# 4. Vorlesung Statistik II
# Einführung in die Varianzanalyse

# Wdhl. Inferenzstatistik



- Psychologische Fragestellungen beziehen sich fast immer auf bestimmte Populationen, die zu groß sind, um sie komplett erfassen und mit Methoden der Deskriptivstatistik auswerten zu können.
- Wir haben uns in Statistik I (Vorlesung 6) für einfache Forschungsfragen genau angeschaut, wie wir auf Basis von Zufallsstichproben trotzdem Aussagen über deskriptivstatistische Maßzahlen in der interessierenden Population treffen können:
    - Zunächst wird der Zufallsprozess bei der Stichprobenziehung mit Methoden der Wahrscheinlichkeitstheorie beschrieben.
    - Basierend darauf ermöglichen Methoden der Inferenzstatistik Rückschlüsse auf die Population.
- Die Aufgabe der angewandten Statistik ist es also (vereinfacht gesagt), statistische Methoden zur Verfügung zu stellen, die es erlauben, für die Psychologie interessante inhaltliche Forschungsfragen in geeignete statistische Fragestellungen zu übersetzen.

# Plan für das restliche Semester



- Für das komplette restliche Semester werden wir uns mit statistischen Modellen beschäftigen, die es uns ermöglichen, noch mehr inhaltliche Fragestellungen untersuchen zu können als bisher.
- Die beiden wichtigsten Modellklassen für die Psychologie sind:
  - Varianzanalytische Modelle
  - Regressionsanalytische Modelle

- Wir werden uns zunächst mit varianzanalytischen Modellen beschäftigen, da diese einen einfachen Einstieg in allgemeinere statistische Modelle ermöglichen und in der Psychologie traditionell (immer noch) sehr häufig angewendet werden.
- In der zweiten Hälfte des Semesters werden wir uns dann mit einigen regressions-analytischen Modellen beschäftigen. Die Gruppe aller regressionsanalytischen Modelle ist sehr, sehr groß und enthält alle varianzanalytischen Modelle als Spezialfall. Fast alle inhaltlichen Fragestellungen in der Psychologie lassen sich mit regressionsanalytischen Verfahren im weiteren Sinne untersuchen.

# Vorbemerkungen Varianzanalyse



- Bislang: Verwendung von **Konfidenzintervallen für Parameterdifferenzen** und ***t*-Tests**, wenn...
    - ...sich die Untersuchungsfrage auf **zwei Populationen** bezieht.
    - ...ein **stetiges Merkmal** untersucht wird.
    - ...die interessierenden Populationsgrößen die **Mittelwerte** sind.

- Jetzt: Statistische Methoden für den Fall, dass...
    - ...sich die Untersuchungsfrage auf **mindestens zwei Populationen** bezieht**.**
    - ...ein **stetiges Merkmal** untersucht wird.
    - ...die interessierenden Populationsgrößen die **Mittelwerte** sind.

➢ **varianzanalytische Verfahren** bzw. **Varianzanalyse (ANOVA)**

# Vorbemerkungen Varianzanalyse



- Die Varianzanalyse wird/wurde in der Psychologie traditionell sehr häufig angewendet.
- Mittlerweile gibt es für fast alle Fragestellungen, in denen die Varianzanalyse zum Einsatz kommen könnte, geeignetere Modelle und Methoden.
- Sie ist allerdings besonders gut als Einführung in statistische Modellierung geeignet, da sie leicht verständlich ist.
- Vor allem können wir durch die Varianzanalyse folgende für die statistische Modellierung generell wichtige Themen kennenlernen, die dann im Rahmen der Regressionsanalyse vertieft werden können:
  - Einfache Modellgleichungen
  - Omnibustests
  - Teststatistiken mit anderen Verteilungen als die t-Verteilung
  - Interaktionseffekte zwischen Variablen
  - Modellierung komplexerer Fragestellungen

# Unabhängige und abhängige Variablen bei Varianzanalysen



- Das interessierende (stetige) Merkmal wird im Rahmen der Varianzanalyse häufig als **abhängige Variable** (**AV**) bezeichnet.
- Die **diskrete Variable**, welche die Populationen definiert und deren Einwirkung auf die abhängige Variable untersucht werden soll, wird als **Faktor** oder **unabhängige Variable** (**UV**) bezeichnet.
- Unter **Stufen** eines Faktors werden die **Ausprägungen des diskreten Merkmals** verstanden.
- Beispiel: Das Merkmal Familienstand kann z.B. vierstufig erhoben werden: verheiratet vs. ledig vs. verwitwet vs. geschieden

# Faktorstufen



- Untersucht man **einen** Faktor, so spricht man von einer **einfaktoriellen** Varianzanalyse. Untersucht man **zwei** Faktoren, dann liegt eine **zweifaktorielle** Varianzanalyse vor usw.
- Jeder Faktor kann unterschiedlich viele Ausprägungen (= Stufen) aufweisen. Man spricht z.B. von einer **einfaktoriellen dreistufigen Varianzanalyse**, wenn der Einfluss eines Faktors mit drei Stufen (z.B. niedrig, mittel, hoch) auf die AV untersucht wird.
- Wenn mehr als ein Faktor untersucht wird, muss in der Beschreibung des Untersuchungsdesigns **spezifiziert werden, wie viele Stufen jeder Faktor** aufweist:
  - z.B. ein **2 x 3 Design** (zwei Faktoren, wobei der erste Faktor zweifach, der zweite dreifach gestuft ist)
  - z.B. ein **3 x 2 x 3 Design** (drei Faktoren, wobei der erste Faktor dreifach, der zweite zweifach, der dritte wiederum dreifach gestuft ist)
- Die Anzahl der Populationen ergibt sich aus dem Produkt der Anzahl der Stufen aller Faktoren. Z.B.: Ein 2 x 3 Design entspricht sechs Populationen

# Weitere Begriffe, die bei Varianzanalyen häufig verwendet werden



- Man spricht im Allgemeinen von **Messwiederholungsdesigns**, wenn das interessierende Merkmal zu mehr als an einem Messzeitpunkt erhoben wird (wird in Statistik II nicht besprochen).

- Man spricht von einem **balancierten** Design, wenn der Umfang der Stichproben aus allen Populationen gleich groß ist, von einem **unbalancierten** Design, wenn mindestens zwei Stichproben ungleich groß sind.

# Beispiel

- Uns interessiert das (stetige) Merkmal Depression. Wir vermuten, dass Personen, die unterschiedlichen Alterskategorien angehören, sich in den durchschnittlichen Depressionswerten voneinander unterscheiden. Es sollen drei Alterskategorien betrachtet werden:
  - Altersgruppe 1: junge Erwachsene
  - Altersgruppe 2: Erwachsene
  - Altersgruppe 3: ältere Erwachsene
- Aus allen drei Populationen soll jeweils eine einfache Zufallsstichprobe mit 20 Personen gezogen werden.
- Fragen:
  - AV?
  - UV / Faktor?
  - Faktorstufen?
  - Design?

