5. Vorlesung Statistik II
Weitere Verfahren im einfaktoriellen varianzanalytischen Modell und zweifaktorielles varianzanalytisches Modell

Plan für heute

- Im ersten Teil der VL werden wir weitere statistische Verfahren im einfaktoriellen varianzanalytischen Modell kennen lernen, mit denen wir häufig interessantere inhaltliche Fragestellungen untersuchen können, als mit dem Omnibustest.

- Im zweiten Teil der VL werden wir das zweifaktorielle varianzanalytische Modell kennenlernen und im Rahmen dessen Interaktionseffekte zwischen Faktoren besprechen.

Beispiel zur Motivation I

- Man vermutet, dass verschiedene Therapieformen unterschiedlich erfolgreich in Bezug auf die Behandlung von Angststörungen sind. Es soll daher untersucht werden, inwiefern sich die mittlere (stetige) Angststärke nach einer Verhaltenstherapie (VT), nach einer Entspannungstherapie (ET) und nach einer tiefenpsychologischen Therapie (TT) unterscheidet.

- Statistisches Modell:

$$Y_{ij} = \mu_j + \epsilon_{ij}, \qquad \text{mit } j = VT, ET, TT \text{ und } i = 1, \dots, n_j \text{ und } Y_{ij} \overset{\text{ind}}{\sim} N(\mu_j, \sigma^2)$$

Beispiel zur Motivation II

- Folgende Fragestellungen könnten uns hier zum Beispiel interessieren:
  1) Zwischen welchen Therapieformen gibt es im Mittel Unterschiede in der Angststärke?
  2) Wie groß sind die mittleren Unterschiede in der Angststärke zwischen den Therapieformen jeweils?
  3) Ist die mittlere Angststärke in mindestens einer der beiden Psychotherapien (VT und TT) niedriger als in der ET?
  4) Können wird davon ausgehen, dass die mittlere Angststärke in der VT niedriger als in der TT ist und in dieser wiederum niedriger als in der ET?

- **Keine** dieser Fragen können wir mithilfe des Omnibus-Tests oder Schätzungen von η^2 beantworten.

Plan

- Wir werden diese Fragestellungen jeweils zunächst in Fragestellungen bezüglich inhaltlich interessierender Größen und daraufhin in Parameter des einfaktoriellen varianzanalytischen Modells in seiner ersten Darstellungsform übersetzen.

- Darauf aufbauend können wir dann inferenzstatistische Verfahren (Konfidenzintervalle und Hypothesentests) herleiten, mithilfe derer wir die Fragestellungen direkt beantworten können.

Neue Begrifflichkeit: Estimand

- Bisher haben wir uns bei der Beantwortung von Fragestellungen immer mit „Estimates“, also Schätzwerten beschäftigt.
- Bevor wir uns nun anschauen, wie wir inhaltliche Fragestellungen jeweils in Fragestellungen bezüglich der Parameter eines statistischen Modells übersetzen, müssen wir zunächst definieren, **welche Größe uns inhaltlich genau interessiert**.
- Diese exakt definierte, inhaltliche Größe nennen wir „**Estimand**“, also „das zu schätzende“.
- In den meisten Fällen wird der Estimand durch **zwei Bestandteile** definiert:

  1. Eine genaue Zielpopulation (vgl. VL Statistik 1)
  2. Eine personenspezifische Zielgröße

Sehr gutes Paper zu Estimands: Lundberg, I., Johnson, R., & Stewart, B. M. (2021). What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory. *American Sociological Review*, *86*(3), 532-565. https://doi.org/10.1177/00031224211004187

Neue Begrifflichkeit: Estimand

- Beispiel: Wir interessieren uns für die durchschnittliche Intelligenz aller Studierenden in Deutschland. In diesem Falle ist der Estimand definiert als

$$\frac{1}{N}\sum_{i=1}^{N} Y_i$$

Mittelwert über alle N Studierenden in Deutschland (Zielpopulation)

Die Intelligenz einer jeden Studentin i (personenspezifische Zielgröße)

- Der Estimand hat zunächst nichts mit den Parametern eines statistischen Modells zu tun. Häufig wird der Estimand in psychologischen Studien aber gar nicht genau definiert. Stattdessen wird sofort in Form von statistischen Parametern „gedacht".
- Das hat zur Folge, dass man in der Schätzung des Estimands an ein bestimmtes statistisches Modell gebunden ist – und damit an alle Annahmen, die dieses Modell definieren.
- Dabei ist es sinnvoller, sich zunächst auf inhaltlicher Ebene (also losgelöst von bestimmten statistischen Modellen) zu überlegen, welche Größe von Interesse ist. Diese Überlegungen sollten außerdem **vor der Datenerhebung** geschehen!
- Aus diesen Überlegungen lassen sich dann statistische Modelle ableiten, die zur Ermittlung des Estimands (und somit zur Beantwortung der inhaltlichen Fragestellung) geeignet sind.

Fragestellung 1

- Zwischen welchen Therapieformen gibt es im Mittel Unterschiede in der Angststärke?
- In diesem Fall interessieren uns die mittleren Unterschiede zwischen allen möglichen Populationen:
  - Der mittlere Unterschied in der Angststärke zwischen VT und TT. Estimand:

$$\frac{1}{N_{VT}} \sum_{i=1}^{N_{VT}} Y_{i\,VT} - \frac{1}{N_{TT}} \sum_{i=1}^{N_{TT}} Y_{i\,TT}$$

  - Der mittlere Unterschied in der Angststärke zwischen VT und ET. Estimand analog.
  - Der mittlere Unterschied in der Angststärke zwischen TT und ET. Estimand analog.
- Ein mögliches Modell für diese Fragestellung ist das varianzanalytische Modell, da wir uns für Mittelwertsunterschiede interessieren. Übersetzt in Modellparameter der Varianzanalyse (erste Darstellungsform):
  - $\mu_{VT} - \mu_{TT}$
  - $\mu_{VT} - \mu_{ET}$
  - $\mu_{TT} - \mu_{ET}$

Fragestellung 1

Da wir uns dafür interessieren, **ob** es **jeweils** Unterschiede gibt, können wir diese Fragestellung durch die Überprüfung der folgenden ungerichteten Hypothesen beantworten:

$$H_0: \mu_{VT} - \mu_{TT} = 0$$

$$H_1: \mu_{VT} - \mu_{TT} \neq 0$$

$$H_0: \mu_{VT} - \mu_{ET} = 0$$

$$H_1: \mu_{VT} - \mu_{ET} \neq 0$$

$$H_0: \mu_{TT} - \mu_{ET} = 0$$

$$H_1: \mu_{TT} - \mu_{ET} \neq 0$$

Fragestellung 2

- Wie groß sind die mittleren Unterschiede in der Angststärke zwischen den Therapieformen jeweils?
- Auch in diesem Fall interessieren uns die mittleren Unterschiede zwischen allen möglichen Populationen:
  - Der mittlere Unterschied in der Angststärke zwischen VT und TT. Der Estimand ist derselbe wie in Fragestellung 1.
  - Der mittlere Unterschied in der Angststärke zwischen VT und ET.
  - Der mittlere Unterschied in der Angststärke zwischen TT und ET.
- Übersetzt in Modellparameter der Varianzanalyse (erste Darstellungsform):
  - $\mu_{VT} - \mu_{TT}$
  - $\mu_{VT} - \mu_{ET}$
  - $\mu_{TT} - \mu_{ET}$
- Da wir uns dafür interessieren, **wie groß** die Unterschiede **jeweils** sind, können wir diese Fragestellung mithilfe von Konfidenzintervallen für die Parameterdifferenzen $\mu_{VT} - \mu_{TT}$, $\mu_{VT} - \mu_{ET}$ und $\mu_{TT} - \mu_{ET}$ beantworten.

