein* Problem dar. In dieser speziellen Hinsicht ist also individuelles Abweichen vom Plan erlaubt.

6.5.1 Spezifikation der Fehlerkovarianzstruktur an Stelle von Zufallseffekten

Zunächst soll am Beispiel eines Modells mit einem festen Zeiteffekt und einem zufälligen Ordinatenabschnitt demonstriert werden, dass dieses hierarchische Regressionsmodell eine zusammengesetzt-symmetrische Struktur für die Kovarianzmatrix der Beobachtungen impliziert und somit äquivalent ist zu einem Modell, das unter Verzicht auf Zufallseffekte eine zusammengesetzt-symmetrische Struktur für die Kovarianzmatrix der Residuen spezifizieren. Wir betrachten das folgende Mikromodell mit der Zeitvariablen *T* und der üblichen Unabhängigkeitsannahme für die Mikroresiduen:

$$Y_{ti} = \pi_{0i} + \pi_{1i} T_{ti} + E_{ti}$$

Auf der Makroebene nehmen wir individuelle, zufällig variierende Ordinatenabschnitte (bei passender Kodierung: Ausgangsniveaus) an, aber eine *konstante* Steigung:

$$\pi_{0i} = \beta_{00} + R_{0i}$$
$$\pi_{1i} = \beta_{10}$$

Durch Einsetzen des Makromodells in das Mikromodell resultiert das kombinierte Modell

$$\begin{aligned} Y_{ti} &= \beta_{00} + R_{0i} + \beta_{01} T_{ti} + E_{ti} \\ &= \beta_{00} + \beta_{01} T_{ti} + R_{0i} + E_{ti} \end{aligned}$$

Für die Kovarianzmatrix der Beobachtungen Y_{ti} erhalten wir die schon in Abschnitt 3.2.1 beschriebene zusammengesetzt-symmetrische Struktur:

- Alle Beobachtungen besitzen dieselbe Varianz:
$$\mathrm{Var}(Y_{ti}) = \mathrm{Var}(R_{0i}) + \mathrm{Var}(E_{ti}) = \tau_{00} + \sigma^2$$
- Für Beobachtungen zum selben Subjekt ergibt sich als Kovarianz:
$$\mathrm{Kov}(Y_{ti}, Y_{t'i}) = \tau_{00}$$
- Für Beobachtungen zu verschiedenen Subjekten ist die Kovarianz gleich null.

Bei einer Studie mit identischem Messwiederholungplan für alle Subjekte kann das eben beschriebene Modell (mit festem Zeiteffekt und zufälligem Ordinatenabschnitt) äquivalent ersetzt werden durch ein Modell, …

- das auf einen Zufallseffekt verzichtet
- und stattdessen eine zusammengesetzt-symmetrische Kovarianzstruktur der Residuen E_{ti} behauptet.

Dies soll an einem simulierten Datensatz[1] mit 150 Fällen und vier Messungen (kodiert mit 0, 1, 2, 3) demonstriert werden, wobei in der simulierten Population die folgenden Parameterwerte gelten:[2]

- mittleres Ausgangsniveau: 0
- feste Steigung: 0,4
- Varianz τ_{00}: 1
- Fehlervarianz σ^2: 1

Die mit folgender Syntax

```
mixed  y  with zeit
  /fixed = zeit
  /random = intercept | subject(id)
  /print = solution testcov.
```

nach bisherigem Muster angeforderte MIXED-Analyse (mit zufälligem Ordinatenabschnitt) liefert gute Schätzwerte für die festen Effekte:[3]

Schätzungen fester Parameter[a]

Parameter	Schätzung	Standardfehler	Freiheitsgrade	T-Statistik	Sig.	Konfidenzintervall 95% Untergrenze	Konfidenzintervall 95% Obergrenze
Konstanter Term	,130084	,107412	253,461	1,211	,227	-,081451	,341618
zeit	,390760	,035556	449,000	10,990	,000	,320883	,460637

a. Abhängige Variable: y.

Auch die beiden Residualvarianzen werden gut geschätzt:

Schätzungen von Kovarianzparametern[a]

Parameter		Schätzung	Standardfehler	Wald Z	Sig.	Konfidenzintervall 95% Untergrenze	Konfidenzintervall 95% Obergrenze
Residuum		,948176	,063282	14,983	,000	,831915	1,080684
Konstanter Term [Subjekt = id]	Varianz	1,066887	,151896	7,024	,000	,807106	1,410283

a. Abhängige Variable: y.

Anschließend wird beschrieben, wie man dieselben Ergebnisse durch die direkte Spezifikation einer zusammengesetzt-symmetrischen Kovarianzmatrix der Residuen (aber *ohne* Zufallseffekt) erzielt:

- Menübefehl **Analysieren > Gemischte Modelle > Linear**
- Wir legen fest, dass die **Subjekte** (Makroeinheiten) über die Variable ID und die **wiederholten** Messungen über die Variable ZEIT definiert werden. Außerdem wählen wir eine **zusammengesetzt-symmetrische** Struktur für die **Kovarianzmatrix bei Messwiederholung** (für die Matrix **R**):

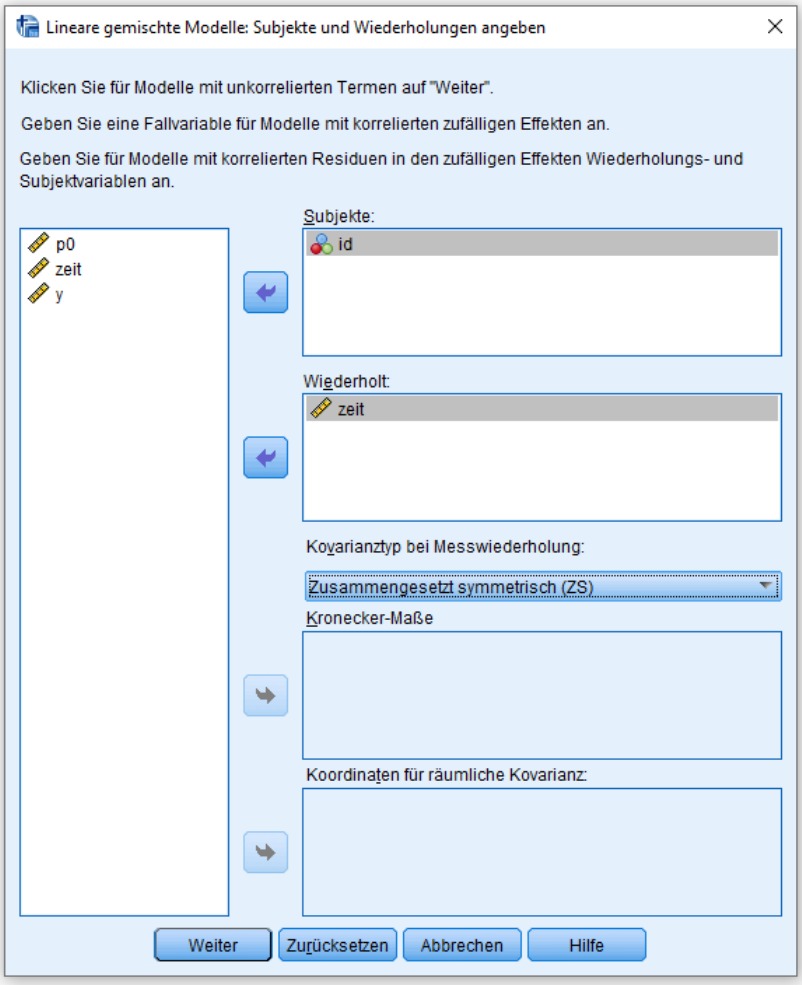
---

1 Die Singer/Willett - Daten eignen sich nicht für die geplante Demonstration, weil hier eine signifikante Varianz der individuellen Wachstumsraten vorliegt (siehe Abschnitt 6.3.3).

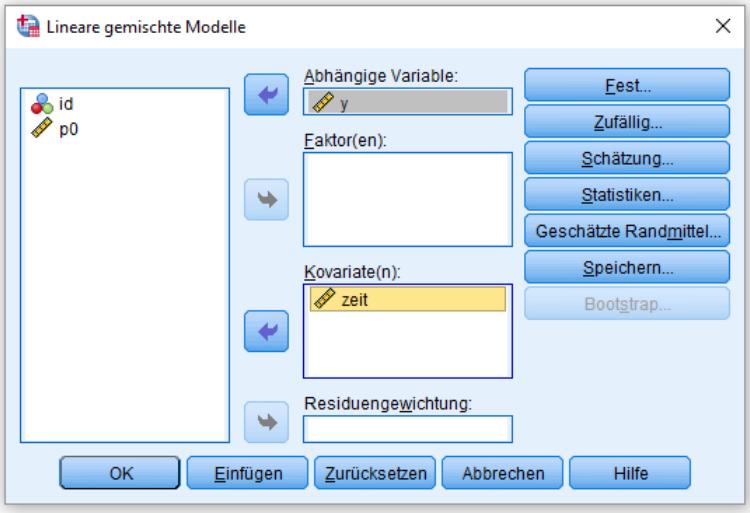
2 An der im Vorwort vereinbaren Stelle finden Sie im Ordner **CS** die folgenden Dateien:

- **CS.sps** (mit dem erzeugenden SPSS-Programm)
- **CS.sav** (Datensatz für einen bestimmten Startwert des Pseudozufallszahlengenerators)
- **CS.spv** (die gleich zu besprechenden Ausgaben)

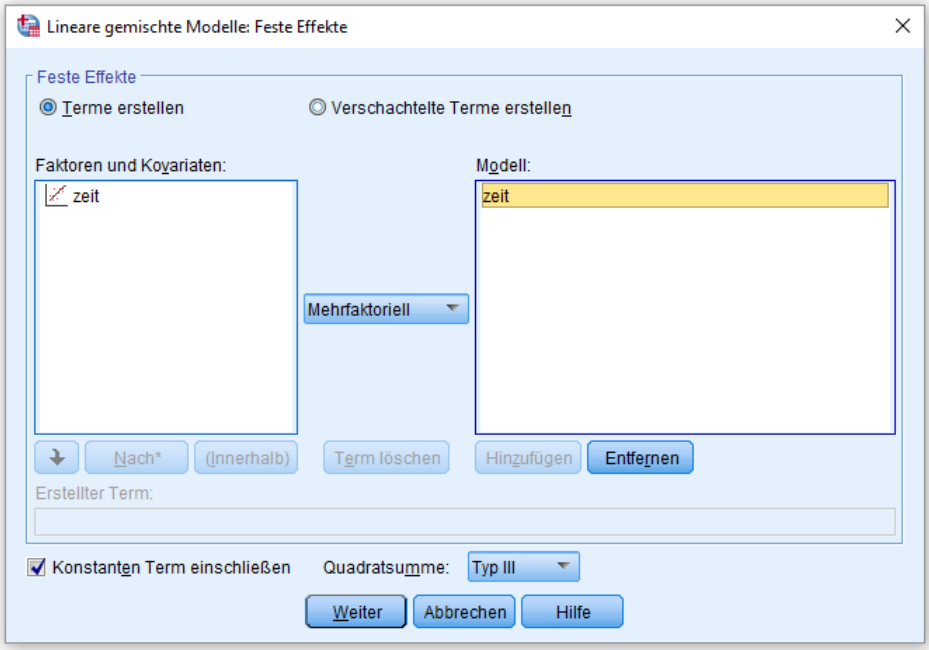
3 In Abschnitt 6.5 verzichten wir der Einfachheit halber auf die **Kenward-Roger - Approximation**, weil sich fast kein Unterschied im Vergleich zur voreingestellten **Satterthwaite-Approximation** zeigt.