# Zwischengliederung



- Bislang:
  - Grundbegriffe
- Jetzt:
  - Statistisches Modell für einen Faktor mit *m* Stufen

# Statistisches Modell



- In **statistischen Modellen** werden die Zufallsvariablen, deren Realisationen für die AV-Werte der Personen in der Stichprobe stehen, durch statistische Modellgleichungen in mehrere Komponenten zerlegt.

- Im einfachsten Fall geht man von zwei Komponenten aus: Die eine Komponente enthält den systematischen Anteil, der die **unabhängige Variable** repräsentiert, die andere Komponente enthält den unsystematischen Anteil, der alle weiteren **zufälligen Einflüsse** repräsentiert.

- Bemerkung: In statistischen Modellen werden die Zufallsvariablen, deren Realisationen für die Merkmalsausprägungen der Personen in der Stichprobe auf der AV stehen, üblicherweise mit $Y$ statt mit $X$ bezeichnet.

# Statistisches Modell für einen Faktor mit m Stufen



- Wir werden in den folgenden Vorlesungen mehrere varianzanalytische statistische Modelle kennenlernen.
- Diese statistischen Modelle sind die allgemeine Grundlage für verschiedene Methoden der Parameterschätzung und verschiedene Hypothesentests. Alle statistischen Verfahren, denen ein solches **varianzanalytisches statistisches Modell** zugrunde liegt, werden **varianzanalytische Verfahren** oder **Varianzanalysen** genannt.
- Zunächst wird das statistische Modell für einen Faktor mit $m$ Stufen eingeführt.
- Das statistische Modell für einen Faktor mit $m$ Stufen wird in zwei Varianten vorgestellt:
    - Die erste Variante ist intuitiver.
    - Die zweite Variante erlaubt eine einfachere Verallgemeinerung auf mehrere Faktoren.

# Statistisches Modell für einen Faktor mit m Stufen: Erste Darstellungsform



- Im Falle **eines *m*-fach gestuften Faktors** lautet die statistische Modellgleichung:

$$Y_{ij} = \mu_j + \varepsilon_{ij}, \qquad \text{mit } j = 1, \dots, m \text{ und } i = 1, \dots, n_j$$

wobei $n_j$ der Umfang der *j*-ten Zufallsstichprobe ist

- $Y_{ij}$ ist eine **Zufallsvariable**, deren Realisation der Wert auf der AV einer Person *i* ist, die aus der *j*-ten von *m* Populationen zufällig gezogen wird.
- $\mu_j$ ist der Erwartungswert der AV in der Stichprobe $j$ und gleichzeitig der Mittelwert der AV in der Population *j* . $\mu_j$ ist für jede Population ein **Parameter** und somit eine **Konstante**.
- $\varepsilon_{ij}$ ist eine **Zufallsvariable** und repräsentiert den so genannten **Fehler**, also alle nicht berücksichtigten Einflüsse auf die AV bezogen auf die Person *i*, die aus der *j*-ten Population zufällig gezogen wurde.

# Beispiel erste Darstellungsform



- Beispiel von vorher: AV Depressionsschwere, $m = 3$ Populationen (1 = junge Erwachsene, 2 = Erwachsene, 3 = ältere Erwachsene), $m = 3$ einfache Zufallsstichproben mit jeweils $n_j = 20$ Personen (d.h. insgesamt 60 Personen).

$$Y_{ij} = \mu_j + \varepsilon_{ij}, \qquad \text{mit } j = 1,2,3, \qquad \text{und } i = 1,2,\ldots,20$$

- $\mu_j$ ist die mittlere Depressionsschwere in Population $j$. Das heißt:
  - $\mu_1$ ist die mittlere Depressionsschwere in der Population der jungen Erwachsenen
  - $\mu_2$ ist die mittlere Depressionsschwere in der Population der Erwachsenen
  - $\mu_3$ ist die mittlere Depressionsschwere in der Population der älteren Erwachsenen
- Die Realisation $y_{ij}$ von $Y_{ij}$ ist die Depressionsschwere der $i$-ten Person der Stichprobe aus der $j$-ten Population. z.B.: Die Realisation $y_{21}$ von $Y_{21}$ ist die Depressionsschwere der zweiten Person in der Stichprobe der jungen Erwachsenen.
- $\varepsilon_{ij}$ repräsentiert die zufällige Abweichung der Depressionsschwere von Person *i* aus der *j*-ten Population vom Mittelwert ihrer Population. z.B.: Die Realisation von $\varepsilon_{21}$ ist die Abweichung der Depressionsschwere $y_{21}$ der zweiten Person in der Stichprobe der jungen Erwachsenen von der mittleren Depressionsschwere $\mu_1$ in der Population der jungen Erwachsenen.

# Statistisches Modell für einen Faktor mit m Stufen: Zweite Darstellungsform



- Wie bereits erwähnt, lässt sich die Zusammensetzung der Zufallsvariable $Y_{ij}$ anhand einer zweiten äquivalenten Modellgleichung darstellen.
- Die Modellgleichung lautet in diesem Fall für einen *m*-fach gestuften kategorialen Faktor:

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}, \qquad \text{wobei } \alpha_j = \mu_j - \mu \text{ mit } j = 1, \ldots, m \text{ und } \mu = \frac{1}{m}\sum_{j=1}^{m} \mu_j$$

- $\mu$ ist der **Erwartungswert** bzw. **Mittelwert** der AV bezogen auf die **Gesamtpopulation**, also der Population, die die Gesamtheit aller Teilpopulationen enthält. $\mu$ ist ein **Parameter** und somit eine **Konstante**.
- $\alpha_j$ ist die **Abweichung** des **Erwartungswerts der Population *j*** vom **Erwartungswert der Gesamtpopulation**. $\alpha_j$ ist für jede Population ein **Parameter** und somit eine **Konstante**.
- $Y_{ij}$ und $\varepsilon_{ij}$ sind Zufallsvariablen und haben die gleiche Interpretation wie in der ersten Darstellungsform.

# Beispiel zweite Darstellungsform



- Beispiel von vorher: AV Depressionsschwere, $m = 3$ Populationen (1 = junge Erwachsene, 2 = Erwachsene, 3 = ältere Erwachsene), $m = 3$ einfache Zufallsstichproben mit jeweils $n_j = 20$ Personen (d.h. insgesamt 60 Personen).

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij},\ \text{mit}\ j = 1,2,3$$

- $\mu$ ist die mittlere Depressionsschwere in der Gesamtpopulation der Erwachsenen.
- $\alpha_j$ ist die Abweichung der mittleren Depressionsschwere in Population $j$ von der mittleren Depressionsschwere in der Gesamtpopulation der Erwachsenen.
  Z.B.: $\alpha_1$ ist die Abweichung der mittleren Depressionsschwere in der Population der jungen Erwachsenen von der mittleren Depressionsschwere in der Gesamtpopulation der Erwachsenen.
- $Y_{ij}$ und $\varepsilon_{ij}$ haben die gleiche Interpretation wie in der ersten Darstellungsform.