Fragestellung 3

- Ist die mittlere Angststärke in mindestens einer der beiden Psychotherapien (VT und TT) niedriger als in der ET?

- In diesem Fall interessieren uns die mittleren Unterschiede zwischen zwei Populationspaaren:
  - Der mittlere Unterschied in der Angststärke zwischen VT und ET.
  - Der mittlere Unterschied in der Angststärke zwischen TT und ET.

- Übersetzt in Modellparameter der Varianzanalyse (erste Darstellungsform):
  - $\mu_{VT} - \mu_{ET}$
  - $\mu_{TT} - \mu_{ET}$

Fragestellung 3

- Da wir uns dafür interessieren, **ob** die mittlere Angststärke in **mindestens einer** der beiden Psychotherapien niedriger als in der ET ist, können wir diese Fragestellung durch die Überprüfung der folgenden zusammengesetzten Hypothesen beantworten:

$$H_0: \mu_{VT} \geq \mu_{ET} \; und \; \mu_{TT} \geq \mu_{ET}$$

$$H_1: \mu_{VT} < \mu_{ET} \; oder \; \mu_{TT} < \mu_{ET}$$

- Dazu überprüfen wir die gerichteten Einzelhypothesen:

$$H_0: \mu_{VT} - \mu_{ET} \geq 0$$

$$H_1: \mu_{VT} - \mu_{ET} < 0$$

$$H_0: \mu_{TT} - \mu_{ET} \geq 0$$

$$H_1: \mu_{TT} - \mu_{ET} < 0$$

Fragestellung 4

- Können wird davon ausgehen, dass die mittlere Angststärke in der VT niedriger als in der TT ist und in dieser wiederum niedriger als in der ET?

- In diesem Fall interessieren uns die mittleren Unterschiede zwischen zwei Populationspaaren:
  - Der mittlere Unterschied in der Angststärke zwischen VT und TT.
  - Der mittlere Unterschied in der Angststärke zwischen TT und ET.

- Übersetzt in Modellparameter:
  - $\mu_{VT} - \mu_{TT}$
  - $\mu_{TT} - \mu_{ET}$

Fragestellung 4

- Da wir uns dafür interessieren, **ob** die mittlere Angststärke in der VT niedriger als in der TT ist **und** in der TT niedriger als in der ET, können wir diese Fragestellung durch die Überprüfung der folgenden zusammengesetzten Hypothesen beantworten:

$$H_0: \mu_{VT} \geq \mu_{TT} \; oder \; \mu_{TT} \geq \mu_{ET}$$

$$H_1: \mu_{VT} < \mu_{TT} \; und \; \mu_{TT} < \mu_{ET}$$

- Dazu überprüfen wir die gerichteten Einzelhypothesen:

$$H_0: \mu_{VT} - \mu_{TT} \geq 0$$
$$H_1: \mu_{VT} - \mu_{TT} < 0$$

$$H_0: \mu_{TT} - \mu_{ET} \geq 0$$
$$H_1: \mu_{TT} - \mu_{ET} < 0$$

Einfache Parameterdifferenzen

- Im Rahmen dieser Vorlesung werden wir uns nur die Konfidenzintervalle und Hypothesentests für einfache Parameterdifferenzen der Form $\mu_j - \mu_k$ genauer anschauen und auch diese nur im balancierten Design.

- Die Analysen sind in R jedoch auch für kompliziertere Parameterkombinationen relativ einfach.

Schätzfunktionen für einfache Parameterdifferenzen

- Die Schätzfunktion für die Parameterdifferenz $\mu_j - \mu_k$ ist einfach die Mittelwertsdifferenz in den Stichproben j und k:

$$\bar{Y}_j - \bar{Y}_k$$

- Der Schätzwert für $\mu_j - \mu_k$ ist also die Realisation

$$\bar{y}_j - \bar{y}_k$$

- Beispiel:
  Wollen wir die Parameterdifferenz $\mu_{VT} - \mu_{TT}$ in unserem Beispiel schätzen, verwenden wir als Schätzwert einfach die Differenz der mittleren Angststärke in der VT-Stichprobe und der mittleren Angststärke in der TT-Stichprobe: $\bar{y}_{VT} - \bar{y}_{TT}$.

Konfidenzintervalle für einfache Parameterdifferenzen

- Ein Konfidenzintervall mit Konfidenzniveau $1-\alpha$ für $\mu_j - \mu_k$ ist (für balancierte Designs)

$$\left[(\bar{Y}_j - \bar{Y}_k) - t_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{S_{pool}^2}{n} + \frac{S_{pool}^2}{n}}, (\bar{Y}_j - \bar{Y}_k) + t_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{S_{pool}^2}{n} + \frac{S_{pool}^2}{n}}\right]$$

wobei $t_{1-\frac{\alpha}{2}}$ das $\left(1-\frac{\alpha}{2}\right)$-Quantil einer t-Verteilung mit $\nu = m(n-1)$ ist.

- Dieses Konfidenzintervall entspricht dem Konfidenzintervall für die Parameterdifferenz zweier Populationen für unabhängige Stichproben aus Statistik I.

- Aber:
    - S_{pool}^2 wird hier aus allen m Stichproben berechnet und nicht nur aus den Stichproben j und k.
    - Der Parameter ν der verwendeten t-Verteilung wird daher anders berechnet.

Hypothesentests für einfache Parameterdifferenzen

- Die Teststatistik für den Hypothesentest bezüglich $\mu_j - \mu_k$ ist (für balancierte Designs)

$$T = \frac{\left(\bar{Y}_j - \bar{Y}_k\right) - \mu_0}{\sqrt{\frac{S^2_{pool}}{n} + \frac{S^2_{pool}}{n}}}$$

- Sie folgt unter der H_0 einer t-Verteilung mit $\nu = m(n-1)$.

- Diese Teststatistik entspricht der Teststatistik des t-Tests für unabhängige Stichproben aus Statistik I.
- Aber:
  - S^2_{pool} wird hier aus allen m Stichproben berechnet und nicht nur aus den Stichproben j und k.
  - Der Parameter ν der verwendeten t-Verteilung wird daher anders berechnet.

Fragestellung 1 in R

- Zwischen **welchen** Therapieformen gibt es im Mittel Unterschiede in der Angststärke?

```
	 Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: User-defined Contrasts


Fit: aov(formula = angst ~ therapie, data = daten)

Linear Hypotheses:
             Estimate Std. Error t value             Pr(>|t|)
VT - TT == 0  -1.9428     0.8019  -2.423               0.0156 *
VT - ET == 0  -9.8561     0.8019 -12.291 <0.0000000000000002 ***
TT - ET == 0  -7.9133     0.8019  -9.868 <0.0000000000000002 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Univariate p values reported)
```

Unkorrigierte p-Werte

Für den Code zur Berechnung in R, siehe Übungsblatt 5

- Testentscheidungen bei einem Signifikanzniveau von $\alpha = 0.005$?

Fragestellung 1 in R

- Auf der Basis der p-Werte entscheiden wir uns für
  - die $H_0: \mu_{VT} - \mu_{TT} = 0$
  - die $H_1: \mu_{VT} - \mu_{ET} \neq 0$
  - die $H_1: \mu_{TT} - \mu_{ET} \neq 0$

- Wir entscheiden uns also dafür, dass sich die mittlere Angststärke in der ET von der mittleren Angststärke sowohl in VT als auch in TT unterscheidet, es aber keinen Unterschied zwischen VT und TT gibt.