- Wir machen **weiter** und legen Y als **abhängige Variable** sowie ZEIT als **Kovariate** fest:

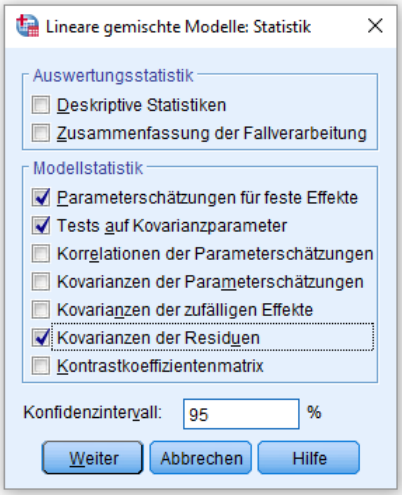


- Wir fügen in der Subdialogbox für die **festen Effekte** ZEIT zum **Modell** hinzu:



  Das Kontrollkästchen für den **konstanten Term** muss auf jeden Fall markiert sein (= Voreinstellung).

- Weil das Modell ohne zufällige Effekte auskommt, muss die zugehörige Subdialogbox nicht aufgesucht werden.
- Schließlich fordern wir in der **Statistik**-Subdialogbox **Parameterschätzer für feste Effekte**, **Tests auf Kovarianzparameter** und die **Kovarianzen der Residuen** an:



Über den **Einfügen**-Schalter der Hauptdialogbox erhält man folgendes Syntax-Äquivalent zu der Modellspezifikation:

```
MIXED y WITH zeit
  /CRITERIA=DFMETHOD(SATTERTHWAITE) CIN(95) MXITER(100) MXSTEP(10) SCORING(1)
    SINGULAR(0.000000000001) HCONVERGE(0, ABSOLUTE) LCONVERGE(0, ABSOLUTE)
    PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=zeit | SSTYPE(3)
  /METHOD=REML
  /PRINT=R SOLUTION TESTCOV
  /REPEATED=zeit | SUBJECT(id) COVTYPE(CS).
```

Zur manuellen Erstellung einer äquivalenten Syntax ist durch Ausnutzung der passenden Voreinstellungen nur wenig Schreibarbeit erforderlich:

```
mixed  y  with zeit
  /fixed = zeit
  /repeated = zeit  | subject(id) covtype(cs)
  /print = solution testcov r.
```

Wir erhalten exakt dieselbe Tabelle für die festen Effekte wie bei der Modellspezifikation mit zufälligem Ordinatenabschnitt:

Schätzungen fester Parameter[a]

Parameter	Schätzung	Std.-Fehler	Freiheitsgrade	T-Statistik	Sig.	Konfidenzintervall 95% Untergrenze	Konfidenzintervall 95% Obergrenze
Konstanter Term	,130084	,107412	253,461	1,211	,227	-,081451	,341618
zeit	,390760	,035556	449,000	10,990	,000	,320883	,460637

a. Abhängige Variable: y.

Auch die Tabelle mit den Kovarianzparametern enthält (bei angepasster Notation) identische Schätzergebnisse:

Schätzungen von Kovarianzparametern[a]

Parameter		Schätzung	Std.-Fehler	Wald Z	Sig.	Konfidenzintervall 95% Untergrenze	Konfidenzintervall 95% Obergrenze
Meßwiederholungen	CS, diagonaler Offset	,948176	,063282	14,983	,000	,831915	1,080684
	CS, Kovarianz	1,066887	,151896	7,024	,000	,769178	1,364597

a. Abhängige Variable: y.

In der folgenden Tabelle wird die intraindividuelle Kovarianzmatrix präsentiert:

R-Matrix[a]

	[zeit = ,00]	[zeit = 1,00]	[zeit = 2,00]	[zeit = 3,00]
[zeit = ,00]	2,015063	1,066887	1,066887	1,066887
[zeit = 1,00]	1,066887	2,015063	1,066887	1,066887
[zeit = 2,00]	1,066887	1,066887	2,015063	1,066887
[zeit = 3,00]	1,066887	1,066887	1,066887	2,015063

Zusammengesetzte Symmetrie

a. Abhängige Variable: y.

6.5.2 Alternative Fehlerkovarianzstrukturen

In Abschnitt 6.5.1 wurde gezeigt, wie ein spezielles Mehrebenenmodell für Längsschnittdaten (mit festem Zeiteffekt und zufälligem Ordinatenabschnitt) alternativ (und äquivalent) über eine bestimmte Struktur für die Residual-Kovarianzmatrix **R** (zusammengesetzte Symmetrie) formuliert werden kann. In der Prozedur MIXED lassen sich neben der zusammengesetzten Symmetrie etliche weitere Strukturen für die Matrix **R** angeben, die (eigenständig oder gemeinsam mit Zufallseffekten) zur realistischen Modellierung dienen können, z. B.:

- **AR(1)**
  Für die Residuen zu äquidistanten Messzeitpunkten bietet sich oft ein autoregressiver Prozess erster Ordnung mit der Autokorrelation ρ an. Abgesehen von einem konstanten Faktor besitzen dann zwei Messungen mit dem zeitlichen Abstand v die Kovarianz ρ^v. Außerdem sind die Varianzen zu allen Messzeitpunkten gleich. Diese Struktur wird gleich in einem Beispiel vorgestellt.

- **Toeplitz**
  Bei äquidistanten Messungen kommt auch die Toeplitz-Struktur in Frage, wobei die Kovarianz zwischen zwei Messungen nur von der zeitlichen Distanz abhängt, aber im Unterschied zur AR(1) - Struktur für jede Zeitdistanz separat geschätzt werden kann. Außerdem sind die Varianzen zu allen Messzeitpunkten gleich.
- **Unstrukturiert**
  Bei hinreichend großer Stichprobe kann man die Kovarianzmatrix (bei k Messungen mit $k(k+1)/2$ Parametern!) frei schätzen lassen, um durch die Exploration ihrer Struktur Aufschlüsse über den datengenerierenden Prozess zu gewinnen. So ist z. B. bei wachsenden Fehlervarianzen ein *random-intercept* - Modell unplausibel, weil es konstante Fehlervarianzen impliziert. In einem solchen Fall kommt eher ein Mehrebenenmodell mit zufälligen Ausgangsniveaus *und* zufälligen Wachstumsraten in Frage, weil es steigende Varianzen der Residuen E_{ti} impliziert. Sinkt die Fehlerkorrelation mit zunehmender zeitlicher Distanz, bietet sich ein autoregressiver Prozess zur Modellierung an.

Eine leichte Fehlspezifikation der Kovarianzstruktur wirkt sich vermutlich auf die Schätzung und Testung der *festen* Regressionskoeffizienten nicht gravierend aus (Hox 2010, S. 103).

Im nächsten Beispiel betrachten wir eine simulierte Population mit linearem Wachstum eines Kriteriums Y über vier Messzeitpunkte (kodiert mit 0, 1, 2, 3), wobei zusätzlich gilt:[1]

- identische Wachstumsrate für alle Individuen (fester Zeiteffekt $\beta_{10} = 0{,}4$)
- zufälliger Koeffizient π_{0i} für das Ausgangsniveau mit festem Effekt eines dichotomen Makroregressors GRUPPE ($\beta_{01} = 1{,}0$) und Residualvarianz $\tau_{00} = 2{,}25$
- Mikroebenenresiduum mit konstanter Varianz 1 und einem autoregressiven Prozess erster Ordnung (Parameter $\rho = 0{,}5$)

Die Modelgleichung der Mikroebene können wir aus Abschnitt 6.5.1 übernehmen:

$$Y_{ti} = \pi_{0i} + \pi_{1i} T_{ti} + E_{ti}$$

Allerdings sind die Mikroresiduen E_{ti} innerhalb einer Person nicht unabhängig, sondern verhalten sich wie ein stationärer autoregressiver Prozess erster Ordnung (konstante Varianz, Autokorrelation ρ mit Betrag < 1). E_{ti} hängt von seinem Vorgänger E_{t-1i} und von einem unabhängig identisch normalverteiltem Anteil δ_{ti} ab, der zur Unterscheidung von den Mikroresiduen als *weißes Rauschen* bezeichnet werden soll:

$$E_{ti} = \rho\, E_{t-1i} + \delta_{ti}$$

Das Modell der Makroebene enthält individuelle Ordinatenabschnitte (wegen der passend gewählten Kodierung gleich dem Ausgangsniveaus), die von der Gruppenvariablen (Abk.: C) abhängen, sowie eine feste Steigung:

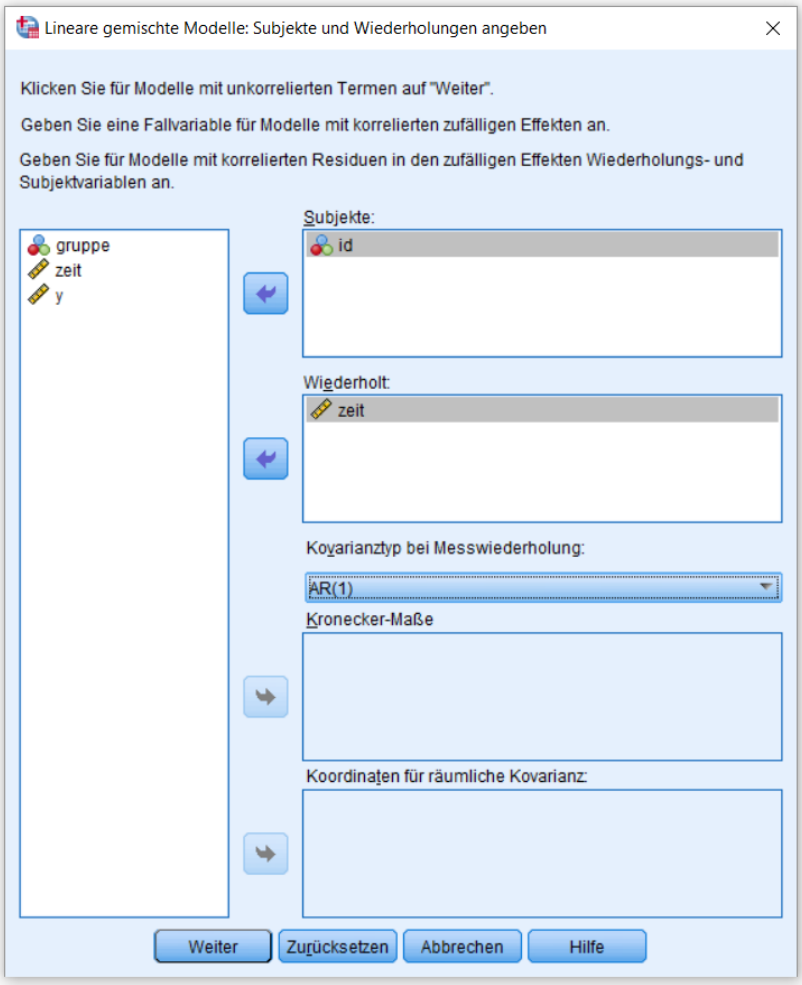
$$\pi_{0i} = \beta_{00} + \beta_{01} C + R_{0i}$$
$$\pi_{1i} = \beta_{10}$$

Anschließend wird die Deklaration des Modells beschrieben:

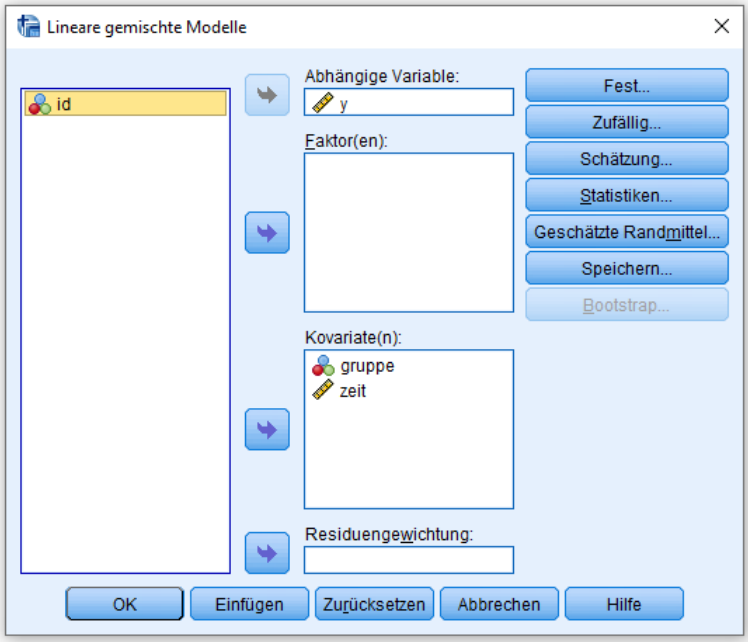
- Menübefehl **Analysieren > Gemischte Modelle > Linear**
- Wir legen fest, dass die **Subjekte** (Makroeinheiten) über die Variable ID und die **wiederholten** Messungen über die Variable ZEIT definiert werden. Außerdem wählen wir eine **AR(1)** - Struktur

[1] An der im Vorwort vereinbaren Stelle finden Sie im Ordner **AR1** die folgenden Dateien:
- **AR1.sps** (mit dem erzeugenden SPSS-Programm)
- **AR1.sav** (Datensatz für einen bestimmten Startwert des Pseudozufallszahlengenerators)
- **AR1.spv** (die gleich zu besprechenden Ausgaben)

für die **Kovarianzmatrix bei Messwiederholung** (für die Mikroebenen-Fehlerkovarianzmatrix **R**):

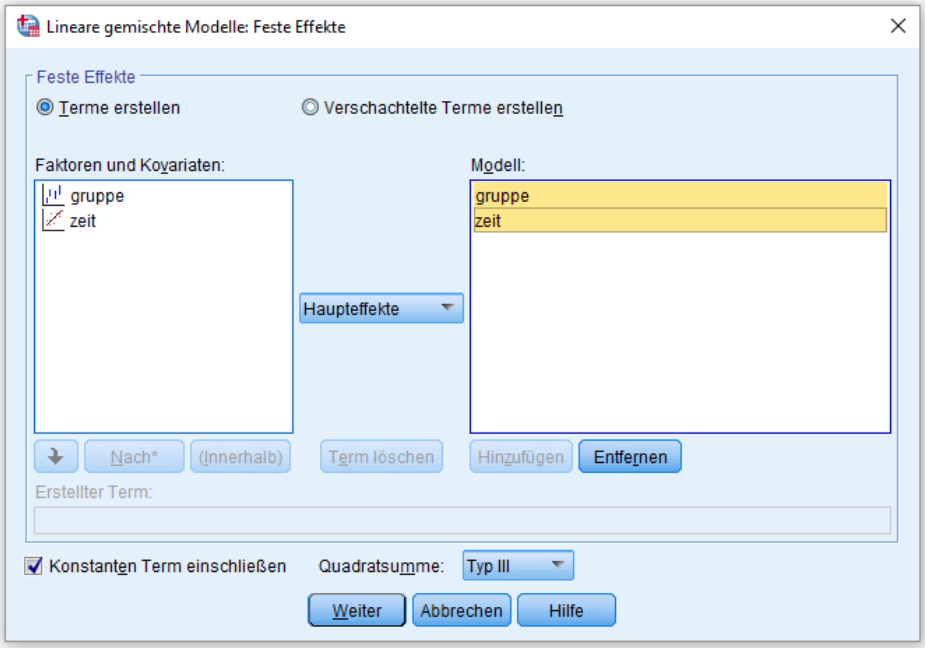


- Wir machen **weiter** und legen Y als **abhängige Variable**, GRUPPE als Faktor sowie ZEIT als **Kovariate** fest:



Weil GRUPPE nur zwei Ausprägungen hat, ist gegen die Behandlung als Kovariate nichts einzuwenden.

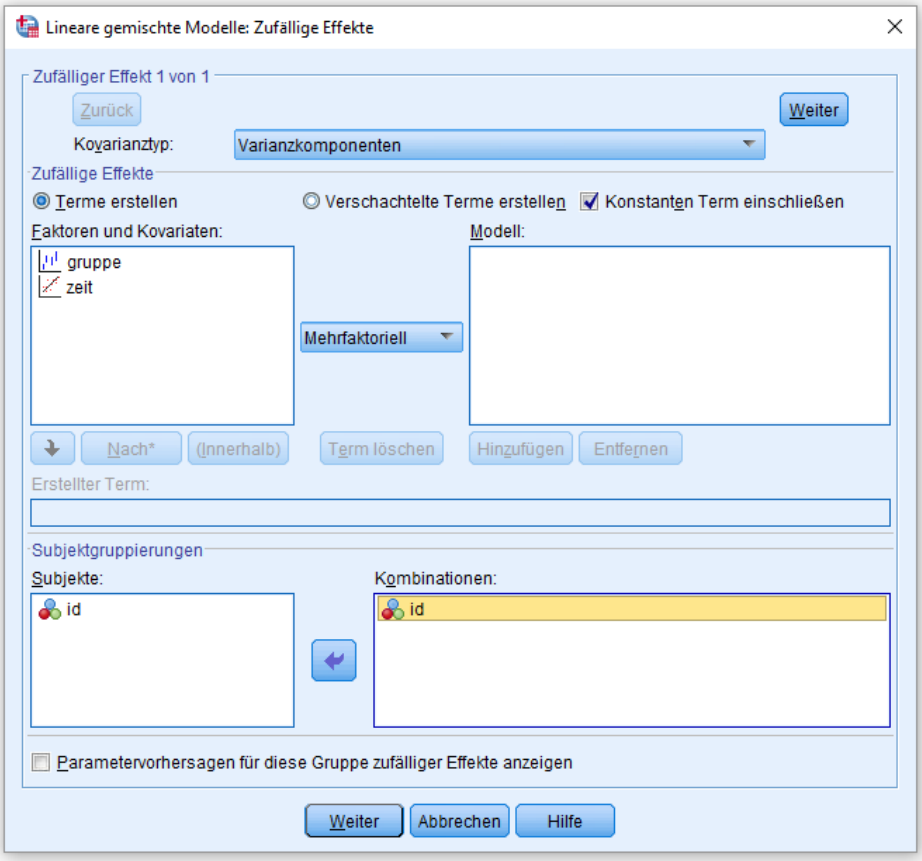
- Wir fügen in folgender Subdialogbox für die **festen Effekte** GRUPPE und ZEIT zum **Modell** hinzu:



Das Kontrollkästchen für den **konstanten Term** muss auf jeden Fall markiert sein (= Voreinstellung). Somit sind die Parameter zu den festen Effekten des Modells deklariert:

- Konstanter Term β_{00}
- GRUPPE β_{01}
- ZEIT β_{10}

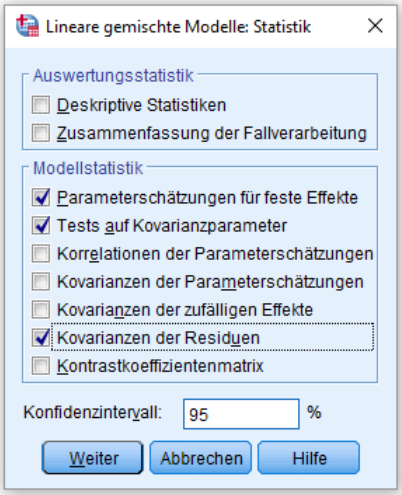
- In der Subdialogbox für die **zufälligen Effekte**



sind folgende Einstellungen vorzunehmen:

- Als **Kovarianztyp** kann die Voreinstellung **Varianzkomponenten** beibehalten werden. Hier geht es um die Struktur der Kovarianzmatrix der zufälligen Effekte. Unser Modell enthält unser Modell nur *einen* zufälligen Effekt (R_{0i}), so dass wir uns über die Kovarianzstruktur keine Gedanken machen müssen.
- Der **konstante Term** wird **eingeschlossen**.
  Ansonsten wird der zufällige Effekt auf die Ausgangsniveaus konstant gleich null gesetzt.
- Die Variable ID wird in die Liste der **Kombinationen** befördert, womit die oben im Dialog **Subjekte und Wiederholungen angeben** vorbereitete Definition der Makroeinheiten abgeschlossen ist.