- **In welchem Fall ist $\alpha_j$ positiv und in welchem Fall negativ?**

# Statistisches Modell für einen Faktor mit m Stufen: 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:

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

  - in der zweiten Darstellung zu:

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

- (beides ist aufwendig zu beweisen)
- Wir haben also in beiden Fällen einen zusätzlichen Parameter $\sigma^2$, der für die Varianz der AV innerhalb der Populationen steht.

# Statistisches Modell für einen Faktor mit m Stufen: Parameter



- Das statistische Modell der Varianzanalyse enthält somit in beiden Darstellungen mehrere Parameter.
  - Erste Darstellung:
    - $\mu_j$ für alle Populationen $j$
    - $\sigma^2$
  - Zweite Darstellung:
    - $\mu$
    - $\alpha_j$ für alle Populationen $j$
    - $\sigma^2$
- 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.

# Zwischengliederung



- Bislang:
  - Grundbegriffe
  - Statistisches Modell für einen Faktor mit *m* Stufen
- Jetzt:
  - Schätzfunktionen für die Parameter des Modells

# Schätzfunktionen für ein einfaktorielles statistisches Modell (I)



- Bei einem einfaktoriellen statistischen Modell in der ersten Darstellung sind die Parameter $\mu_j$ und $\sigma^2$unbekannt.
- Es lässt sich zeigen, dass **erwartungstreue**, **konsistente** und **effiziente Schätzfunktionen** für die unbekannten Größen $\mu_j$ und $\sigma^2$ existieren.
- Für die unbekannten Erwartungswerte $\mu_j$ verwenden wir jeweils die Schätzfunktion

$$\hat{\mu}_j = \bar{Y}_j = \frac{1}{n_j}\sum_{i=1}^{n} Y_{ij}$$

- Die Realisation dieser Schätzfunktion, also der Schätzwert für $\mu_j$, ist einfach der Mittelwert der AV in der aus Population $j$ gezogenen Stichprobe.
- Die Schätzfunktionen für die Parameter $\mu$ und $\alpha_j$ der zweiten Darstellung können aus den Schätzfunktionen für $\mu_j$ abgeleitet werden.

# Schätzfunktionen für ein einfaktorielles statistisches Modell (II)



- Als Schätzfunktion für $\sigma^2$ verwenden wir in beiden Darstellungen die gepoolte Schätzfunktion $S_{pool}^2$.
- Wir erinnern uns: Im Falle von unabhängigen Stichproben wurde aus dem (gewichteten) Mittelwert der beiden Schätzfunktionen $S_1^2$ und $S_2^2$ die gepoolte Schätzfunktion $\hat{\sigma}^2 = S_{pool}^2$ gebildet.
- Dieses Prinzip wird für die Varianzanalyse übernommen.
- Im balancierten Design mit $n_1 = n_2 = \cdots = n_m = n$ gilt:

$$\hat{\sigma}^2 = S_{pool}^2 = \frac{1}{m(n-1)} \left( \sum_{j=1}^{m} \sum_{i=1}^{n} (Y_{ij} - \bar{Y}_j)^2 \right)$$

- Es gilt:

$$E(S_{pool}^2) = \sigma^2$$

# Zwischengliederung



- Bislang:
  - Grundbegriffe
  - Statistisches Modell für einen Faktor mit *m* Stufen
  - Schätzfunktionen für die Parameter des Modells
- Jetzt:
  - Ein erster Hypothesentest

# Omnibustest



- Wir lernen nun einen ersten Hypothesentest im Rahmen der einfaktoriellen Varianzanalyse kennen, den sogenannten **Omnibustest**.

- Die Hypothesen, die mit diesem Hypothesentest überprüft werden, sind relativ uninteressant.

- Der Test wird jedoch (leider) sehr häufig verwendet.

# Die inhaltlichen Hypothesen bei einem Omnibustest



- Beim Omnibustest wird unter der **Nullhypothese** die Aussage getroffen, dass sich die Mittelwerte in den Populationen **nicht unterscheiden**.
- Unter der **Alternativhypothese** wird die Aussage getroffen, dass sich **mindestens einer** der Mittelwerte in den Populationen von einem anderen **unterscheidet.**
- Falls wir uns im Rahmen des Omnibustests für die Alternativhypothese entscheiden, können wir also nicht sagen, **welche** der Mittelwerte sich unterscheiden. Dies macht den Omnibustest in den meisten Fällen uninteressant.

# Statistische Hypothesen für den Omnibustest I



- In Abhängigkeit davon, welche Variante des statistischen Modells verwendet wird, unterscheiden sich die **statistischen Hypothesen des Omnibustests**. Beide Formulierungen sind jedoch inhaltlich äquivalent.
- 1. Art der Formulierung der statistischen Hypothesen für den Omnibustest:

$H_0: \mu_1 = \mu_2 = \ldots = \mu_k \ldots = \mu_m$

$H_1: \mu_j \neq \mu_k \text{ für mindestens ein Paar } j, k \text{ mit } j \neq k$

- 2. Art der Formulierung der statistischen Hypothesen für den Omnibustest:

$H_0: \alpha_j = 0$ für alle $j$

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

# Statistische Hypothesen für den Omnibustest II



- Die statistischen Hypothesen des Omnibustests lassen sich als zusammengesetzte Hypothesen darstellen. Für ein Beispiel mit m = 3 sieht dies in der ersten Darstellungsform folgendermaßen aus.

- Zusammengesetzte Hypothesen:

$$H_0: \mu_1 = \mu_2 \; und \; \mu_1 = \mu_3 \; und \; \mu_2 = \mu_3$$

$$H_1: \mu_1 \neq \mu_2 \; oder \; \mu_1 \neq \mu_3 \; oder \; \mu_2 \neq \mu_3$$

- Einzelne Hypothesen:

$$H_{01}: \mu_1 = \mu_2, \, H_{02}: \mu_1 = \mu_3, \, (H_{03}: \mu_2 = \mu_3)$$

$$H_{11}: \mu_1 \neq \mu_2, \, H_{12}: \mu_1 \neq \mu_3, \, (H_{13}: \mu_2 \neq \mu_3)$$

Bemerkung: Aufgrund der logischen Abhängigkeit der Einzelhypothesen kann eine der Einzelhypothesen auch jeweils weggelassen werden.

- Die Alternativhypothese des Omnibustests entspricht also einer **zusammengesetzten Alternativhypothese mit „oder“**. Der Omnibustest prüft diese zusammengesetzte Hypothese mit **einem einzigen** Hypothesentest (siehe Vorlesung 1).

- Damit kommt es zu keiner $\boldsymbol{\alpha}$-Fehler-Kumulierung, das Signifikanzniveau des Omnibustests muss nicht korrigiert werden und die Power ist höher im Vergleich zur Testung mit einzelnen Hypothesentests und Korrektur des Signifikanzniveaus.