- Bemerkung: Da wir uns für alle Einzelhypothesen interessieren, haben wir das Signifikanzniveau der einzelnen Hypothesentests nicht korrigiert. In diesem Fall ist die interessante zu kontrollierende Größe die FDR und nicht die FWER (siehe „Jelly Beans" Beispiel in Vorlesung 2). Daher ist es wie immer wichtig, neben der Verwendung eines niedrigen Signifikanzniveaus ($\alpha = 0.005$) auch eine hohe Power ($1 - \beta = 0.8$) durch Erhebung einer großen Stichprobe sicherzustellen.

Fragestellung 2 in R

- Wie groß sind die mittleren Unterschiede in der Angststärke zwischen den Therapieformen jeweils?

```
  Simultaneous Confidence Intervals

Multiple Comparisons of Means: User-defined Contrasts


Fit: aov(formula = angst ~ therapie, data = daten)

Quantile = 1.9626
95% confidence level


Linear Hypotheses:
              Estimate lwr      upr
VT - TT == 0  -1.9428  -3.5167  -0.3690
VT - ET == 0  -9.8561  -11.4300 -8.2823
TT - ET == 0  -7.9133  -9.4872  -6.3395
```

Untere Grenzen der KIs

Obere Grenzen der KIs

- Interpretation?

Fragestellung 2 in R

- Die plausiblen Werte für den Unterschied in der mittleren Angststärke zwischen VT und TT liegen zwischen -3.52 und -0.37.

- Die plausiblen Werte für den Unterschied in der mittleren Angststärke zwischen VT und ET liegen zwischen -11.43 und -8.28.

- Die plausiblen Werte für den Unterschied in der mittleren Angststärke zwischen TT und ET liegen zwischen -9.49 und -6.34.

Fragestellung 3 in R

- Ist die mittlere Angststärke in **mindestens einer** der beiden Psychotherapien (VT und TT) niedriger als in der ET?

```
                 Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: User-defined Contrasts


Fit: aov(formula = angst ~ therapie, data = daten)

Linear Hypotheses:
             Estimate Std. Error t value        Pr(<t)
VT - ET >= 0  -9.8561     0.8019 -12.291 <0.0000000001 ***
TT - ET >= 0  -7.9133     0.8019  -9.868 <0.0000000001 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Adjusted p values reported -- single-step method)
```

Korrigierte p-Werte mit der **Tukey Methode**, da zusammengesetzte Alternativhypothese mit „oder“

- Testentscheidungen bei einem Signifikanzniveau von $\alpha^* = 0.005$?

Fragestellung 3 in R

- Auf Basis der mit der **Tukey Methode** korrigierten p-Werte entscheiden wir uns für
  - die $H_1\!: \mu_{VT} - \mu_{ET} < 0$
  - die $H_1\!: \mu_{TT} - \mu_{ET} < 0$

- Wir entscheiden uns also für die zusammengesetzte Alternativhypothese, dass die mittlere Angststärke in **mindestens einer** der beiden Psychotherapien (VT und TT) niedriger ist als in der ET.
- Bemerkung 1: Da wir die Hypothesen in diesem Beispiel im Rahmen des einfaktoriellen varianzanalytischen Modells aufgestellt haben, können wir (im Gegensatz zum Beispiel aus Vorlesung 1) die Tukey Methode verwenden.
- Bemerkung 2: Es wäre hier ebenfalls möglich, die Bonferroni Methode zur Korrektur der p-Werte zu verwenden. Die Power für den zusammengesetzten Hypothesentest wäre dann aber niedriger, da die Bonferroni-Methode mögliche Abhängigkeiten zwischen den Hypothesentests nicht explizit berücksichtigen kann.

Fragestellung 3 in R

- Bemerkung 3: Interessieren wir uns zusätzlich zu der zusammengesetzten Hypothese auch noch für die Einzelhypothesen, also **welche** der Psychotherapien (VT und TT) besser funktionieren als die ET, ist dafür das Signifikanzniveau des zusammengesetzten Hypothesentests α^* nicht relevant. Die interessante zu kontrollierende Größe für die Interpretation der Einzelhypothesen ist wieder die FDR und nicht die FWER (siehe „Jelly Beans“ Beispiel in Vorlesung 2).

Fragestellung 4 in R

- Können wird davon ausgehen, dass die mittlere Angststärke in der VT niedriger als in der TT ist **und** in dieser wiederum niedriger als in der ET?

```
         Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: User-defined Contrasts


Fit: aov(formula = angst ~ therapie, data = daten)

Linear Hypotheses:
             Estimate Std. Error t value                Pr(<t)
VT - TT >= 0  -1.9428     0.8019  -2.423                0.0078 **
TT - ET >= 0  -7.9133     0.8019  -9.868 <0.0000000000000002 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Univariate p values reported)
```

Unkorrigierte p-Werte, da zusammengesetzte Alternativhypothese mit „und“

- Testentscheidungen bei einem Signifikanzniveau von $\alpha^* = 0.005$?

Fragestellung 4 in R

- Auf der Basis der unkorrigierten p-Werte entscheiden wir uns für
  - die $H_0: \mu_{VT} - \mu_{TT} \geq 0$
  - die $H_1: \mu_{TT} - \mu_{ET} < 0$

- Wir entscheiden uns also für die zusammengesetzte Nullhypothese. Wir gehen **nicht** davon aus, dass die Angststärke in der VT im Mittel niedriger als in der TT ist und in dieser wiederum im Mittel niedriger als in der ET.

Effektstärken und Stichprobenplanung

- Effektstärken: Für einfache Parameterdifferenzen $\mu_j - \mu_k$ kann das schon aus Statistik I bekannte

$$\delta = \frac{\mu_j - \mu_k}{\sigma}$$

mit Schätzfunktion

$$\hat{\delta} = \frac{\bar{Y}_j - \bar{Y}_k}{\sqrt{S^2_{pool}}}$$

verwendet werden.

- Stichprobenplanung: Für einfache Parameterdifferenzen kann mithilfe von δ jeweils auf die in Statistik I besprochenen Verfahren für Hypothesentests und Konfidenzintervalle im Zwei-Stichprobenfall zurückgegriffen werden.

- Bemerkung: Sind die Annahmen des einfaktoriellen varianzanalytischen Modells erfüllt, ist die wahre Power mit der so berechneten Stichprobengröße sogar noch etwas höher als geplant.

Bemerkungen

- Falls mehrere Hypothesen innerhalb eines Modells getestet werden, kann es vorkommen, dass die Testentscheidungen logisch nicht konsistent sind. Beispielsweise kann es sein, dass wir uns auf der Basis der p-Werte für die Hypothesen

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_0: \mu_2 - \mu_3 = 0$$

$$H_1: \mu_1 - \mu_3 \neq 0$$

  entscheiden.

- Die heute besprochenen varianzanalytischen Verfahren werden in der Literatur u.a. auch a-priori Vergleiche, geplante Vergleiche, a-posteriori Vergleiche, posthoc Vergleiche oder Kontrastanalysen genannt.

- In der Literatur spricht man bei der Korrektur von p-Werten manchmal nur von der Tukey Methode, wenn die FWER für alle möglichen paarweisen Vergleiche kontrolliert wird.

Fazit

Vorgehen bei der Untersuchung von Forschungsfragen mithilfe des (einfaktoriellen) varianzanalytischen Modells**:**

- Genau überlegen, welche inhaltlichen Größen in der Population interessieren (d.h. den Estimand definieren).
- Feststellen, ob die Forschungsfrage im Rahmen eines (einfaktoriellen) varianzanalytischen Modells untersucht werden kann.
- Falls ja, diese Größen in Parameter des Modells übersetzen.
- Das entsprechende inferenzstatistische Verfahren auf diese Parameter anwenden.
- Inhaltliche Abwägung zwischen Hypothesentestung und Intervallschätzung.
- **Auch wenn Hypothesen getestet werden sollen, wird das passende varianzanalytische Verfahren fast nie der Omnibustest sein.**

Zwischengliederung

- Bislang:
  - Weitere statistische Verfahren im einfaktoriellen varianzanalytischen Modell
- Jetzt:
  - Zweifaktorielles varianzanalytisches Modell

Vorbemerkung: Zweifaktorielles varianzanalytisches Modell

- Mithilfe von varianzanalytischen Modellen, die mehr als einen Faktor enthalten, lassen sich so genannte **Interaktionen** untersuchen.
- Unter einer Interaktion versteht man im Rahmen von zweifaktoriellen varianzanalytischen Modellen die **Wechselwirkung zwischen zwei Faktoren**.
- Eine solche Wechselwirkung liegt vor, wenn der **Zusammenhang** zwischen **einem** der Faktoren und der **AV** davon abhängt, welche **Ausprägung** auf dem **anderen** Faktor vorliegt.