- Schließlich fordern wir in der **Statistik**-Subdialogbox noch **Parameterschätzer**, **Tests für die Kovarianzparameter** und die **Kovarianzen der Residuen** an:



Über den **Einfügen**-Schalter der Hauptdialogbox erhält man folgendes Syntax-Äquivalent zu der Modellspezifikation:

```
MIXED y WITH gruppe zeit
  /CRITERIA=DFMETHOD(SATTERTHWAITE) CIN(95) MXITER(100) MXSTEP(10) SCORING(1)
    SINGULAR(0.000000000001) HCONVERGE(0, ABSOLUTE) LCONVERGE(0, ABSOLUTE)
    PCONVERGE(0.000001, ABSOLUTE)
  /FIXED=gruppe zeit | SSTYPE(3)
  /METHOD=REML
  /PRINT=R SOLUTION TESTCOV
  /RANDOM=INTERCEPT | SUBJECT(id) COVTYPE(VC)
  /REPEATED=zeit | SUBJECT(id) COVTYPE(AR1).
```

Zur manuellen Erstellung einer äquivalenten Syntax ist durch Ausnutzung der passenden Voreinstellungen nur wenig Schreibarbeit erforderlich:

```
mixed y with gruppe zeit
  /fixed = gruppe zeit
  /random = intercept | subject(id)
  /repeated = zeit | subject(id) covtype(ar1)
  /print = r solution testcov.
```

In der Modellspezifikation

Modelldimension[a]

		Anzahl Ausprägungen	Kovarianzstruktur	Anzahl Parameter	Subjektvariablen	Anzahl Fälle
Feste Effekte	Konstanter Term	1		1		
	gruppe	1		1		
	zeit	1		1		
Zufallseffekte	Konstanter Term	1	Varianz-Komponenten	1	id	
Wiederholte Effekte	zeit	4	Autoregressiv erster Ordnung	2	id	150
Gesamt		8		6		

a. Abhängige Variable: y.

protokolliert SPSS sechs Parameter:

- Drei Parameter für feste Effekte (β_{00}, β_{01}, β_{10})
- Die Varianzen π_{00} des Zufallseffekts R_{0i}
- Den autoregressiven Parameter ρ
- Die Varianz des weißen Rauschens δ_{ti}

In der Tabelle mit den festen Effekten

Schätzungen fester Parameter[a]

						Konfidenzintervall 95%	
Parameter	Schätzung	Standardfehler	Freiheitsgrade	T-Statistik	Sig.	Untergrenze	Obergrenze
Konstanter Term	-,032174	,210970	167,676	-,153	,879	-,448674	,384326
gruppe	,937868	,276391	147,962	3,393	,001	,391685	1,484050
zeit	,439451	,034945	160,616	12,575	,000	,370439	,508462

a. Abhängige Variable: y.

erhalten wir realistische Schätzungen für die mittleren Ausgangsniveaus in den beiden Gruppen (wahre Werte: 0 bzw. 1) sowie für den festen Zeiteffekt.

Die Tabelle mit den Kovarianzparametern

Schätzungen von Kovarianzparametern[a]

						Konfidenzintervall 95%	
Parameter		Schätzung	Standardfehler	Wald Z	Sig.	Untergrenze	Obergrenze
Meßwiederholungen	AR1, diagonal	,924958	,154381	5,991	,000	,666885	1,282900
	AR1, Rho	,469525	,092266	5,089	,000	,270574	,630014
Konstanter Term [Subjekt = id]	Varianz	2,395676	,355260	6,743	,000	1,791438	3,203720

a. Abhängige Variable: y.

finden sich brauchbare Schätzungen für die Varianz der Mikroebenenresiduen (wahrer Wert: 1), für die Autokorrelation erster Ordnung (wahrer Wert: 0,5) sowie für die Makroebenenresidualvarianz τ_{00} (wahrer Wert: 2,25).

Die Fehlerkovarianzen zu den Beobachtungszeitpunkten sinken mit wachsendem Abstand:

R-Matrix[a]

	[zeit = ,00]	[zeit = 1,00]	[zeit = 2,00]	[zeit = 3,00]
[zeit = ,00]	,924958	,434291	,203910	,095741
[zeit = 1,00]	,434291	,924958	,434291	,203910
[zeit = 2,00]	,203910	,434291	,924958	,434291
[zeit = 3,00]	,095741	,203910	,434291	,924958

Autoregressiv erster Ordnung

a. Abhängige Variable: y.

Im konstruierten Beispiel zeigen bei einer Vernachlässigung der autoregressiven Fehlerstruktur die Schätzer zu den festen Effekten kaum Verzerrungen:

Schätzungen fester Parameter[a]

Parameter	Schätzung	Standardfehler	Freiheitsgrade	T-Statistik	Sig.	Konfidenzintervall 95% Untergrenze	Obergrenze
Konstanter Term	-,030976	,210017	161,174	-,147	,883	-,445717	,383764
gruppe	,940942	,278014	148	3,385	,001	,391552	1,490331
zeit	,443700	,028709	449	15,455	,000	,387278	,500121

a. Abhängige Variable: y.

Allerdings erreicht das korrekte Modell eine deutlich bessere Fit-Beurteilung:

Korrektes Model
mit AR(1) - Fehlerkovarianzstruktur

Informationskriterien[a]

Eingeschränkte -2 Log Likelihood	1826,911
Akaike-Informationskriterium (AIC)	1832,911
Hurvich und Tsai (IC)	1832,952
Bozdogan-Kriterium (CAIC)	1849,087
Bayes-Kriterium von Schwarz (BIC)	1846,087

Die Informationskriterien werden in einem möglichst kleinen Format angezeigt.

a. Abhängige Variable: y.

Modell mit unterstellter
Unabhängigkeit der Fehler

Informationskriterien[a]

Eingeschränkte -2 Log Likelihood	1857,636
Akaike-Informationskriterium (AIC)	1861,636
Hurvich und Tsai (IC)	1861,656
Bozdogan-Kriterium (CAIC)	1872,420
Bayes-Kriterium von Schwarz (BIC)	1870,420

Die Informationskriterien werden in einem möglichst kleinen Format angezeigt.

a. Abhängige Variable: y.

Weil es dazu nur *einen* zusätzlichen Parameter benötigt (die Autokorrelation erster Ordnung), fallen auch die Kriterien mit Bestrafung für eine hohe Parameterzahl (z. B. AIC, BIC) deutlich günstiger aus.

6.6 Beispiel 2: Analyse von zeitvariablen Regressoren

Bei der Analyse von Cluster-Stichproben sorgt man durch das gruppeninterne Zentrieren von Mikroregressoren dafür, dass Mehrebenenmodelle bzgl. der kausalen Interpretierbarkeit der Effekte von Mikroregressoren denselben Staus besitzen wie das Fixed-Effects - Modell (vgl. Abschnitt 3.4.1). Eine analoge Zentrierung von Mikroregressoren kommt auch bei Panel-Studien in Frage.

Für den Mikroregressor Zeit ist eine Zentrierung jedoch zumindest bei einem für alle Fälle identischen Beobachtungsplan nicht erforderlich. Wir konnten daher im bisher betrachteten Panel-Beispiel (Veränderung der Wortbildungsleistung) auf eine Diskussion von Zentrierungsmethoden für Mikroregressoren verzichten. Wenn bei allen Personen (Makroeinheiten) dieselben Ausprägungen des Mikroregressors Zeit realisiert werden, kann es nicht zu einer für die kausale Interpretierbarkeit schädlichen Korrelation des Mikroregressors mit unbeobachteten Makromerkmalen kommen. Für jede Ausprägung der Zeitvariablen haben alle Makrovariablen bzw. Personmerkmale (ob beobachtet oder nicht) dieselbe Verteilung, sofern nicht fehlende Werte mit verletzter MAR-Bedingung zum Problem werden. Es kann zu einem kausalen Fehlschluss kommen, wenn z.B. im Verlauf einer Studie (mit steigendem Wert der Zeitvariablen) in Abhängigkeit von einem unbeobachteten Makromerkmal immer mehr Fälle aussteigen. Dann unterscheiden sich die bedingten Verteilungen des Makromerkmals gegeben eine Ausprägung der Zeitvariablen. Wenn das unbeobachtete Makromerkmal einen Einfluss auf das Kriterium ausübt, ist statistisch ein Scheineffekt der Zeit zu erwarten.

Bei anderen zeitvariablen Regressoren ist das interne (z.B. intraindividuelle) Zentrieren empfehlenswert, wie das folgende Beispiel aus Baltes-Götz (2016) illustriert.[1] Im Beispiel geht es um die Wirksamkeit von Maßnahmen zur Waldbrandprävention in einer fiktiven Welt. Für 80 Regionen werden über 10 Jahre beobachtet:

- Ausgaben für die Waldbrandprävention, erfasst in den Variablen X_{it}
- Die von Waldbränden betroffenen Flächen, erfasst in den Variablen Y_{it}

Für die Y_{it}-Werte in den Regionen und Jahren gilt in der Kunstwelt das folgende wahre Modell:

$$Y_{it} = \alpha_i - 0{,}5\, X_{it} + R_{it} \qquad (1)$$

Dabei steht α_i für die erwartete jährliche Schadfläche in der Region *i* bei Verzicht auf Präventionsmaßnahmen ($X_{it} = 0$). In der Kunstwelt gehen die α_i-Werte auf drei zeitstabile Regressoren Z_1, Z_2 und Z_3 sowie ein Residuum zurück:[2]

$$\alpha_i = Z_{1_i} + Z_{2_i} + Z_{3_i} + R_{\alpha_i} \qquad (2)$$

Der Anschaulichkeit halber sollen den zeitkonstanten Regressoren die folgenden Bedeutungen angeheftet werden:

- Z_1: Natürliche Ursachen für Waldbrände
- Z_2: Touristische Nutzung des Waldes
- Z_3: Brandstiftung (z.B. zur Gewinnung von Bauland)

Die Modellgleichung (1) wird komplettiert durch unbekannte zeitvariable Einflüsse im Residuum R_{it}.