# Zwischengliederung

- Bislang:
  - Grundbegriffe
  - Statistisches Modell für einen Faktor mit *m* Stufen
  - Schätzfunktionen für die Parameter des Modells
  - Hypothesen für den Omnibus-Test

- Jetzt:
  - Teststatistik des Omnibustests
  - Wahrscheinlichkeitsverteilung der Teststatistik des Omnibustests unter der $H_0$
  - Durchführung des Omnibustests

# Quadratsummen



- Die Teststatistik des Omnibustests setzt sich aus **Quadratsummen** zusammen.

- Unter einer Quadratsumme versteht man im Allgemeinen eine **Summe von quadrierten Abweichungen**.

- Da der Omnibustest für gewöhnlich kein für uns interessanter Test ist, verzichten wir auf die genaue Herleitung der Quadratsummen und somit der Teststatistik (siehe Exkurs im Anhang). Wir fokussieren uns auf die Aussagen der Teststatistik und Tests.

# Teststatistik des Omnibustests



- Vereinfacht gesagt, wird in der Teststatistik folgendes Verhältnis betrachtet:

$$T = \frac{Varianz\ zwischen\ den\ Gruppen}{Varianz\ innerhalb\ der\ Gruppen}$$

- Sie folgt **unter der Nullhypothese** einer sogenannten $F$-Verteilung mit Parametern

  $\nu_1 = df_{zw}$ und $\nu_2 = df_{inn}$:

  $T \sim F(\nu_1, \nu_2)$ bzw. $T \sim F(df_{zw}, df_{inn})$

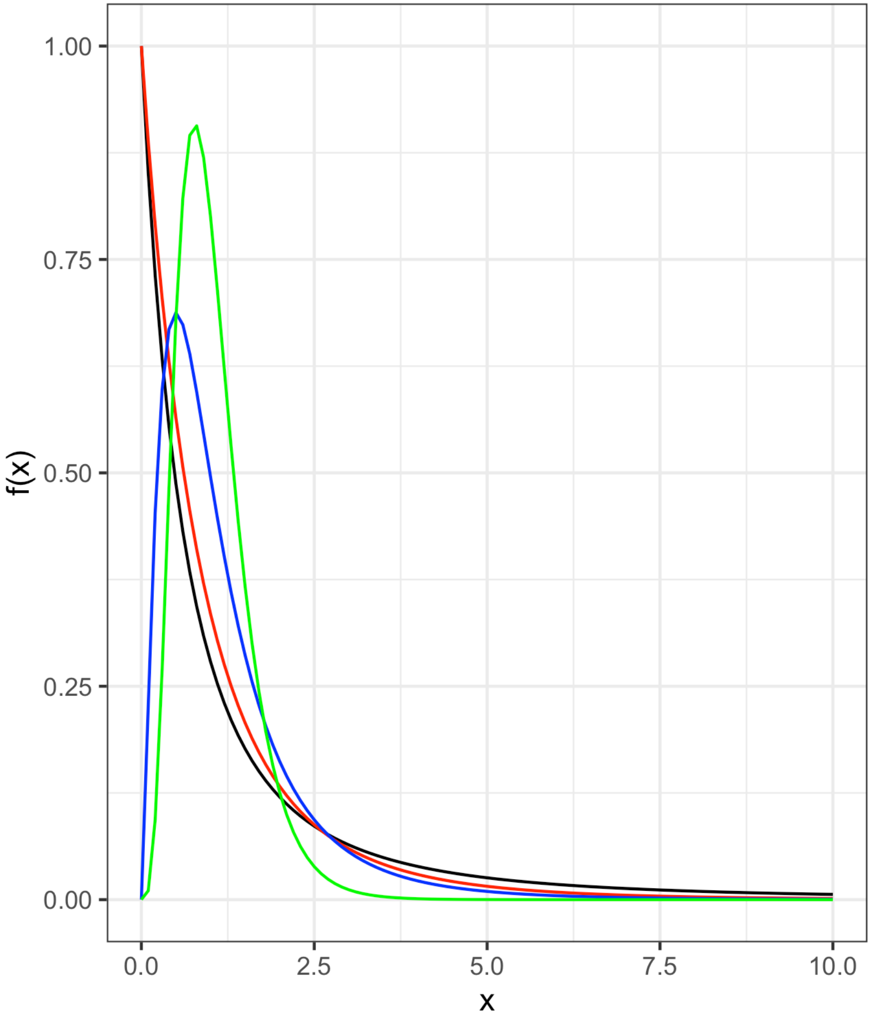
- $df_{zw} = m - 1$ sind die Freiheitsgrade zwischen den Gruppen.
- $df_{inn} = m(n - 1)$ sind die Freiheitsgrade innerhalb der Gruppen.

- In der Teststatistik wird überprüft, ob die **mittlere Abweichung der Gruppenmittelwerte vom Gesamtmittelwert** größer ist als die **mittlere Abweichung der einzelnen Werte innerhalb einer Gruppe von ihrem Gruppenmittelwert**.

# Eigenschaften der $\boldsymbol{F}$-Verteilung



Dichtefunktionen für unterschiedliche Parameterwerte




- Die $F$-Verteilung ist eine stetige Wahrscheinlichkeitsverteilung.
- Ihre genaue Form hängt von **zwei Parametern** $\nu_1$ und $\nu_2$ ab.
- Der Träger einer $F$-verteilten Zufallsvariable ist das Intervall **von null bis plus unendlich**.
- Der **Erwartungswert** einer $F$-verteilten Zufallsvariable $X$ ist für $\nu_2 > 2$

$$E(X) = \frac{\nu_2}{\nu_2 - 2}$$

- Es existieren unendlich viele *F*-Verteilungen, je nachdem welche Werte die Parameter $\nu_1$ und $\nu_2$ annehmen.

# Verteilungsfunktion der F-Verteilung



- Die Wahrscheinlichkeit dafür, dass sich eine $F$-verteilte Zufallsvariable $X$ in einem Wert kleiner oder gleich einem Wert $x$ realisiert, kann mithilfe der Verteilungsfunktion $F$ berechnet werden.
- Mithilfe der Verteilungsfunktion können auch beliebige Quantile bestimmt werden.
- Beispiele in R für eine $F$ -Verteilung mit $\nu_1 = 2$ und $\nu_2 = 10$:
  - Bestimmung von $P(X \leq 1) = F(1)$:

    ```
    > pf(1, df1=2, df2=10)
    [1] 0.5981224
    ```

  - Bestimmung des 0.995-Quantils, also des Wertes $x$, für den $P(X \leq x) = 0.995$ gilt:

    ```
    > qf(0.995, df1=2, df2=10)
    [1] 9.426999
    ```

# Kritischer Bereich des Omnibustests

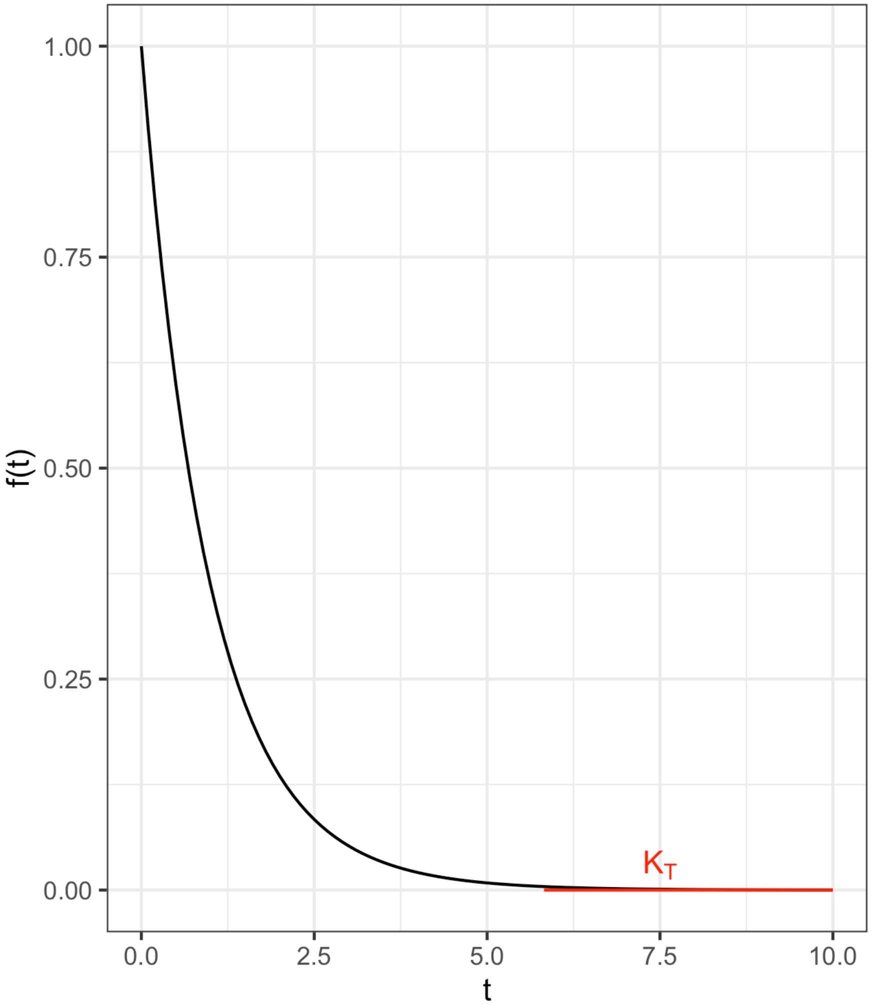


- Unter Geltung der Nullhypothese wird erwartet, dass sich die Teststatistik $T$ in einem Bereich, der **nahe dem Wert eins liegt**, realisiert, da unter der Nullhypothese gilt, dass die Varianz zwischen den Gruppen gleich der Varianz innerhalb der Gruppen ist.
- Falls die Alternativhypothese gilt, würde man erwarten, dass die Realisation der Teststatistik $\mathrm{T}$ größer als eins ist, da in diesem Fall die Varianz zwischen den Gruppen größer als die Varianz innerhalb der Gruppen ist.
- Falls die Alternativhypothese gilt, kann man also sagen, dass die Variable, die die Gruppen definiert (also die unabhängige Variable), die Varianz der abhängigen Variable (zu Teilen) erklären kann.
- **Je stärker** die Realisation $\mathrm{t}$ der Teststatistik $\mathrm{T}$ **in positiver Richtung vom Wert 1 abweicht**, **desto stärker** sprechen die Daten für die **Alternativhypothese**.
- Der **kritische Bereich** des Omnibustests liegt also **auf der rechten Seite**.

# Graphische Veranschaulichung: $\boldsymbol{F}$-Verteilung, kritischer Bereich und Testentscheidung



Dichtefunktion der Teststatistik
(unter Geltung der $H_0$)




- Beispiel: $m = 3$ Gruppen, in jeder Gruppe $n = 20$ Beobachtungen.
- Das heißt:
  $df_{zw} = m - 1 = 3 - 1 = 2$ und
  $df_{inn} = m(n-1) = 3(20-1) = 57$
- Unter der Nullhypothese gilt also

$$T \sim F(2, 57)$$

- Kritischer Bereich für $\alpha = 0.005$:

```
> qf(0.995, df1=2, df2=57)
[1] 5.822804
```

- Das heißt: $K_T = [5.82,\ \infty[$
  Wenn $t \in K_T$, dann $H_1$ annehmen
  Wenn $t \notin K_T$, dann $H_0$ annehmen