Beispiel

Es soll auf Basis von zufällig ausgewählten Personen die folgende inhaltliche Frage untersucht werden:

- Hängt es von der Art des Unterrichtsfachs ab, wie effektiv (bezogen auf die Wissens-vermittlung) die eingesetzten Präsentationsmethoden sind?
- UV 1 (zweifach gestuft); Faktor A: „Fach“: Geschichte vs. Mathe
- UV 2 (zweifach gestuft); Faktor B: „Methode“: Tafel vs. Powerpoint
- AV: „Wissen“: Wissenstest zur Erfassung des fachspezifischen Wissens (Wertebereich von null bis 100)

Anzahl der Populationen für das Beispiel

- Die **Anzahl der Faktoren** und die **Anzahl der Stufen** bestimmen die **Anzahl der Populationen**, aus denen die Zufallsstichproben gezogen werden.
- Bezogen auf unser Beispiel können vier unterschiedliche Populationen betrachtet werden:
  - Population der Personen, bei denen im Mathematikunterricht die Tafel verwendet wird.
  - Population der Personen, bei denen im Mathematikunterricht PP-Folien verwendet werden.
  - Population der Personen, bei denen im Geschichtsunterricht die Tafel verwendet wird.
  - Population der Personen, bei denen im Geschichtsunterricht PP-Folien verwendet werden.

Statistisches Modell für ein zweifaktorielles Design: Erste Darstellungsform

- Im Falle eines *r*-fach und eines *c*-fach gestuften Faktors lautet die statistische Modellgleichung in der ersten Darstellung:

$$Y_{ijk} = \mu_{jk} + \varepsilon_{ijk} \text{ mit } i = 1, \ldots, n_{jk};\; j = 1, \ldots, r \text{ und } k = 1, \ldots, c$$

- Y_{ijk} ist eine **Zufallsvariable**, deren Realisation der AV-Wert einer Person *i* ist, die aus der Population jk von *r* x *c* Populationen zufällig gezogen wird.
- μ_{jk} ist der **Erwartungswert** der AV in der Stichprobe jk und gleichzeitig der Mittelwert der AV in der Population jk. μ_{jk} ist ein **Parameter** und somit eine **Konstante**.
- ε_{ijk} repräsentiert den so genannten **Fehler**, d.h. alle nicht berücksichtigten Einflüsse auf die AV bezogen auf die Person *i*, die aus der Population jk zufällig gezogen wurde.
- ε_{ijk} ist eine **Zufallsvariable**.

Beispiel erste Darstellungsform

- Beispiel von vorher: AV = Wissen, $r = 2$ Stufen auf dem Faktor „Fach“ (1 = Geschichte, 2 = Mathe), $c = 2$ Stufen auf dem Faktor „Methode“ (1 = Tafel, 2 = PP).

$$Y_{ijk} = \mu_{jk} + \varepsilon_{ijk}$$

- μ_{jk} ist das mittlere Wissen in Population jk. Das heißt:
  - μ_{11} ist das mittlere Wissen in der Population der Personen, bei denen im Geschichtsunterricht die Tafel verwendet wird.
  - μ_{12} ist das mittlere Wissen in der Population der Personen, bei denen im Geschichtsunterricht PP-Folien verwendet werden.
  - μ_{21} ist das mittlere Wissen in der Population der Personen, bei denen im Matheunterricht die Tafel verwendet wird.
  - μ_{22} ist das mittlere Wissen in der Population der Personen, bei denen im Matheunterricht PP-Folien verwendet werden.

Statistisches Modell für ein zweifaktorielles Design: Zweite Darstellungsform I

Bevor wir die statistische Modellgleichung in der zweiten Darstellung einführen, zunächst ein paar Definitionen:

- Mittlere AV über alle Populationen hinweg:

$$\mu = \frac{1}{r \cdot c} \sum_{j=1}^{r} \sum_{k=1}^{c} \mu_{jk}$$

- Mittlere AV über alle Populationen mit Faktorstufe j auf dem ersten Faktor:

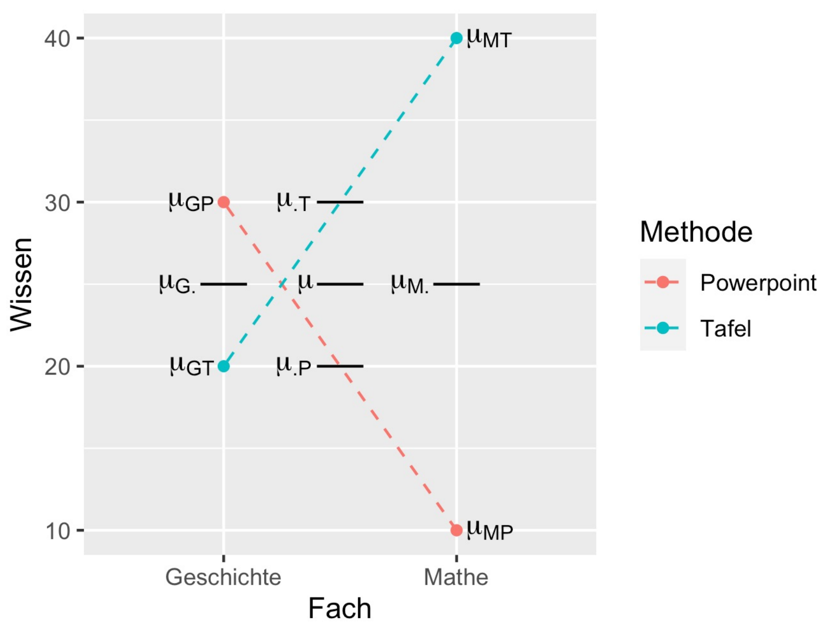
$$\mu_{j\cdot} = \frac{1}{c} \sum_{k=1}^{c} \mu_{jk}$$

- Mittlere AV über alle Populationen mit Faktorstufe k auf dem zweiten Faktor:

$$\mu_{\cdot k} = \frac{1}{r} \sum_{j=1}^{r} \mu_{jk}$$

Statistisches Modell für ein zweifaktorielles Design: Zweite Darstellungsform II

	T	PP	
G	μ_{11}	μ_{12}	$\mu_{1\cdot}$
M	μ_{21}	μ_{22}	$\mu_{2\cdot}$
	$\mu_{\cdot 1}$	$\mu_{\cdot 2}$	μ



Für unser Beispiel:

- μ ist das mittlere Wissen aller Personen.
- $\mu_{1\cdot}$ ist das mittlere Wissen der Personen im Geschichtsunterricht.
- $\mu_{2\cdot}$ ist das mittlere Wissen der Personen im Matheunterricht.
- $\mu_{\cdot 1}$ ist das mittlere Wissen der Personen, bei denen die Tafel verwendet wurde.
- $\mu_{\cdot 2}$ ist das mittlere Wissen der Personen, bei denen PP-Folien verwendet wurden.