Für den Präventionsaufwand in der Region *i* im Jahr *t* sorgt in der Kunstwelt die folgende Gleichung:

$$X_{it} = Z_{1_i} + Z_{2_i} + S_i + R_{X_{it}} \qquad (3)$$

Die Waldbrandrisiken Z_1 und Z_2 (nicht jedoch Z_3) beeinflussen den Präventionsaufwand. Außerdem sind weitere Eigenschaften der Regionen für den Präventionsaufwand verantwortlich (z.B. Bedeutung des Waldes für die Wirtschaft), die in Gleichung (3) durch das Symbol S_i vertreten werden. Schließlich werden zeitvariable Einflüsse (z.B. Haushaltsbelastungen durch exogene Krisen) angenommen, die in Gleichung (3) durch das Symbol $R_{X_{it}}$ vertreten sind.

Als politische Entscheidungsgrundlage ist ein möglichst genaues Wissen um die Wirksamkeit der Präventionsmaßnahmen hochrelevant. Zur Schätzung kommen verschiedene Modelle in Frage, deren Vor- und Nachteile anschließend demonstriert werden. Laut Standardannahme über die simulierte Welt sind die zeitstabilen Regressoren Z_1, Z_2 und Z_3 *nicht* beobachtet worden. Gelegentlich werden Varianten der Kunstwelt betrachtet.

Zuerst analysieren wir die Daten in der Standardkonfiguration mit dem folgenden POLS-Modell (*Pooled Ordinary Least Squares*):

$$Y_{it} = \alpha + \beta X_{it} + \tilde{R}_{it}$$

Im Modell wird vernachlässigt, dass die Regressoren X_{it} mit den Residuen $\tilde{R}_{it}$ korrelieren, weil letztere auch die individuellen Effekte α_i enthalten:

$$\tilde{R}_{it} = \alpha_i + R_{it}$$

Infolgedessen ist das geschätzte Regressionsgewicht (-0,157) im Vergleich zum wahren Effekt (von -0,5) erheblich verzerrt:

---

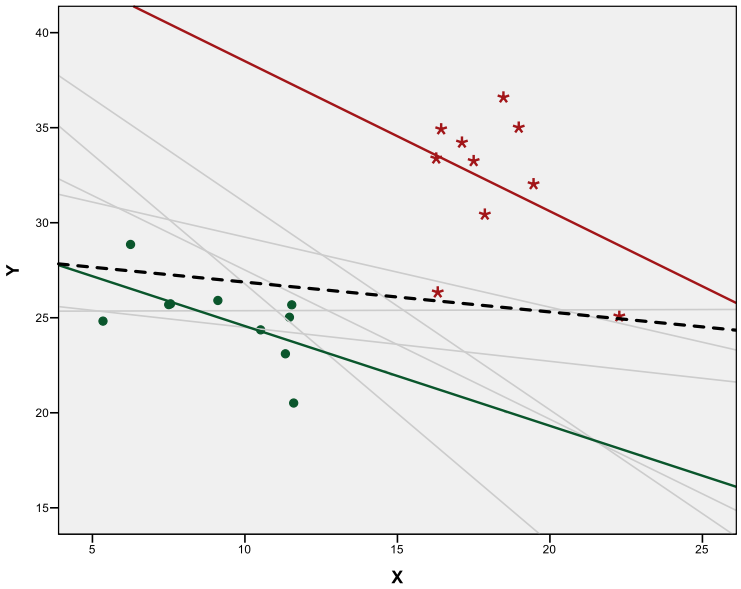
[1] SPSS-Dateien mit Syntax, Daten und Ausgaben zur Simulation sind im Ordner **Waldbrandprävention** an der im Vorwort vereinbarten Stelle zu finden.

[2] Im Text werden Variablen oft ohne Indizes für Regionen und Jahre notiert (z.B. Z_1 statt Z_{1_i}), wenn es der Klarheit nicht schadet. Weil das Anwendungsbeispiel nur *einen* zeitvariablen Regressor enthält, bleibt sein Symbol ohne Nummer (z.B. X_{it} statt $X_{1_{it}}$).

$$\hat{Y}_{it} = 28{,}445 - 0{,}157\, X_{it}$$
$$(0{,}572) \qquad (0{,}034)$$

Außerdem werden die Standardfehler (eingeklammerte Werte in der unteren Ergebniszeile) unterschätzt, weil das Modell von 80×10 *unabhängigen* Fällen ausgeht, also die Abhängigkeit der 10 aus einer Region stammenden Beobachtungen ignoriert. Im Datensatz ist weniger Information über die Parameter vorhanden als vom statistischen Kalkül angenommen, und infolgedessen wird für die Parameterschätzungen eine zu geringe Streuung berechnet.

In der folgenden Abbildung sind für 10 Regionen die Regressionen der Schadenshöhe auf den Präventionsaufwand zu sehen, wobei mit einer Ausnahme eine kostendämpfende Wirkung der Prävention abzulesen ist. Die mit den Daten *aller* Regionen geschätzte POLS-Regressionsgerade ist als unterbrochene Linie eingezeichnet.



Für zwei Regionen mit einem relativ niedrigen bzw. hohen Schadensniveau sind die Regressionsgeraden hervorgehoben und die Einzelbeobachtungen als Kreise bzw. Sterne eingetragen. So wird optisch plausibel, warum die POLS-Regressionsgerade keine nennenswerte Präventionswirkung vermuten lässt, obwohl in den meisten Regionen ein deutlicher Effekt besteht.

Im Beispiel ist durch die Anwendung eines POLS-Modells auf Paneldaten der Effekt eines zeitvariablen Regressors betragsmäßig *unterschätzt* worden. Unter angepassten Bedingungen kann es auch zu einer Überschätzung kommen. Um das Risiko verzerrter Schätzungen zu minimieren, bieten sich folgende Maßnahmen an:

- Alle relevanten Regressoren beobachten und in das Modell einbeziehen
  Dieses Ziel sollte immer angestrebt werden, ist aber in der Regel nicht vollständig realisierbar.
- Verwendung einer Panel-adäquaten Modellierung
  Bei einem Mikroregressor (mit Varianz auf der Mikro- und der Makroebene) kann man durch eine Panel-adäquate Modellierung zwischen dem (meist primär interessierenden) Within-Effekt und dem Between-Effekt unterscheiden. So gelangt man zu einer Schätzung des Within-Effekts, die nicht durch Korrelationen zwischen dem Mikroregressor und ausgelassenen *zeitstabilen* Regressoren verzerrt sein können.

6.7 Fehlerhaftes Mehrebenenmodell mit einem unzentrierten Mikroregressor

Während das in der ökonometrischen Methodendiskussion zur Analyse von Panel-Datei favorisierte Fixed-Effects - Modell (siehe z.B. Brüderl 2010) nur Mikroregressoren berücksichtigt, kann ein Mehrebenenmodell Mikro- und Makroregressoren enthalten. Seine generelle Form kann für eine Panel-Stichprobe genauso formuliert werden wie für eine Cluster-Stichprobe:

$$Y_{it} = \beta_{0\mathrm{i}} + \beta_1 X_{1_{it}} + \ldots + \beta_J X_{J_{it}} + R_{it}$$
$$\beta_{0\mathrm{i}} = \gamma_{00} + \gamma_{01} Z_{1_i} + \ldots + \gamma_{0K} Z_{K_i} + U_{0i}$$

Die zeitvariablen Regressoren ($X_1, \ldots, X_J$) gehören zur Ebene der Zeitpunkte, und die zeitstabilen Regressoren ($Z_1, \ldots, Z_K$) zur Ebene der Individuen.

Weil im aktuellen Beispiel die Zeit und mögliche Polynome davon nicht als Mikroregressoren auftreten, ersparen wir uns die im bisherigen Verlauf von Abschnitt 6 verwendete, etwas unhandliche Bezeichnungskonventionen nach Raudenbush & Bryk (2002, S. 162). Wir verwenden stattdessen eine Notation analog zu Abschnitt 3, wobei aber *t* als Index für die Mikroeinheiten (Zeitpunkte) und *i* als Index für die Makroeinheiten (Gebiete) verwendet wird.