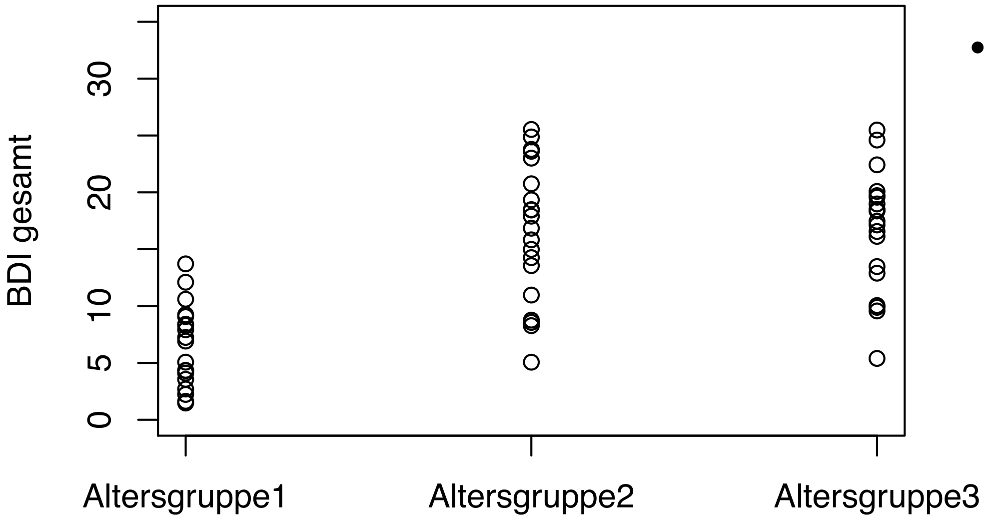
# Voraussetzungen des Omnibustests

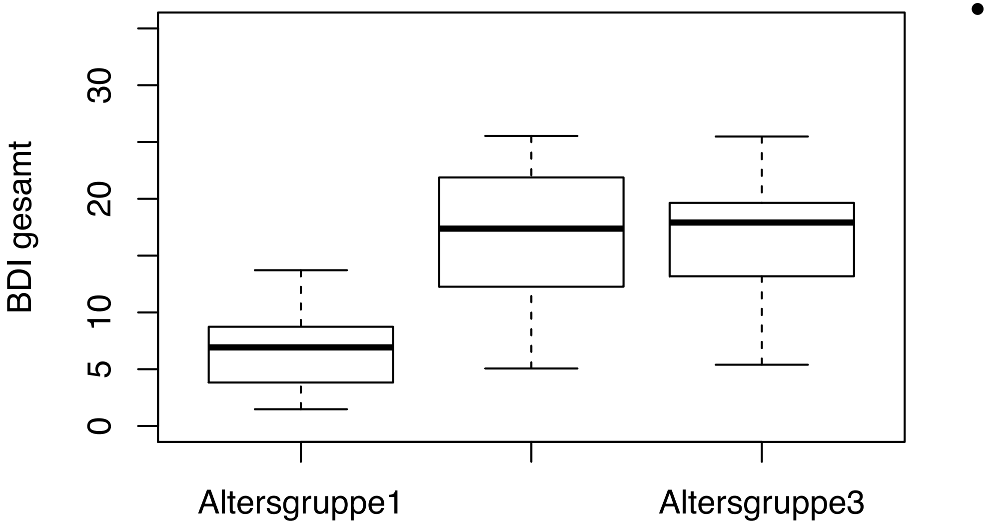


- Der Omnibustest ist wie alle Verfahren im Rahmen des varianzanalytischen Modells (zunächst) daran gebunden, dass die Annahmen des Modells richtig sind.
- Die Überprüfung der Annahmen werden wir noch besprechen.
- Im folgenden Beispiel gehen wird davon aus, dass die Voraussetzungen erfüllt sind.

# Graphische Darstellung der Daten






- Beispiel von vorher: AV Depressionsschwere, $m = 3$ Populationen (1 = junge Erwachsene, 2 = Erwachsene, 3 = ältere Erwachsene), $m = 3$ einfache Zufallsstichproben mit jeweils $n_j = n = 20$ Personen (d.h. insgesamt 60 Personen).

- Was fällt auf?

# Deskriptive Statistik

```
> length(Altersgruppe1)
[1] 20
> length(Altersgruppe2)
[1] 20
> length(Altersgruppe3)
[1] 20
```

```
> mean(Altersgruppe1)
[1] 6.01837
> mean(Altersgruppe2)
[1] 16.64277
> mean(Altersgruppe3)
[1] 16.73504
```

```
> sd(Altersgruppe1)
[1] 4.029811
> sd(Altersgruppe2)
[1] 6.076885
> sd(Altersgruppe3)
[1] 5.184188
```

- Was fällt auf?

# Omnibus-Test in R



*p*-Wert

```
> fit <- aov(BDI ~ Altersgruppe)
> summary(fit)
             Df Sum Sq Mean Sq F value   Pr(>F)
Altersgruppe  2   1518   759.1   28.45 2.7e-09 ***
Residuals    57   1521    26.7
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

$t$

$df_{zw}$ $qs_{zw}$ $qs_{inn}$ $mqs_{inn}$

$df_{inn}$ $mqs_{zw}$

**Exkurs**

$$t = \frac{qs_{zw}/df_{zw}}{qs_{inn}/df_{inn}} = \frac{mqs_{zw}}{mqs_{inn}} = \frac{1518/2}{1521/57} = \frac{759.1}{26.7} = 28.45$$

- Wie lauten die inhaltlichen Hypothesen für den Omnibustest?
- Wie wird der *p*-Wert berechnet?
- Testentscheidung und Interpretation?

# Zwischengliederung



- Bislang:
  - Durchführung des Omnibustests
- Jetzt:
  - Effektgröße im einfaktoriellen varianzanalytischen Modell

# Effektgröße im einfaktoriellen varianzanalytischen Modell



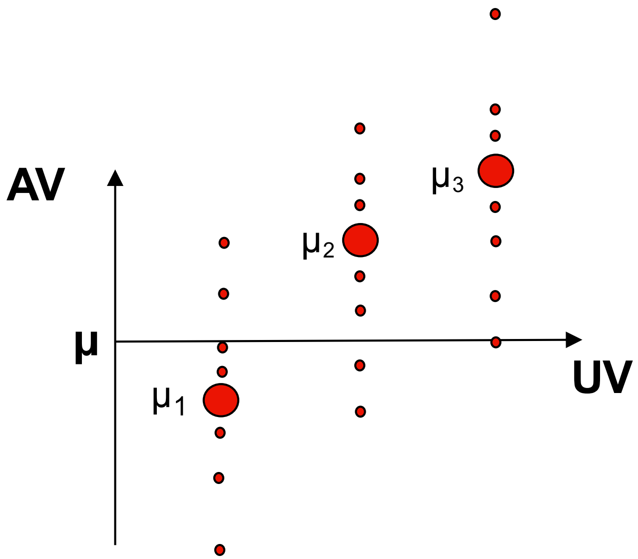
- Eine im Rahmen des einfaktoriellen varianzanalytischen Modells häufig verwendete Effektgröße ist $\eta^2$.

- $\eta^2$ ist definiert als das Verhältnis der Varianz $\sigma^2_{zw}$ und der Gesamtvarianz $\sigma^2_{ges}$:

$$\eta^2 = \frac{\sigma^2_{zw}}{\sigma^2_{ges}}$$

# Inhaltliche Interpretation von $\sigma_{zw}^2$





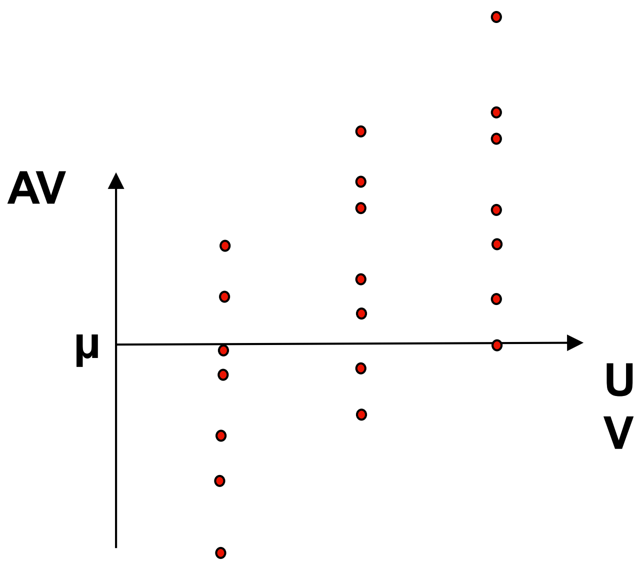