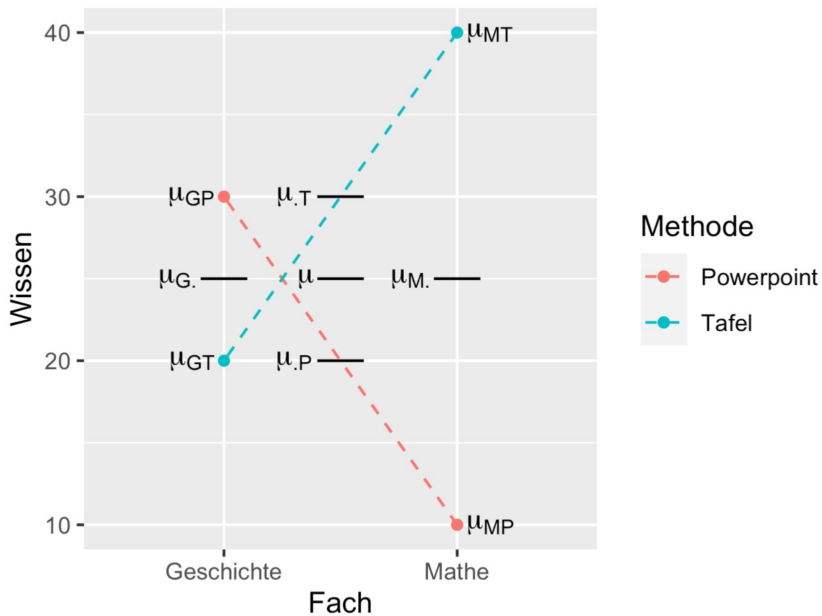
Statistisches Modell für ein zweifaktorielles Design: Zweite Darstellungsform III

Im Falle von zwei Faktoren, die *r*-fach bzw. *c*-fach gestuft sind, lautet die statistische Modellgleichung in der zweiten Darstellung:

$$Y_{ijk} = \mu + \alpha_j + \beta_k + \gamma_{jk} + \varepsilon_{ijk}, \qquad \text{mit } i = 1, \dots, n_{jk};\ j = 1, \dots, r \ \text{ und } \ k = 1, \dots, c$$

- $\alpha_j = \mu_{j.} - \mu$ repräsentiert den **Effekt der Kategorie *j*** des Faktors A auf die AV.
- $\beta_k = \mu_{.k} - \mu$ repräsentiert den **Effekt der Kategorie *k*** des Faktors B auf die AV.
- $\gamma_{jk} = \mu_{jk} - \mu - \alpha_j - \beta_k$ repräsentiert den **Interaktionseffekt** bezogen auf die Population *jk*.

Beispiel zweite Darstellungsform



$$Y_{ijk} = \mu + \alpha_j + \beta_k + \gamma_{jk} + \varepsilon_{ijk}$$

- μ repräsentiert das durchschnittliche Wissen aller Personen.
- α_j repräsentiert den Effekt des Fachs *j* auf das Wissen.
- β_k repräsentiert den Effekt der Methode *k* auf das Wissen.
- γ_{jk} repräsentiert den Effekt der Wechselwirkung zwischen Fach *j* und Methode *k* auf das Wissen.

Statistisches Modell für ein zweifaktorielles Design: Annahmen

- In beiden Darstellungsformen treffen wir zudem die Annahme, dass die AV in allen Populationen die gleiche Varianz aufweist und ihr Histogramm innerhalb jeder Population durch die Dichtefunktion einer Normalverteilung approximiert werden kann.
- Falls wir aus allen Populationen jeweils eine einfache Zufallsstichprobe ziehen, führt diese Annahme auf der Ebene des statistischen Modells
  - in der ersten Darstellung zu:

$$Y_{ijk} \overset{\text{ind}}{\sim} N\left(\mu_{jk}, \sigma^2\right) \text{ bzw. } \varepsilon_{ijk} \overset{\text{iid}}{\sim} N(0, \sigma^2)$$

  - in der zweiten Darstellung zu:

$$Y_{ijk} \overset{\text{ind}}{\sim} N\left(\mu + \alpha_j + \beta_k + \gamma_{jk}, \sigma^2\right) \text{ bzw. } \varepsilon_{ijk} \overset{\text{iid}}{\sim} N(0, \sigma^2)$$

- (Beides kann aufwendig mathematisch bewiesen werden)
- Wir haben also in beiden Fällen einen zusätzlichen Parameter σ^2, der für die Varianz der AV innerhalb der Populationen steht.

Statistisches Modell für ein zweifaktorielles Design: Parameter

- Das statistische Modell der Varianzanalyse enthält somit in beiden Darstellungen mehrere Parameter.
  - Erste Darstellung:
    - μ_{jk} für alle Populationen jk
    - σ^2
  - Zweite Darstellung:
    - μ
    - α_j für alle Populationen j
    - β_k für alle Populationen k
    - γ_{jk} für alle Populationen jk
    - σ^2
- Alle diese Parameter können geschätzt werden und für alle diese Parameter können statistische Hypothesentests konstruiert werden. Dies gilt auch für Kombinationen der Parameter, z.B. für Parameterdifferenzen.
- Je nach konkreter Fragestellung muss dann entschieden werden, welche Parameter geschätzt werden sollen bzw. welche Hypothesen getestet werden sollen.

Gliederung

- Bislang:
  - Zweifaktorielles varianzanalytisches Modell

- Jetzt:
  - Schätzfunktionen für die Parameter des Modells

Schätzfunktionen für die Parameter

- Wir beschränken uns auf die erste Darstellungsform und balancierte Designs mit $n_{jk} = n$ für alle jk.
- Für die unbekannten Erwartungswerte μ_{jk} verwenden wir jeweils die Stichprobenmittelwerte:

$$\hat{\mu}_{jk} = \bar{Y}_{jk} = \frac{1}{n}\sum_{i=1}^{n} Y_{ijk}$$

- Die Realisation dieser Schätzfunktion, also der Schätzwert für μ_{jk}, ist einfach der Mittelwert der AV in der aus Population jk gezogenen Stichprobe.
- Als Schätzfunktion für σ^2 verwenden wir wieder die gepoolte Schätzfunktion S_{pool}^2:

$$\hat{\sigma}^2 = S_{pool}^2 = \frac{1}{r \cdot c(n-1)}\left(\sum_{j=1}^{r}\sum_{k=1}^{c}\sum_{i=1}^{n}\left(Y_{ijk} - \bar{Y}_{jk}\right)^2\right)$$

Gliederung

- Bislang:
  - Zweifaktorielles varianzanalytisches Modell
  - Schätzfunktionen für die Parameter des Modells

- Jetzt:
  - Omnibustests im zweifaktoriellen varianzanalytischen Modell

Omnibustests

- Im zweifaktoriellen varianzanalytischen Modell können **drei verschiedene Omnibustests** durchgeführt werden. Die statistischen Hypothesen werden gewöhnlich für die Parameter der zweiten Darstellungsform formuliert.