In der Makrogleichung werden die individuellen Effekte partiell durch die zeitstabilen Regressoren erklärt. Die unerklärten Anteile U_{0i} (die Makroresiduen) bilden zusammen mit den Mikroresiduen R_{it} das Zweikomponenten-Residuum der zusammengesetzten Modellgleichung:

$$Y_{it} = \gamma_{00} + \beta_1 X_{1_{it}} + \ldots + \beta_J X_{J_{it}} + \gamma_{01} Z_{1_i} + \ldots + \gamma_{0K} Z_{K_i} + U_{0i} + R_{it} \qquad (4)$$

Für die Makroresiduen U_{0i} wird angenommen, dass sie mit allen Regressoren unkorreliert sind:

$$\mathrm{Cor}\left(X_{j_{it}}, U_{0i}\right) = \mathrm{Cor}\left(Z_{k_i}, U_{0i}\right) = 0$$

Zwar wird diese Annahme selten vollständig gerechtfertigt sein, doch kann ihre Plausibilität durch die Aufnahme von potentiell relevanten zeitstabilen Regressoren in das Modell erhöht werden.[1]

Damit die Koeffizienten zu den Mikroregressoren konsistent geschätzt werden können, wird zudem (in allen Mehrebenenmodellen für Paneldaten) die *strikte Exogenität* benötigt, also die Unkorreliertheit der zum Zeitpunkt *t* beobachteten Mikroregressoren mit den zu einem beliebigen Zeitpunkt *s* beobachteten Mikroresiduen:

$$\mathrm{Cor}\left(X_{j_{it}}, R_{is}\right) = 0$$

Ist ein zeitvariabler Regressor X_j mit den Makroresiduen korreliert, dann ist sein Schätzer in Relation zum primär interessierenden Within-Effekt verzerrt. Im Waldbrandbeispiel liegt genau diese Situation vor, und wir erhalten für den Regressor Präventionsaufwand (mit dem wahren Within-Effekt -0,5) eine verzerrte Effektschätzung:

$$\begin{array}{rcrcl} \hat{Y}_{it} & = & 30{,}940 & - & 0{,}314\ X_{it} \\ & & (0{,}943) & & (0{,}055) \end{array}$$

Für Korrelationen zwischen den Makroresiduen und den zeitvariablen Regressoren im Mehrebenenmodell sind vor allem ausgelassene zeitkonstante Regressoren verantwortlich. Folglich kann das Problem durch die Erfassung und Berücksichtigung von relevanten zeitkonstanten Regressoren reduziert werden. In der simulierten Kunstwelt entstehen die individuellen Effekte α_i aus drei zeitstabilen Regressoren und einem Zufallsanteil (siehe Gleichung (2)), wobei die zeitstabilen Regressoren Z_1, Z_2 und Z_3 bislang als unbeobachtet galten. Nun verschieben wir in der künstlichen Welt vorübergehend die Grenze des Wissens und nehmen den zeitstabilen Regressor Z_1 in das Modell auf. Im erweiterten Modell rückt der Schätzer zum Effekt des zeitvariablen Regressors näher an den wahren Within-Effekt von -0,5 heran:

$$\begin{array}{rcrcrcl} \hat{Y}_{it} & = & 29{,}552 & - & 0{,}382\ X_{it} & + & 0{,}663\ Z_{1_i} \\ & & (0{,}990) & & (0{,}058) & & (0{,}193) \end{array}$$

[1] Wird die *Zeit* als variabler Regressor einbezogen (über eine metrische Variable oder eine Serie von Dummy-Variablen), dann drohen *keine* Korrelationen mit den Makroresiduen, weil die Zeitvariablen keine interindividuelle Varianz besitzen.

6.8 Within-Between - Modell

Ist im Mehrebenenmodell für einen (unzentrierten) Mikroregressor die kritische Annahme der Unkorreliertheit mit den Makroresiduen falsch, dann resultiert eine verzerrte Effektschätzung. Diese Schwäche ist aber leicht zu beheben, indem für jeden potentiell betroffenen zeitvariablen Regressor X_j analog zu Abschnitt 3.4.1.1.2 unter Verwendung der Mittelwertsvariablen $\bar{X}_{j\,i.}$

$$\bar{X}_{j\,i.} = \frac{1}{T}\sum_{t=1}^{T} X_{j\,it}$$

die intraindividuell zentrierte Variante $(X_{j\,it} - \bar{X}_{j\,i.})$ zusammen mit $\bar{X}_{j\,i.}$ in das Modell einbezogen wird. Wir gehen im Waldbrandbeispiel davon aus, dass nur ein Mikroregressor vorhanden ist und dass die zeitstabilen Regressoren Z_1, Z_2 und Z_3 *nicht* beobachtet worden sind. Damit ergibt sich das folgende Modell mit dem Symbol $\beta_1^{(W)}$ für den Within-Effekt und $\beta_1^{(B)}$ für den Between-Effekt des Regressors:

$$Y_{it} = \gamma_{00} + \beta_1^{(W)}(X_{1\,it} - \bar{X}_{1\,i.}) + \beta_1^{(B)}\bar{X}_{1\,i.} + U_{0i} + R_{it} \qquad (5)$$

Dieses Modell wird von vielen Autoren (z.B. Allison 2009, S. 23ff) als *Hybrid-Modell* bezeichnet. In diesem Manuskript wird der Name **Within-Between (WB) - Modell** verwendet.

Mit den Variablen `DX` (für die intraindividuell zentrierte Waldbrandprävention) und `MX` (für die individuell gemittelte Waldbrandprävention) liefert das folgende MIXED-Kommando

```
mixed y with dx mx
  /criteria dfmethod(kenwardroger)
  /fixed dx mx
  /random intercept | subject(id)
  /print solution testvov.
```

für die simulierten Waldbranddaten folgende Schätzwerte zum WB-Modell:

Schätzungen fester Parameter[a]

Parameter	Schätzung	Standardfehler	Freiheitsgrade	T-Statistik	Sig.	Konfidenzintervall 95% Untergrenze	Konfidenzintervall 95% Obergrenze
Konstanter Term	27,616554	1,264706	78,000	21,836	,000	25,098719	30,134389
DX	-,498275	,074518	719	-6,687	,000	-,644573	-,351977
MX	-,104237	,076582	78,000	-1,361	,177	-,256700	,048227

a. Abhängige Variable: Y.

Für $\beta^{(W)}$ (wahrer Wert: -0,5) resultiert eine präzise Schätzung.

Intraindividuell zentrierte Zeitpunktregressoren sind mit allen zeitkonstanten Regressoren unkorreliert, also auch mit dem zugehörigen Mittelwertsregressor (Kreft & de Leeuw 1998, 135f). Damit bleibt der Within-Schätzer unverändert, wenn man aus dem WB-Modell den Mittelwertsregressor entfernt oder eine andere Veränderung bei der Ausstattung mit zeitkonstanten Regressoren vornimmt.

Während einige Autoren den Between-Effekt als wenig erhellend oder als irrelevant beurteilen (Allison 2009, S. 25; Brüderl 2010, S. 977), sehen andere dahinter einen Prozess von eigenständigen Interesse (z.B. Snijders & Bosker 2012). Bei Forschungsarbeiten in der Mehrebentradition ist es üblich, zeitvariable Regressoren als intraindividuell zentrierte Variable *und* als Mittelwertsvariable aufzunehmen (siehe z.B. Hedeker & Gibbons 2006, S. 72ff). Man möchte Niveaueffekte separieren von Veränderungseffekten (z.B. bei der Höhe von Steuern oder bei der Dosierung von Medikamenten).

6.9 Hausman-Test

Ersetzt man im WB-Modell den intraindividuell zentrierte Mikroregressor $(X_{1_{it}} - \bar{X}_{1_{i.}})$ durch den unzentrierten Mikroregressor $X_{1_{it}}$, dann resultiert das sogenannte **Correlated Random Effects (CRE) - Modell** (siehe Wooldridge 2013, S. 479ff):[1]

$$Y_{it} = \gamma_{00} + \beta_1 X_{1_{it}} + \delta_{01}\bar{X}_{1_{i.}} + \gamma_{01} Z_{1_i} + U_{0i} + R_{it} \qquad (6)$$

Der Regressionskoeffizient zu $\bar{X}_{1_{i.}}$ schätzt die Differenz $\left(\beta_1^{(B)} - \beta_1^{(W)}\right)$ zwischen dem Within- und dem Between-Effekt des Mikroregressors X_1, und sein Signifikanztest erlaubt eine Beurteilung der Nullhypothese

$$\mathrm{H}_0: \beta_1^{(B)} = \beta_1^{(W)}$$

Im Waldbrandbeispiel zaubern wir nun die bisher als unbekannt betrachteten zeitstabilen Regressoren $(Z_1,\ Z_2,\ Z_3)$ aus dem Hut und nehmen sie in das CRE-Modell auf. Das MIXED-Kommando

```
mixed y with x mx z1 z2 z3
  /print = solution testcov
  /fixed = x mx z1 z2 z3
  /random = intercept | subject(id).
```

liefert die folgenden Parameterschätzungen:

Schätzungen fester Parameter[a]

Parameter	Schätzung	Standardfehler	Freiheitsgrade	T-Statistik	Sig.	Konfidenzintervall 95% Untergrenze	Obergrenze
Konstanter Term	13,479036	1,182136	75	11,402	,000	11,124101	15,833972
X	-,498275	,074518	719	-6,687	,000	-,644573	-,351977
MX	,068010	,093280	432,524	,729	,466	-,115328	,251349
Z1	,884470	,121225	75	7,296	,000	,642979	1,125962
Z2	,780846	,107431	75,000	7,268	,000	,566832	,994861
Z3	1,125459	,100787	75	11,167	,000	,924681	1,326236

a. Abhängige Variable: Y.

Weil der Test zum Mittelwertsregressor $\bar{X}_{1_{i.}}$ die Signifikanz deutlich verfehlt, kann die Nullhypothese der Identität von Within- und Between-Effekt akzeptiert werden. Folglich kann auf den Mittelwertsregressor sowie auf die Unterscheidung zwischen dem Within- und dem Between-Effekt von X_1 verzichtet werden. Das MIXED-Kommando

```
mixed y with x z1 z2 z3
  /print = solution testcov
  /fixed = x z1 z2 z3
  /random = intercept | subject(id).
```

liefert präzise Schätzungen der wahren Parameter in der simulierten Wert:

[1] Im Namen des Modells kommt zum Ausdruck, dass Korrelationen zwischen den individuellen Effekten und den zeitvariablen Regressoren erlaubt sind. In diesem Abschnitt enthalten die Modellgleichungen der Übersichtlichkeit halber generell nur *einen* zeitvariablen und *einen* zeitkonstanten Regressor.

Schätzungen fester Parameter[a]

Parameter	Schätzung	Standardfehler	Freiheitsgrade	T-Statistik	Sig.	Konfidenzintervall 95% Untergrenze	Konfidenzintervall 95% Obergrenze
Konstanter Term	13,479456	1,180806	75,347	11,415	,000	11,127348	15,831565
X	-,454844	,044788	178,770	-10,155	,000	-,543225	-,366462
Z1	,917952	,112053	94,372	8,192	,000	,695480	1,140424
Z2	,805507	,101847	87,130	7,909	,000	,603079	1,007934
Z3	1,123844	,100649	75,393	11,166	,000	,923358	1,324330

a. Abhängige Variable: Y.

Jetzt sind wir wieder bei einem Mehrebenenmodell mit einem unzentrierten Mikroregressor angekommen (vgl. Abschnitt 6.7), das aber als gültig angenommen werden kann.