- $\sigma_{zw}^2$ ist die Varianz der Erwartungswerte der *m* Populationen um den Gesamt-erwartungswert $\mu$:

$$\sigma_{zw}^2 = \frac{1}{m}\sum_{j=1}^{m}(\mu_j - \mu)^2$$

- Je größer die Abweichungen der Erwartungswerte $\mu_j$ mit $j = 1, \ldots, m$ vom Gesamterwartungswert $\mu$, desto größer ist $\sigma_{zw}^2$.

# $\sigma_{ges}^2$






- $\sigma_{ges}^2$ ist die Gesamtvarianz der AV-Werte der einzelnen Personen in allen *m* Populationen um den Gesamterwartungswert $\mu$.

- Man kann zeigen, dass $\sigma_{ges}^2 = \sigma_{zw}^2 + \sigma^2$ ist.

- $\sigma_{ges}^2$ hängt zum einen davon ab, wie stark sich die Erwartungswerte der *m* Populationen unterscheiden (also von $\sigma_{zw}^2$) und zum anderen davon, wie stark die Streuung innerhalb der Populationen ist (also von $\sigma^2$)

# Inhaltliche Interpretation von $\eta^2$



- Wegen $\sigma_{ges}^2 = \sigma_{zw}^2 + \sigma^2$ ist $\sigma_{zw}^2 < \sigma_{ges}^2$ und somit kann $\eta^2 = \frac{\sigma_{zw}^2}{\sigma_{ges}^2}$ Werte zwischen 0 und 1 annehmen.

- Inhaltliche Interpretation von $\eta^2$:
  $\eta^2 = \frac{\sigma_{zw}^2}{\sigma_{ges}^2}$ ist der Anteil an der Gesamtvarianz der AV in der Population, der durch die UV (bzw. durch den Faktor) erklärt werden kann.

- $\eta^2$ ist eine Kombination von Parametern des einfaktoriellen varianzanalytischen Modells und somit selbst ein Parameter. Wir können ihn also mithilfe der Daten schätzen oder Hypothesen über ihn testen.

# Schätzfunktion und Schätzwert für $\eta^2$



- Als Schätzfunktionen für $\sigma^2_{zw}$ und $\sigma^2_{ges}$ können wir jeweils die entsprechenden Stichprobenkennwerte verwenden (vereinfachte Formeln mit balanciertem Design):

$$\hat{\sigma}^2_{zw} = \frac{1}{m}\sum_{j=1}^{m}(\bar{Y}_j - \bar{Y})^2 = \frac{1}{m \cdot n}\sum_{j=1}^{m} n \cdot \left(\bar{Y}_j - \bar{Y}\right)^2$$

$$\hat{\sigma}^2_{ges} = \frac{1}{m \cdot n}\sum_{j=1}^{m}\sum_{i=1}^{n}(Y_{ij} - \bar{Y})^2$$

- Als Schätzfunktion $\hat{\eta}^2$ für $\eta^2$ ergibt sich damit:

$$\hat{\eta}^2 = \frac{\hat{\sigma}^2_{zw}}{\hat{\sigma}^2_{ges}}$$

- Die Realisation $\hat{\eta}^2_{Wert}$ der Schätzfunktion $\hat{\eta}^2$ ist der Schätzwert für $\eta^2$

# Berechnung von $\hat{\eta}^2_{Wert}$ in R für unser Beispiel