- Omnibustest für die Interaktion:

  $H_0: \gamma_{jk} = 0$ für alle Kombinationen *jk*

  $H_1: \gamma_{jk} \neq 0$ für mindestens eine Kombination *jk*

- Omnibustest für den Faktor A (UV 1):

  $H_0: \alpha_j = 0$ für alle *j*

  $H_1: \alpha_j \neq 0$ für mindestens ein *j*

- Omnibustest für den Faktor B (UV 2):

  $H_0: \beta_k = 0$ für alle *k*

  $H_1: \beta_k \neq 0$ für mindestens ein *k*

Omnibustest für die Interaktion I

- Omnibustest für die Interaktion:

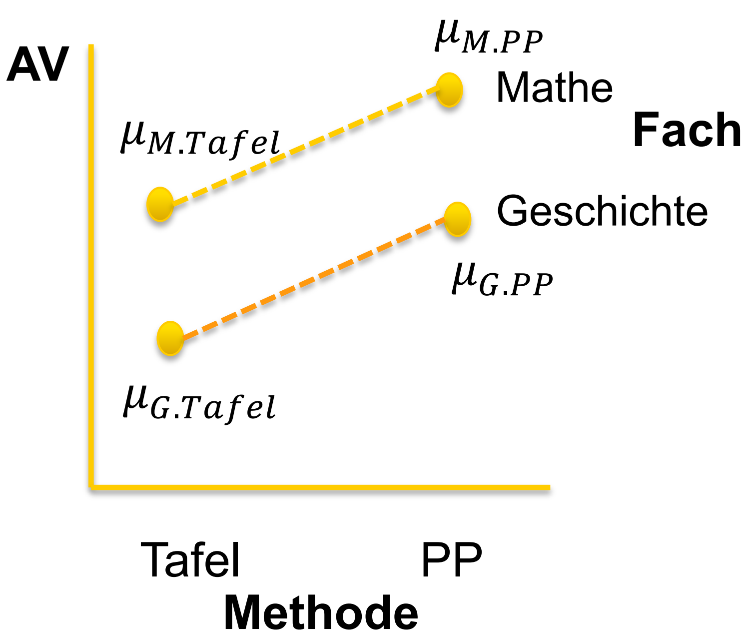
  $H_0: \gamma_{jk} = 0$ für alle Kombinationen *jk*

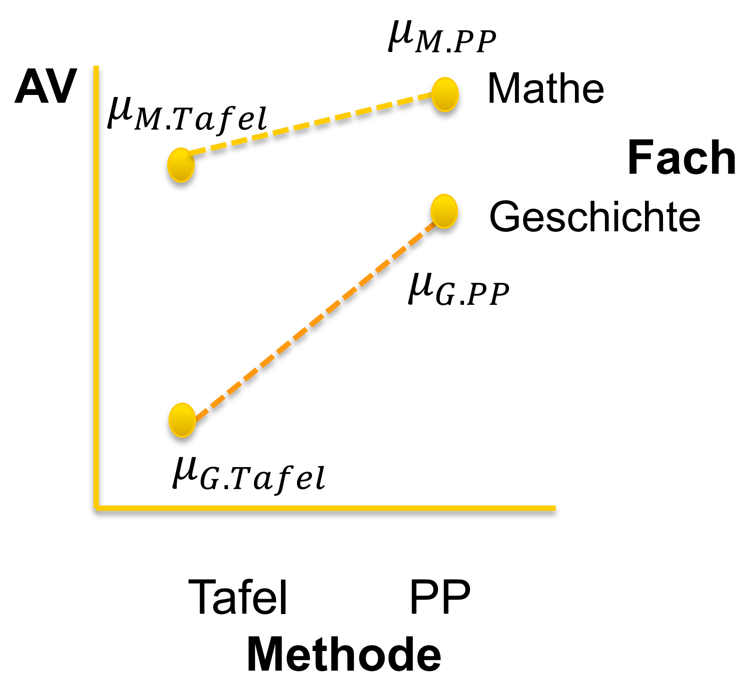
  $H_1: \gamma_{jk} \neq 0$ für mindestens eine Kombination *jk*

- Die H_0 des Omnibustests für die Interaktion gilt genau dann, wenn die Differenzen der Erwartungswerte μ_{jk} über die Faktorstufen des jeweils anderen Faktors konstant sind (etwas aufwendig zu beweisen).
- In diesem Fall liegt **keine Interaktion** d.h. keine Wechselwirkung der Faktoren vor: Der Einfluss der Faktoren auf die AV ist über die Stufen des jeweils anderen Faktors hinweg gleich.
- Die H_1 des Omnibustests für die Interaktion gilt genau dann, wenn die Differenzen der Erwartungswerte μ_{jk} über die Faktorstufen des jeweils anderen Faktors nicht konstant sind.
- In diesem Fall liegt eine **Interaktion** d.h. eine Wechselwirkung der Faktoren vor: Der Einfluss der Faktoren auf die AV unterscheidet sich über die Stufen des jeweils anderen Faktors hinweg. Man kann die Einflüsse der beiden Faktoren auf die AV nicht getrennt voneinander betrachten!

Omnibustest für die Interaktion II

Graphisch betrachtet liegt eine Interaktion also genau dann vor, wenn die **Erwartungswert-Profile nicht parallel** zueinander sind:





$$\mu_{M.Tafel} - \mu_{G.Tafel} = \mu_{M.PP} - \mu_{G.PP}$$
$$\mu_{M.PP} - \mu_{M.Tafel} = \mu_{G.PP} - \mu_{G.Tafel}$$

$$\mu_{M.Tafel} - \mu_{G.Tafel} \neq \mu_{M.PP} - \mu_{G.PP}$$
$$\mu_{M.PP} - \mu_{M.Tafel} \neq \mu_{G.PP} - \mu_{G.Tafel}$$

Omnibustest für den Haupteffekt von Faktor A

- Omnibustest für den Faktor A (UV 1):
  $H_0: \alpha_j = 0$ für alle *j*
  $H_1: \alpha_j \neq 0$ für mindestens ein *j*

- Wegen $\alpha_j = \mu_{j.} - \mu$ bedeutet die H_0 des Omnibustest für den Faktor A, dass es keine Unterschiede zwischen den gemittelten Erwartungswerten $\mu_{j.}$ der Faktorstufen des Faktors A gibt.

- Die H_1 des Omnibustest für den Faktor A bedeutet, dass es mindestens einen Unterschied zwischen den gemittelten Erwartungswerten $\mu_{j.}$ der Faktorstufen des Faktors A gibt. In diesem Fall sagt man auch, dass ein **Haupteffekt des Faktors A** vorliegt.

- Wichtig:
  - Die H_0 bedeutet **nicht unbedingt**, das der Faktor A keinen Einfluss auf die AV hat.
  - Die H_1 bedeutet **nicht unbedingt**, dass der Faktor A auf allen Faktorstufen des Faktors B einen Einfluss auf die AV hat.

Omnibustest für den Haupteffekt von Faktor B

- Omnibustest für den Faktor B (UV 2):
  $H_0: \beta_k = 0$ für alle *k*
  $H_1: \beta_k \neq 0$ für mindestens ein *k*

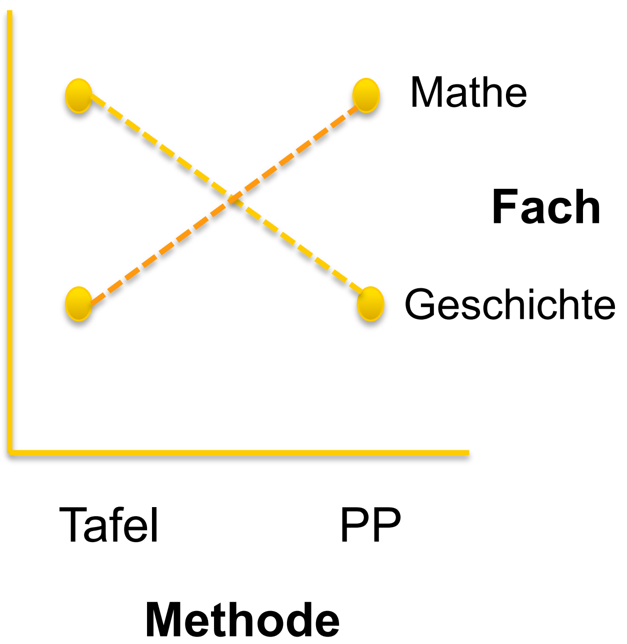
- Wegen $\beta_k = \mu_{.k} - \mu$ bedeutet die H_0 des Omnibustest für den Faktor B, dass es keine Unterschiede zwischen den gemittelten Erwartungswerten $\mu_{.k}$ der Faktorstufen des Faktors B gibt.