Traditionell wird der sogenannte *Hausman-Spezifikationstest* dazu verwendet, um in Panel-Studien für zeitvariable Regressoren den Within- und den Between-Effekt auf Identität zu beurteilen (Hausman 1978). Mit den Tests zu den Mittelwertsregressoren bietet das CRE-Modell eine bequeme und vielfach empfohlene Alternative zum Hausman-Test (Allison 2009, S. 25; Wooldridge 2013, S. 480).

6.10 Vergleich zu traditionellen Verfahren für Längsschnittdaten

6.10.1 Vorteile der HLM-Technik

Im Vergleich zur traditionellen Messwiederholungsvarianzanalyse sind vor allem die folgenden Vorzüge der HLM-Technik bei der Analyse von Längsschnittdaten zu nennen:

- Es sind individuelle Beobachtungspunkte (hinsichtlich Anzahl und zeitlicher Positionierung) erlaubt.
- Gilt für alle Untersuchungsteilnehmer derselbe Beobachtungsplan, sind beide Verfahren anwendbar. Während die Messwiederholungsvarianzanalyse jedoch *komplette* Fälle (mit Daten zu *allen* Messzeitpunkten) benötigt, können bei der Mehrebenenanalyse auch unvollständige Fälle berücksichtigt werden. Während der bei einer traditionellen Messwiederholungsanalyse angewandte fallweise Ausschluss fehlender Werte nur unter der MCAR-Bedingung zulässig ist, gelangt die Mehrebenenanalyse durch die Verwendung von Maximum-Likelihood - Methoden auch unter der MAR-Bedingung zu unverzerrten Schätzungen (Hox 2010, S. 107; Raudenbush & Bryk 2002, S. 199f). Bei verletzter MAR-Bedingung kann eine HLM-Modellierung jedoch zu verzerrten Schätzungen führen. Wenn z.B. im Verlauf einer Studie mit der Zeit als einzigem Mikroregressor in Abhängigkeit von einem unbeobachteten Makromerkmal sukzessive immer mehr Fälle aussteigen, dann unterscheiden sich die bedingten Verteilungen des Makromerkmals gegeben eine Ausprägung der Zeitvariablen. Wenn das unbeobachtete Makromerkmal einen Einfluss auf das Kriterium ausübt, ist statistisch ein Scheineffekt des Mikroregressors Zeit zu erwarten. In dieser Situation ist die Messwiederholungsvarianzanalyse mit der hier prinzipiell realisierten fallweisen Behandlung fehlender Werte im Vorteil. Zwar leidet die Generalisierbarkeit der Ergebnisse (die externe Validität) unter der veränderten Stichprobenzusammensetzung, doch bleibt die interne Validität der Zeiteffektbeurteilung erhalten.
- Die Berücksichtigung zeitabhängiger Kovariaten ist bei der traditionellen Messwiederholungsanalyse nicht möglich, bei der Mehrebenenanalyse aber eine Selbstverständlichkeit (über Merkmale auf der Mikroebene).
- Bei einer Messwiederholungsanalyse wird zur Modellierung der Abhängigkeit zwischen den Messungen desselben Falles entweder die Sphärizität angenommen (univariater Ansatz) oder eine frei schätzbare Kovarianzmatrix zugelassen (multivariater Ansatz). Bei der Mehrebenenanalyse sind zwischen diesen Extremen flexible Kovarianzstrukturen modellier- und testbar (Hox 2010, S. 103).

6.10.2 Messwiederholungsvarianzanalyse als spezielle Mehrebenenanalyse

In diesem Abschnitt wird demonstriert, dass die traditionelle Messwiederholungsvarianzanalyse als spezielle Mehrebenenanalyse aufgefasst werden kann. Zur Demonstration werden künstliche Daten[1] verwendet mit ...

- einem dreistufigen Messwiederholungsfaktor mit festem Effekt
- einem zufälligen Ordinatenabschnitt

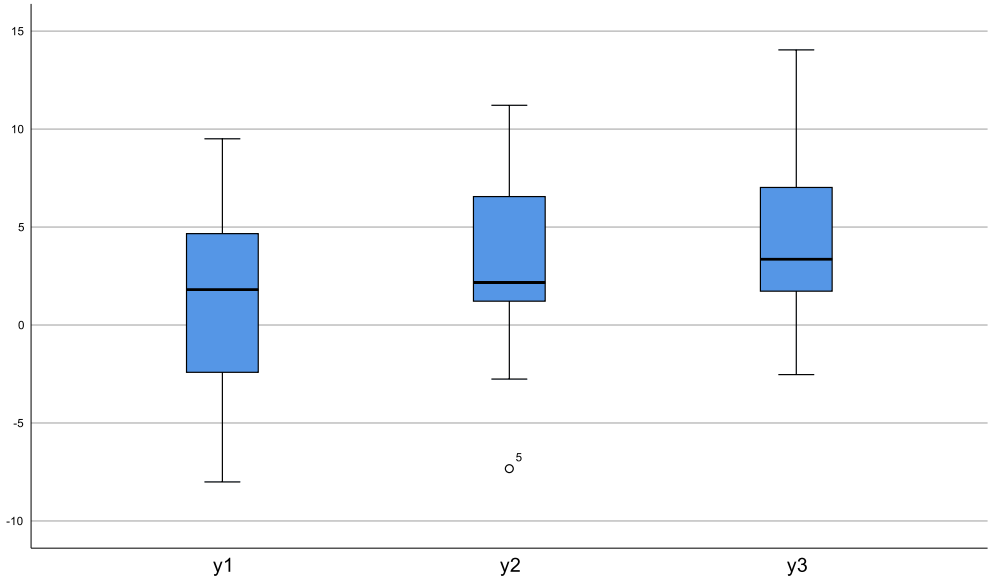
Mit den Indikator-Kodiervariablen D_1 und D_2 für den Messwiederholungsfaktor lautet das korrekte kombinierte Modell:

$$Y_{ti} = \beta_{00} + \beta_{10} D_1 + \beta_{20} D_2 + R_{0i} + E_{ti}$$

Die wahren Parameterausprägungen in der künstlichen Population sind:

- β_{00}: 0
- β_{10}: 1
- β_{20}: 2
- Varianz τ_{00}: 9
- Fehlervarianz σ^2: 9

In einer Stichprobe mit 30 Fällen zeigt sich ein starker Effekt des Messwiederholungsfaktors:



Bei vollständiger Datenmatrix (keine fehlenden Werte) liefern die SPSS-Prozedur GLM bei einer Messwiederholungsvarianzanalyse mit der univariaten Methode und die Prozedur MIXED bei einer passend spezifizierten Mehrebenenanalyse (fester Zeiteffekt, zufälliger Ordinatenabschnitt) und Verwendung der REML-Schätzmethode identische Testergebnisse.

Unter Verwendung der Daten im Breit-Format eignet sich das folgende GLM-Kommando zur Analyse des Messwiederholungsfaktors:

```
glm y1 y2 y3
  /wsfactor=Zeit 3
  /wsdesign=Zeit.
```

Die Nullhypothese der Sphärizität wird für die Beispieldaten akzeptiert:

[1] SPSS-Dateien mit Syntax, Daten und Ausgaben zur Simulation sind im Ordner **MWH-VA und HLM** an der im Vorwort vereinbarten Stelle zu finden.

Mauchly-Test auf Sphärizität[a]

Maß: MEASURE_1

Innersubjekteffekt	Mauchly-W	Approx. Chi-Quadrat	df	Sig.	Epsilon[b] Greenhouse-Geisser	Huynh-Feldt	Untergrenze
Zeit	,964	1,037	2	,595	,965	1,000	,500

Prüft die Nullhypothese, daß sich die Fehlerkovarianz-Matrix der orthonormalisierten transformierten abhängigen Variablen proportional zur Einheitsmatrix verhält.

a. Design: Konstanter Term
Innersubjektdesign: Zeit

b. Kann zum Korrigieren der Freiheitsgrade für die gemittelten Signifikanztests verwendet werden. In der Tabelle mit den Tests der Effekte innerhalb der Subjekte werden korrigierte Tests angezeigt.

Für den Messwiederholungsfaktor (Zeit) wird die Nullhypothese deutlich zurückgewiesen ($F_{2,98} = 7{,}627$; $p = 0{,}001$):

Tests der Innersubjekteffekte

Maß: MEASURE_1

Quelle		Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Sig.
Zeit	Sphärizität angenommen	160,594	2	80,297	7,627	,001
	Greenhouse-Geisser	160,594	1,930	83,216	7,627	,001
	Huynh-Feldt	160,594	2,000	80,297	7,627	,001
	Untergrenze	160,594	1,000	160,594	7,627	,010
Fehler(Zeit)	Sphärizität angenommen	610,588	58	10,527		
	Greenhouse-Geisser	610,588	55,965	10,910		
	Huynh-Feldt	610,588	58,000	10,527		
	Untergrenze	610,588	29,000	21,055		

Unter Verwendung der Daten im Lang-Format liefert das folgende MIXED-Kommando

```
mixed y by zeit
  /fixed=zeit | sstype(3)
  /print=solution testcov
  /random=intercept | subject(nr) covtype(vc).
```

dasselbe Testergebnis:

Tests auf feste Effekte, Typ III[a]

Quelle	Zähler-Freiheitsgrade	Nenner-Freiheitsgrade	F-Wert	Sig.
Konstanter Term	1	29	18,535	,000
zeit	2	58	7,627	,001

a. Abhängige Variable: y.

Außerdem erhalten wir u.a. die folgenden Informationskriterien:

Informationskriterien[a]

Eingeschränkte -2 Log Likelihood	499,350
Akaike-Informationskriterium (AIC)	503,350
Hurvich und Tsai (IC)	503,493
Bozdogan-Kriterium (CAIC)	510,282
Bayes-Kriterium von Schwarz (BIC)	508,282

Die Informationskriterien werden in einem möglichst kleinen Format angezeigt.

a. Abhängige Variable: y.