```
library(DescTools)

> EtaSq(fit)
                eta.sq
Altersgruppe 0.499571
```

```
> summary(fit)
             Df Sum Sq Mean Sq F value  Pr(>F)
Altersgruppe  2   1518   759.1   28.45 2.7e-09 ***
Residuals    57   1521    26.7
```

Interpretation:

Der Schätzwert für den Anteil an der Gesamtvarianz der Depressionsschwere in der Population, der durch das Alter erklärt werden kann, beträgt 50%.

# Zwischengliederung



- Bislang:
  - Effektgröße im einfaktoriellen varianzanalytischen Modell

- Jetzt:
  - Annahmen im einfaktoriellen varianzanalytischen Modell

# Wdh.: Annahmen im einfaktoriellen varianzanalytischen Modell I



- Eine erste Annahme des einfaktoriellen varianzanalytischen Modells wird durch die Modellgleichung selbst getroffen (hier nur Darstellung 1):

$$Y_{ij} = \mu_j + \varepsilon_{ij} \qquad \text{mit } j = 1, \dots, m \text{ und } i = 1, \dots, n_j$$

- In Worten: Wir gehen davon aus, dass der AV-Wert jeder Person in jeder Gruppe dem Gruppenerwartungswert $\mu_j$ plus einer Abweichung von diesem entspricht.

- Diese Annahme trifft immer zu. Wir müssen sie also nicht überprüfen.

# Wdh.: Annahmen im einfaktoriellen varianzanalytischen Modell II



- 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:

$$Y_{ij} \overset{\text{ind}}{\sim} N(\mu_j, \sigma^2) \text{ bzw. } \varepsilon_{ij} \overset{\text{iid}}{\sim} N(0, \sigma^2)$$

  - in der zweiten Darstellung zu:

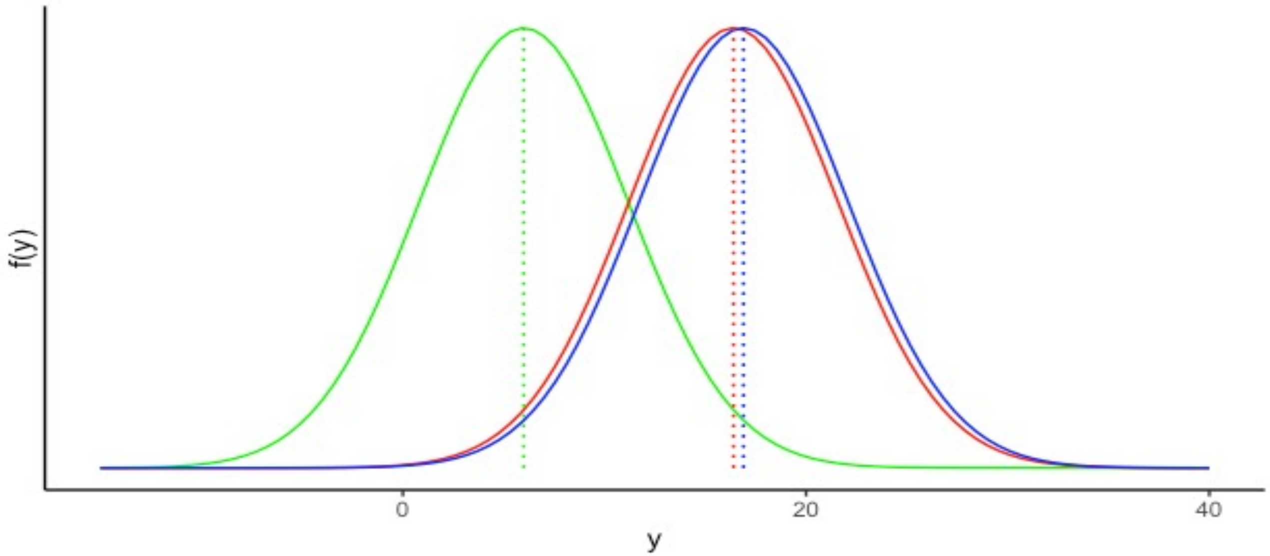
$$Y_{ij} \overset{\text{ind}}{\sim} N(\mu + \alpha_j, \sigma^2) \text{ bzw. } \varepsilon_{ij} \overset{\text{iid}}{\sim} N(0, \sigma^2)$$

- Inhaltlich bedeutet dies, dass die Abweichungen der AV-Werte aller Personen in allen Stichproben von ihrem Gruppenerwartungswert **unabhängig** sind und innerhalb jeder Stichprobe einer **Normalverteilung** mit **gleicher Varianz** $\sigma^2$ folgen.

# Veranschaulichung der Annahmen



- Bsp.: Mittlere Depression bei jungen Erwachsenen, Erwachsenen und alten Erwachsenen
- $\hat{\mu}_{1\_Wert} = \bar{y}_1 = 6.0$, $\hat{\mu}_{2\_Wert} = \bar{y}_2 = 16.6$, $\hat{\mu}_{3\_Wert} = \bar{y}_3 = 16.7$, $\hat{\sigma}^2_{Wert} = s^2_{pool} = 26.7$
- Veranschaulichung der geschätzten Verteilungen der Depression in den 3 Altersgruppen:




- Annahmen des varianzanalytischen Modells: Eine Normalverteilung pro Population
- Die Erwartungswerte der Normalverteilungen dürfen sich unterscheiden.
- Die Varianzen der Normalverteilungen müssen gleich sein.

# Bedeutung der Annahmen



- Die Unabhängigkeit der $Y_{ij}$ können wir durch die Ziehung einer einfachen Zufallsstichprobe sicherstellen.

- Die beiden anderen auf der letzten Folie beschriebenen Annahmen (die Normalverteilungsannahme und die Varianzgleichheitsannahme) können falsch sein.

- Nur wenn alle Annahmen erfüllt sind, weisen die varianzanalytischen Verfahren, die wir zur Parameterschätzung und Hypothesentestung im Rahmen des varianzanalytischen Modells verwenden, die von uns gewünschten Eigenschaften auf, z.B.
  - das von uns gewünschte Konfidenzniveau im Fall von Konfidenzintervallen.
  - das von uns gewünschte Signifikanzniveau im Fall von Hypothesentests.

# Überprüfung der Annahmen I



- Problem: Um sicher sagen zu können, ob die Annahmen erfüllt sind, müssten wir die Populationen kennen.

- Es gibt deskriptive und inferenzstatistische Methoden, um die Annahmen auf der Basis unserer Stichproben zu überprüfen:

  - Normalverteilungsannahme:
    - deskriptiv: Histogramme der AV in allen Stichproben anschauen.
    - inferenzstatistisch: Kolmogorov-Smirnov-Test oder Shapiro-Wilk Test innerhalb jeder der einzelnen Stichproben (Nullhypothese jeweils: Normalverteilungsannahme gilt).

  - Varianzgleichheitsannahme:
    - deskriptiv: Box-Plots oder Varianzen der AV in den Stichproben vergleichen.
    - inferenzstatistisch: Levene-Test (Nullhypothese: Varianzen sind gleich).

# Überprüfung der Annahmen II



- Aber:
  - die deskriptivstatistischen Verfahren sind sehr subjektiv.
  - Die inferenzstatistischen Verfahren haben in kleinen Stichproben eine niedrige Power und erkennen in großen Stichproben auch geringfügige Verletzungen der Annahmen. Außerdem treffen Sie teilweise selbst weitere Annahmen.
  - Wir müssen sowieso davon ausgehen, dass die Annahmen nicht exakt erfüllt sind. Warum sollten z.B. die Varianzen in den Populationen exakt gleich sein?
- Außerdem: Man kann zeigen, dass die meisten im Rahmen dieser VL besprochenen varianzanalytischen Verfahren **in großen Stichproben** (bei balancierten Designs) trotz Verletzung der Annahmen approximativ die von uns gewünschten Eigenschaften aufweisen.
- Fazit: In kleinen Stichproben können wir die Annahmen nicht sinnvoll überprüfen und in großen Stichproben ist es in den meisten Fällen nicht so wichtig, ob sie erfüllt sind.
- **Also wie immer: Große Stichproben erheben!**

# Varianzgleichheit und der Omnibustest



- Für den Omnibustest gilt bei großem $n$ im balancierten Design:
  - Ist nur die Normalverteilungsannahme verletzt, nicht aber die Annahme der Varianzgleichheit, ist das Signifikanzniveau approximativ gleich dem gewählten $\alpha$.
  - Ist die Annahme der Varianzgleichheit verletzt, ist das Signifikanzniveau **nicht unbedingt** approximativ gleich dem gewählten $\alpha$ (unabhängig davon, ob die Normalverteilungsannahme gilt).
- Im Falle einer Verletzung der Varianzgleichheitsannahme kann aber auf eine Variante des Omnibustests zurückgegriffen werden, die weniger anfällig gegenüber ungleichen Varianzen ist. Sie verwendet eine etwas andere Teststatistik und eine korrigierte Formel für $\nu_2$ bei der Berechnung des p-Werts bzw. des kritischen Bereichs. In R:

```
oneway.test(AV ~ UV, daten)
```

- Da wir von (zumindest geringfügig) ungleichen Varianzen in den Populationen ausgehen müssen, spricht nichts dagegen, einfach ohne Überprüfung der Annahmen immer diese Variante zu wählen. Der korrigierte Omnibustest hat dann für großes $n$ trotz Verletzung aller Annahmen ein approximatives Signifikanzniveau gleich dem gewählten $\alpha$.

# Varianzgleichheit und weitere varianzanalytische Verfahren

- Falls einfache Parameterdifferenzen $\mu_j - \mu_k$ von Interesse sind und man davon ausgehen kann, dass sich die Varianzen der AV in den Populationen stark unterscheiden, sollten die aus Statistik I bekannten klassischen Konfidenzintervalle und klassischen t-Tests für unabhängige Stichproben verwendet werden, da diese weniger anfällig für die Verletzung der Varianzhomogenitätsannahme sind.

Erklärung:

- Die Verfahren aus Statistik I verwenden zur Schätzung von $\sigma^2$ nur die Daten aus den Populationen j und k. Sind die Annahmen des varianzanalytischen Modells erfüllt, ist eine Schätzung von $\sigma^2$ im Rahmen des varianzanalytischen Modell genauer, da hier für die Berechnung von $S_{pool}^2$ mehr Daten genutzt werden können.

- Ist die Annahme der Varianzgleichheit nicht erfüllt, kann dies für die Verfahren aus Statistik I zumindest für große Stichproben vernachlässigt werden (Vorlesung 13 in Statistik I). Für die varianzanalytischen Verfahren gilt dies jedoch nicht und stark ungleiche Varianzen können auch bei großen Stichproben problematisch sein.

# Plan nächste Woche



- Bislang:
  - Grundbegriffe
  - Statistisches Modell für einen Faktor mit m Stufen.
  - Schätzfunktionen für die Parameter des Modells
  - Hypothesen für den Omnibustest.
  - Teststatistik des Omnibustests.
  - Wahrscheinlichkeitsverteilung der Teststatistik des Omnibustests unter der H0
  - Durchführung des Omnibustests
  - Effektgröße für den Omnibustest
  - Modellannahmen des Omnibustests
- Nächste Woche:
  - Weitere statistische Verfahren im einfaktoriellen varianzanalytischen Modell
  - Zweifaktorielles varianzanalytisches Modell