- Die H_1 des Omnibustest für den Faktor B bedeutet, dass es mindestens einen Unterschied zwischen den gemittelten Erwartungswerten $\mu_{.k}$ der Faktorstufen des Faktors B gibt. In diesem Fall sagt man auch, dass ein **Haupteffekt des Faktors B** vorliegt.

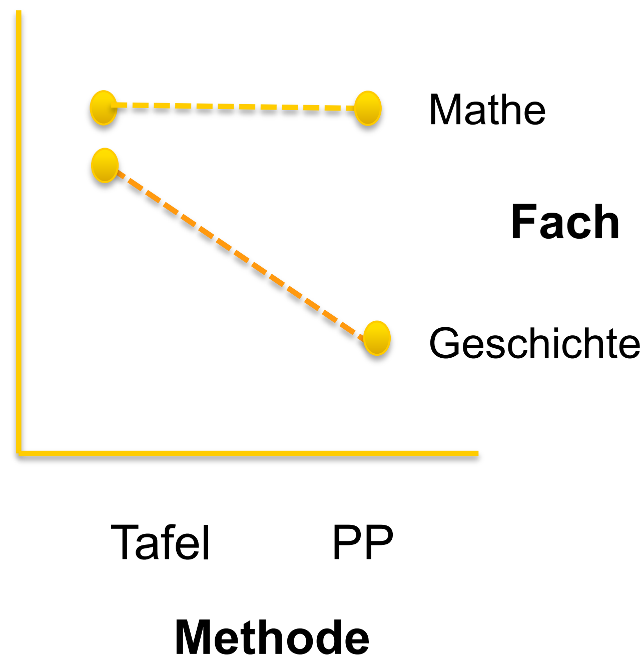
- Wichtig:
  - Die H_0 bedeutet **nicht unbedingt**, das der Faktor B keinen Einfluss auf die AV hat.
  - Die H_1 bedeutet **nicht unbedingt**, dass der Faktor B auf allen Faktorstufen des Faktors A einen Einfluss auf die AV hat.

Graphische Veranschaulichung Haupteffekte

Kein Haupteffekt Faktor Methode, aber Einfluss der Methode in beiden Fächern:



Haupteffekt Faktor Methode, aber kein Einfluss der Methode im Fach Mathe:



Fazit Omnibustests im zweifaktoriellen Modell

- Die Hypothesen des Omnibustests für die Interaktion können in Fällen interessant sein, in denen man sich dafür interessiert, ob es eine Wechselwirkung zwischen zwei Faktoren bei der Beeinflussung einer AV gibt.

- **Aber:** In solchen Fällen interessiert man sich meistens für spezifische Wechselwirkungen und sollte diese dann auch direkt überprüfen. Die Alternativhypothese des Omnibustests für die Interaktion besagt lediglich, dass irgendeine Form der Interaktion vorliegt.

- Die Hypothesen der beiden Omnibustests für die Haupteffekte sind bei vorliegender Interaktion für sich genommen nicht sinnvoll interpretierbar und auch bei nicht vorliegender Interaktion eher uninteressant, da wir in diesem Fall nur wissen, dass es irgendwelche Mittelwertsunterschiede gibt.

- Bemerkung: Da im zweifaktoriellen varianzanalytischen Modell mehrere Omnibustests durchgeführt werden können, kann es in sehr seltenen Fällen notwendig sein, die p-Werte der Omnibustests mit der Bonferroni Methode zu korrigieren.
  Beispiel: $H_1\colon \alpha_j \neq 0 \; oder \; \beta_k \neq 0$ (für mindestens ein *j* oder mindestens ein *k*)

Beispiel Omnibustests in R


```
              Df Sum Sq Mean Sq F value   Pr(>F)    
Fach           1   7825    7825  19.700 1.51e-05 ***
Methode        1    820     820   2.065   0.1523    
Fach:Methode   1   1711    1711   4.308   0.0392 *  
Residuals    196  77853     397                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- Testentscheidungen?

Gliederung

- Bislang:
  - Zweifaktorielles varianzanalytisches Modell
  - Schätzfunktionen für die Parameter des Modells
  - Omnibustests im zweifaktoriellen varianzanalytischen Modell
- Jetzt:
  - Weitere Verfahren im zweifaktoriellen varianzanalytischen Modell

Weitere varianzanalytische Verfahren

- Genau wie im einfaktoriellen Modell interessieren wir uns nur sehr selten für die Omnibushypothesen.
- Genau wie dort können wir jedoch für die uns tatsächlich interessierenden Größen direkt Konfidenzintervalle und Hypothesentests berechnen.
- Wir werden dies anhand einiger Beispiel-Fragestellungen illustrieren.
- Dabei werden wir alle uns interessierenden Größen in Parameter des Modells in der ersten Darstellungsform übersetzen.
- Ob es sinnvoll ist, im Falle der Durchführung mehrerer einzelner Hypothesentests die p-Werte zu korrigieren, richtet sich nach den gleichen Prinzipien wie bisher.

Fragestellung 1 I

- Zwischen welchen Populationen gibt es Unterschiede im mittleren Wissen?
- Übersetzung in statistische Alternativhypothesen:

$H_1: \mu_{G.Tafel} - \mu_{G.PP} \neq 0$

$H_1: \mu_{M.PP} - \mu_{G.PP} \neq 0$

$H_1: \mu_{M.Tafel} - \mu_{G.PP} \neq 0$

$H_1: \mu_{M.PP} - \mu_{G.Tafel} \neq 0$

$H_1: \mu_{M.Tafel} - \mu_{G.Tafel} \neq 0$

$H_1: \mu_{M.Tafel} - \mu_{M.PP} \neq 0$

- Bemerkung: In der Literatur spricht man bei der Korrektur von p-Werten manchmal nur von der Tukey Methode, wenn die FWER für alle möglichen paarweisen Vergleiche kontrolliert wird.

Fragestellung 1 II


```
         Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means:


Fit: aov(formula = Wissen ~ Fach_Methode, data = daten1)

Linear Hypotheses:
                                   Estimate Std. Error t value Pr(>|t|)
Geschichte_Tafel - Geschichte_PP == 0   -1.800      3.986  -0.452   0.9693
Mathe_PP - Geschichte_PP == 0          -18.360      3.986  -4.606   <0.001 ***
Mathe_Tafel - Geschichte_PP == 0        -8.460      3.986  -2.122   0.1496
Mathe_PP - Geschichte_Tafel == 0       -16.560      3.986  -4.155   <0.001 ***
Mathe_Tafel - Geschichte_Tafel == 0     -6.660      3.986  -1.671   0.3419
Mathe_Tafel - Mathe_PP == 0              9.900      3.986   2.484   0.0658 .
---
```

Testentscheidungen?

Korrektur p-Werte?

Fragestellung 2 I

- Ist das mittlere Wissen im Geschichtsunterricht mit PP-Folien höher als in allen anderen Populationen?

- Übersetzung in statistische Alternativhypothesen:

$$H_1: \mu_{G.PP} - \mu_{G.Tafel} > 0$$

$$H_1: \mu_{G.PP} - \mu_{M.Tafel} > 0$$

$$H_1: \mu_{G.PP} - \mu_{M.PP} > 0$$

Fragestellung 2 II


```
         Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means:


Fit: aov(formula = Wissen ~ Fach_Methode, data = daten1)