Aus der HLM-Modellspezifikation ergibt sich für die Kovarianzmatrix der Beobachtungen eine zusammengesetzt-symmetrische Struktur. Wie in Abschnitt 6.5 erläutert, kann diese Kovarianzstruktur über

einen zufälligen Ordinatenabschnitt oder durch direkte Spezifikation der Fehlerkovarianzen (in der **R**-Matrix) erreicht werden. Die letztgenannte Auswertungstechnik kann über die folgende Syntax realisiert werden:

```
mixed y by zeit
  /fixed=zeit | sstype(3)
  /print=r solution testcov
  /repeated=zeit | subject(nr) covtype(cs).
```

Dabei lohnt es sich, die **R**-Matrix ausgeben zu lassen:

R-Matrix[a]

	[zeit = 1]	[zeit = 2]	[zeit = 3]
[zeit = 1]	19,785940	9,258562	9,258562
[zeit = 2]	9,258562	19,785940	9,258562
[zeit = 3]	9,258562	9,258562	19,785940

Zusammengesetzte Symmetrie

a. Abhängige Variable: y.

Durch direkte Spezifikation lässt sich bei der SPSS-Prozedur MIXED für die Kovarianzmatrix der Residuen alternativ zur zusammengesetzt-symmetrischen Struktur auch eine komplett freie Struktur anfordern (bei k Messungen mit $k(k+1)/2$ zu schätzenden Parametern). In Bezug auf dieses Modell mit saturierter Fehlerkovarianzmatrix stellt jedes Modell mit weniger Parametern für die Kovarianzen der Residuen eine Einschränkung dar, so dass ein Modellvergleich per Likelihood-Quotienten - Test möglich ist (Hox 2010, S. 103).

Im Beispiel führt die Freigabe der Kovarianzstruktur zur folgenden Schätzung für die **R**-Matrix:

R-Matrix[a]

	[zeit = 1]	[zeit = 2]	[zeit = 3]
[zeit = 1]	20,973839	9,786841	7,960465
[zeit = 2]	9,786841	18,467218	10,028380
[zeit = 3]	7,960465	10,028380	19,916762

Unstrukturiert

a. Abhängige Variable: y.

Es werden 4 zusätzliche Parameter benötigt:

- Bei einer R-Matrix mit CS-Struktur sind eine Varianz und eine Kovarianz zu schätzen.
- Bei einer R-Matrix mit UN-Struktur und 3 Messzeitpunkten sind $\frac{3 \cdot 4}{2} = 6$ Parameter zu schätzen.

Das -2LL - Maß für den Anpassungsfehler sinkt aber kaum:

Informationskriterien[a]

Eingeschränkte -2 Log Likelihood	498,269
Akaike-Informationskriterium (AIC)	510,269
Hurvich und Tsai (IC)	511,319
Bozdogan-Kriterium (CAIC)	531,065
Bayes-Kriterium von Schwarz (BIC)	525,065

Die Informationskriterien werden in einem möglichst kleinen Format angezeigt.

a. Abhängige Variable: y.

Bei 4 Freiheitsgraden ist die Prüfgröße für den Modellvergleich (499,35 - 498,269 = 1,081) weit von der Signifikanzgrenze entfernt (p = 0,897).[1] Dieses Ergebnis war aufgrund des oben berichteten Sphärizitätstests zu erwarten.

[1] Zur Berechnung der Überschreitungswahrscheinlichkeit zur -2LL - Differenz taugt z. B. folgendes Kommando:
compute p = 1 - CDF.CHISQ(1.081, 4).

Es wundert auch nicht, dass die Informationskriterien AIC und BIC klar für das einfachere Modell votieren.

Wenngleich die Messwiederholungsvarianzanalyse im konkreten Beispiel äquivalente Ergebnisse geliefert hat, sprechen doch weiterhin gravierende Vorteile für die Mehrebenenanalyse:

- Bessere Behandlung fehlender Werte
  Die Messwiederholungsvarianzanalyse benötigt vollständige Fälle und zwingt zum fallweisen Ausschluss fehlender Werte. Dies führt zu verzerrten Schätzungen, wenn die strenge MCAR-Bedingung *nicht* erfüllt ist (vgl. Hox 2010, S. 106ff). Die Mehrebenenanalyse kann demgegenüber fehlende Werte tolerieren und liefert auch unter den leichter zu erfüllenden MAR-Anforderungen unverzerrte Parameterschätzungen.
- Die Toleranz gegenüber fehlenden Werten kann auch als Unterstützung individueller Erhebungspläne beschrieben werden.
- Unterstützung von zeitabhängigen Kovariaten (siehe z.B. Baltes-Götz 2016)

Literatur

Allison, P. D. (2009). *Fixed Effects Regression Models*. Los Angeles, CA: Sage.

AL-Marshadi, A.H. (2007). The New Approach to Guide the Selection of the Covariance Structure in Mixed Model. *Research Journal of Medicine and Medical Sciences*, 2(2), 88-97.

Baltes-Götz, B. (2013). *Behandlung fehlender Werte in SPSS und Amos*. Online-Dokument: http://www.uni-trier.de/index.php?id=23239

Baltes-Götz, B. (2016). Regressionsmodelle für Paneldaten. In A. Geissler & M. Schneider (Hrsg.). *Zwischen artes liberales und artes digitales*, S. 157-181. Marburg: Tectum-Verlag.

Baltes-Götz, B. (2018). *Mediator- und Moderatoranalyse mit SPSS und PROCESS*. Online-Dokument: http://www.uni-trier.de/index.php?id=22528

Baltes-Götz, B. (2019). *Lineare Regressionsanalyse mit SPSS*. Online-Dokument: http://www.uni-trier.de/index.php?id=22489

Brüderl, J. (2010). Kausalanalyse mit Paneldaten. In C. Wolf & H. Best (Hrsg.). *Handbuch der sozialwissenschaftlichen Datenanalyse*, S. 963–994, Wiesbaden: VS Verlag.

Cohen, J., Cohen, P., West, S. G. & Aiken, L. S. (2003). *Applied multiple regression/correlation analyses for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

De Leeuw, J. & Meijer, E. (2008). Introduction to Multilevel Analysis. In: De Leeuw, J. & Meijer, E. (Eds.). *Handbook of Multilevel Analysis*. New York: Springer.

Eid, M., Gollwitzer, M. & Schmitt, M. (2011). *Statistik und Forschungsmethoden* (2. Aufl.). Weinheim: Beltz.

Field, A. (2013). *Discovering Statistics Using SPSS* (4th ed.). London: Sage.

Goldstein, H. (2003). *Multilevel Statistical Models* (3rd ed.). London: Arnold.

Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica*, 46(6), pp. 1251-1271.

Heck, R.H., Thomas, S.L. & Tabata, L.N. (2010). *Multilevel and Longitudinal Modeling with IBM SPSS*. New York: Taylor & Francis.

Heck, R.H., Thomas, S.L. & Tabata, L.N. (2012). *Multilevel Modeling of Categorical Outcomes Using IBM SPSS*. New York: Taylor & Francis.

Hedeker, D. & Gibbons, R. D. (2006). *Longitudinal Data Analysis*. New York: Wiley & Sons.

Hox, J.J. (2010). *Multilevel Analysis. Techniques and Applications* (2nd ed.). New York: Routledge.

Hox, J.J., Moerbeek, M., & van de Schoot, R. (2018). *Multilevel Analysis. Techniques and applications* (3rd ed.). New York: Routledge.

Judd, C.M., Westfall, J., & Kenny, D.A. (2017). Experiments with More Than One Random Factor Designs, Analytic Models, and Statistical Power. *Annu. Rev. Psychol.*, 6817.1–17.25.

Kreft, I. & de Leeuw, J (1998). *Introducing Multilevel Modeling*. London: Sage.

Leyland, A.H. (2004). A review of multilevel modelling in SPSS. Online-Dokument: http://www.mlwin.com/softrev/reviewspss.pdf

Lüdtke, O., Marsh, H.W., Robitzsch, A., Trautwein, U., Asparouhov, T. & Muthén, B. (2008). The Multilevel Latent Covariate Model: A New, More Reliable Approach to Group-Level Effects in Contextual Studies. *Psychological Methods*, 13(3), 203-229.

Maas, C.J.M & Hox, J.J (2005). Sufficient Sample Sizes for Multilevel Modeling. *Methodology*, 1(3), pp. 86–92

Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J. & Craven, R. G. (2008). The big-fish-little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychology Review, 20*, 319-350.

Norušis, M.J. (2008). *SPSS 16.0. Advanced Statistical Procedures Companion*. Upper Saddle River, NJ: Prentice Hall.

Rasbash, J., Steele, F., Browne,W.J. & Goldstein, H. (2019). *A User's Guide to MLwiN. Version 3.03*. Online-Dokument: http://www.bristol.ac.uk/cmm/media/software/mlwin/downloads/manuals/3-03/manual-web.pdf

Raudenbush, S.W. & Bryk, A.S. (2002). *Hierarchical Linear Models* (2nd ed.). Thousand Oaks: CA: Sage.

Raudenbush, S.W., Bryk, A.S., Cheong, Y.F., Congdon, R.T & du Toit, M. (2011). *HLM 7: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15, pp. 351–357.

Singer, J. D. (1998). Using SAS PROC MIXED to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models. *Journal of Educational and Behavioral Statistics*, 23(4), 323-355.

Singer, J. D. & Willett, J.B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford University Press.

Snijders, T.A.B. (2005). *Power and sample size in multilevel modeling*. Online-Dokument (abgerufen am 28.12.2017): www.stats.ox.ac.uk/~snijders/PowerSampleSizeMultilevel.pdf
Erschienen in: B.S. Everitt & D.C. Howell (Eds.), Encyclopedia of Statistics in Behavioral Science. Volume 3, pp. 1570 - 1573. Chicester, UK: Wiley.

Snijders, T.A.B. & Bosker, R.J. (2012). *Multilevel Analysis* (2nd ed.). London: Sage.

Tabachnick, B.G. & Fidell, L.S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson.

Weiss, J. (2010). *Statistical Methods in Ecology*. Lecture 28. Online-Dokument (abgerufen: 28.12.2012): http://www.unc.edu/courses/2010fall/ecol/563/001/docs/lectures/lecture28.htm

West, B.T., Welch, K.B. & Galecki, A.T. (2015). *Linear Mixed Models. A Practical Guide Using Statistical Software* (2nd ed.). Boca Raton, FL: CRC Press.

Willett, J.B. (1988). Questions and answers in the measurement of change. In E. Rothkopf (Ed.), *Review of research in education (1988-89)* (pp. 345-422). Washington, DC: American Educational Research Association.

Wooldridge, J. M. (2013). *Introductory Econometrics: A modern approach* (5th ed.). Cengage Learning.

Index