Linear Hypotheses:
                                      Estimate Std. Error t value Pr(>t)
Geschichte_PP - Geschichte_Tafel <= 0    1.800      3.986   0.452 0.5635
Geschichte_PP - Mathe_Tafel <= 0         8.460      3.986   2.122 0.0448 *
Geschichte_PP - Mathe_PP <= 0           18.360      3.986   4.606 <0.001 ***
---
```

Testentscheidungen?

Korrektur p-Werte?

Fragestellung 3 I

- Angenommen, wir wissen aus Voruntersuchungen, dass im Matheunterricht die Tafel und im Geschichtsunterricht PP-Folien besonders effektiv sind. Nun vermuten wir, dass der Unterschied zwischen den Methoden im Matheunterricht größer ist als im Geschichtsunterricht. Wir gehen also von einer spezifischen Interaktion aus.

- Übersetzung in eine statistische Alternativhypothese:

$$H_1\colon \mu_{M.Tafel} - \mu_{M.PP} > \mu_{G.PP} - \mu_{G.Tafel}$$

bzw.:

$$H_1\colon \mu_{M.Tafel} - \mu_{M.PP} - \mu_{G.PP} + \mu_{G.Tafel} > 0$$

Fragestellung 3 II


```
         Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: User-defined Contrasts


Fit: aov(formula = Wissen ~ Fach_Methode, data = daten1)

Linear Hypotheses:
                                                              Estimate Std. Error t value Pr(>t)
Mathe_Tafel - Mathe_PP - Geschichte_PP + Geschichte_Tafel <= 0   8.100      5.637   1.437 0.0762 .
---
```

Testentscheidungen?

Gliederung

- Bislang:
  - Zweifaktorielles varianzanalytisches Modell
  - Schätzfunktionen für die Parameter des Modells
  - Omnibustests im zweifaktoriellen varianzanalytischen Modell
  - Weitere Verfahren im zweifaktoriellen varianzanalytischen Modell

- Jetzt:
  - Exkurs: Stichprobenplanung
  - Exkurs: Varianzanalytische Modelle mit mehr als zwei Faktoren
  - Exkurs: Messwiederholungsdesigns
  - Fazit: Varianzanalytische Modelle

Exkurs: Stichprobenplanung

- Omnibustests:

  Stichprobenplanung für die Omnibustests ist im mehrfaktoriellen varianzanalytischen Modell noch etwas komplizierter als im einfaktoriellen Modell. Da die statistischen Hypothesen des Omnibustests so gut wie nie relevante inhaltliche Forschungsfragen beantworten können, verzichten wir hier auf die Beschreibung sowohl für den einfachen einfaktoriellen Fall als auch für den komplizierteren mehrfaktoriellen Fall.

- Weitere varianzanalytische Verfahren:

  Für einfache Parameterdifferenzen kann (genauso wie im einfaktoriellen varianzanalytischen Modell) auf die in Statistik I besprochenen Verfahren für Hypothesentests und Konfidenzintervalle im Zwei-Stichprobenfall zurückgegriffen werden.

Exkurs: Varianzanalytische Modelle mit mehr als zwei Faktoren

- Die Erweiterung varianzanalytischer Modelle auf mehr als zwei Faktoren funktioniert analog zur Erweiterung von einem auf zwei Faktoren.

- Mit jedem weiteren Faktor können, neben einem zusätzlichen Haupteffekt, weitere Interaktionen betrachtet werden. Beispiel: Bei 3 Faktoren A, B, C sind Omnibustests für die 4 Interaktionen AxB, AxC, BxC und AxBxC möglich.

- Die inhaltliche Interpretation der Omnibustests wird damit immer komplizierter.
  Wie auch in den bisher besprochenen Modellen, sind die theoretisch möglichen Omnibustests zur Beantwortung der meisten inhaltlichen Forschungsfragen irrelevant.

- Hypothesentests und Intervallschätzung für Parameterdifferenzen (oder Kombinationen von Parameterdifferenzen) funktionieren genauso wie im zweifaktoriellen Modell. Die Anzahl der für die Forschungsfrage zu betrachtenden Populationen kann unter Umständen aber sehr groß werden.

Exkurs: Messwiederholungsdesigns I

- Varianzanalytische Modelle werden deutlich komplizierter, wenn die Faktorstufen eines oder mehrerer Faktoren voneinander abhängig sind. Dieser Fall tritt am häufigsten in sogenannten **Messwiederholungsdesigns** auf.
- Da die Omnibustests (wie auch in den bisher besprochenen Modellen) zur Beantwortung der meisten inhaltlichen Forschungsfragen in Messwiederholungsdesigns nicht relevant sind, lohnt es sich für uns nicht, varianzanalytische Modelle mit abhängigen Faktoren zu besprechen.
- Hypothesentests und Parameterschätzung für Parameterdifferenzen können in der Praxis, auch im abhängigen Fall, sehr einfach mit einer Kombination der statistischen Verfahren aus Statistik I für zwei unabhängige oder abhängige Stichproben untersucht werden. Wie bereits erwähnt, sind die Verfahren aus Statistik I bei großen Stichproben sogar robust gegenüber möglichen Verletzungen statistischer Annahmen.
- Noch kompliziertere Forschungsfragen mit Messwiederholungsdesigns sollten besser im Rahmen sogenannter „gemischter linearer Modelle“ (Vorlesung Master) untersucht werden.

Exkurs: Messwiederholungsdesigns II

- **Klassisches Beispiel:**
  2 x 3 Design mit einem unabhängigen Faktor Therapie (Verhaltenstherapie, keine Therapie), einem abhängigen Faktor Zeit (Beginn der Therapie, Ende der Therapie, Follow-Up 6 Monate nach der Therapie) und der stetigen AV Depressionsschwere.

- **Sinnvolle Hypothese, falls gezeigt werden soll, dass die Verhaltenstherapie „wirkt“:**
  Die mittlere Depressionsschwere mit Verhaltenstherapie behandelter Patienten nimmt vom Beginn der Therapie über das Ende der Therapie bis zum Follow-Up immer weiter ab. Gleichzeitig ist die mittlere Depressionsschwere mit Verhaltenstherapie behandelter Patienten sowohl am Ende der Therapie, als auch beim Follow-Up niedriger als bei Patienten ohne Therapie.

- Zusammengesetzte Alternativhypothese mit „und“:

$H_{11}: \mu_{VT.Beginn} - \mu_{VT.Ende} > 0$ -> t-Test für abhängige Stichproben

$H_{12}: \mu_{VT.Ende} - \mu_{VT.FollowUp} > 0$ -> t-Test für abhängige Stichproben

$H_{13}: \mu_{VT.Ende} - \mu_{KT.Ende} < 0$ -> t-Test für unabhängige Stichproben

$H_{14}: \mu_{VT.FollowUp} - \mu_{KT.FollowUp} < 0$ -> t-Test für unabhängige Stichproben

Fazit: Varianzanalytische Modelle

- Varianzanalytische Modelle sind statistische Modelle zur Beschreibung einer stetigen AV in mehreren Populationen, die sich durch die diskreten Ausprägungen einer oder mehrerer UVs zusammensetzen.
- Im Rahmen der varianzanalytischen Modelle können eine Vielzahl von inhaltlich relevanten psychologischen Fragestellungen untersucht werden.
- Die in Statistik I besprochenen statistischen Verfahren im Ein- oder Zwei-Stichprobenfall können als wichtige Spezialfälle von statistischen Verfahren im Rahmen einfacher varianzanalytischer Modelle aufgefasst werden.
- In den restlichen Vorlesungen von Statistik II werden wir uns mit regressionsanalytischen Modellen beschäftigen. Diese beinhalten die varianzanalytischen Modelle als Spezialfall. Darüber hinaus können wir damit untersuchen:
  - Stetige AV, eine stetige UV: Einfache lineare Regression
  - Stetige AV, mehrere stetige und/oder diskrete UVs: Multiple lineare Regression
  - Binäre AV, mehrere stetige und/oder diskrete UVs: Logistische Regression
  - …