

Multivariate Verfahren

Master Psychologie ALU Freiburg, 1. Fachsemester, WS 2016/17; Dozent: Dr. Rainer Leonhart

Disclaimer: Dieses Skript wurde von einer Studentin nach bestem Wissen und Gewissen angelegt. Später bestand sie die Modulprüfung mit der bestmöglichen Note. Der Kausalzusammenhang zwischen diesen beiden Ereignissen ist nicht geklärt. Haftung für die Richtigkeit oder Vollständigkeit des Skriptes kann deshalb nicht übernommen werden.

Gelb markierte Stellen repräsentieren Fragen, die zum Zeitpunkt der letzten Änderung des Skriptes nicht geklärt waren. Für viele davon finden sich entsprechende Fragen und Antworten im q2a Multivariate Verfahren unter dem Usernamen **wraith**.

Blau markierte Stellen waren im Wintersemester 2016/2017 nicht klausurrelevant.

Kursiv geschriebene Stellen beziehen sich zumeist auf mündliche Überlieferungen nach Rainer Leonhart und sind nicht klausurrelevant, helfen aber teilweise beim Verständnis.

Inhalt

Signifikanzprüfung	2
Allgemeines Lineares Modell (ALM/GLM)	5
Datenvorbereitung	7
Fehlerhafte Daten	7
Fehlende Werte	9
Faktorenanalyse	12
Exploratorische Faktorenanalyse	12
Konfirmatorische Faktorenanalyse	18
Diskriminanzanalyse	19
Regressionsanalyse	23
Einfache (Lineare) Regression	23
Multiple (Lineare) Regression	23
Logistische Regression	29
Multiple Varianzanalyse (MANOVA)	33
MANOVA mit Messwiederholung	39
Kovarianzanalyse (MANCOVA)	44
Hierarchische Clusteranalyse	47
Strukturgleichungsmodelle	51
Hierarchische Lineare Modelle (Mehrebenenmodelle)	59

Signifikanzprüfung

- Fast immer wird geprüft, ob sich Elemente des ALM (z.B. Mittelwertdifferenz, Korrelation etc.) statistisch bedeutsam von 0 unterscheiden (= Prüfung der **Nullhypothese** als Grundlage der inferenzstatistischen Hypothesenprüfung)
- Natürlich immer nur dann nötig, wenn nicht alle Elemente der Population in Analyse einbezogen
- **Unterschiede zwischen Verfahren**
 - Anzahl Variablen
 - Anzahl Gruppen
 - Komplexität der statistischen Modelle
 - Skalenniveaus
 - Interaktionseffekte
 - Mehrebenenmodellierung
 - ...
- **Grundgedanke**
 - **Zentraler Grenzwertsatz:** In einer Population mit einer endlichen Varianz und einem Mittelwert nähert sich die Verteilung der Mittelwerte aus gleich großen Stichproben mit N unabhängigen Beobachtungen einer Normalverteilung an, deren Varianz der Varianz der Population geteilt durch N und deren Mittelwert dem Populationsmittelwert entspricht. Die Annäherung ist immer stärker, je größer N ist.
 - Mittelwerte zufällig aus Population gezogener Stichproben streuen um Populationsmittelwert -> theoretische Verteilung der Kennwerte -> inferenzstatistische Aussagen unter Voraussetzung der Gültigkeit der H_0 möglich
 - Man schätzt aus den Stichprobenkennwerte die Populationskennwerte.
 - Sind zwei Stichprobenmittelwerte sehr ähnlich, ist es sehr wahrscheinlich, dass sie aus derselben Population stammen; je unterschiedlicher, desto unwahrscheinlicher. Diese bedingte Wahrscheinlichkeit wird anhand der Wahrscheinlichkeitsverteilung für die Differenz der Stichprobenmittelwerte bei Gültigkeit der H_0 ($p(\text{Differenz} | H_0)$) berechnet.
 - Liegt unter der Voraussetzung, dass in Population keine Mittelwertdifferenz/kein Zusammenhang vorliegt, die Wahrscheinlichkeit für die beobachtete (**empirische**) oder eine noch größere Mittelwertdifferenz unter dem festgelegten Grenzwert (**α -Niveau**), so muss die H_0 verworfen werden (-> **Alternativhypothese H_1** angenommen).
 - Ansonsten wird H_0 angenommen.
- **Fehler** (gegenläufig)
 - **α -Fehler/Fehler 1. Art:** Ablehnung der richtigen Nullhypothese
 - **β -Fehler/Fehler 2. Art:** Beibehaltung der falschen Nullhypothese
- **Wahrscheinlichkeitsfunktionen**
 - **Zufallsvariable X :** Ordnet Ergebnissen eines Zufallsexperiments eine reelle Zahl zu; in diesem Fall seine Wahrscheinlichkeit
 - Zwei Arten
 - **Diskrete Wahrscheinlichkeitsverteilung** (Binomialverteilung, Bernoulliverteilung, Hypergeometrische Verteilung)
 - Zufallsexperiment hat diskrete Variable X -> Ergebnisse abzählbar
 - Beispiel: Würfeln
 - **Stetige Wahrscheinlichkeitsverteilung** (Normalverteilung, Standardnormalverteilung, t-Verteilung, F-Verteilung, χ^2 -Verteilung)
 - Stetige Variable X -> unendlich viele Elementarereignisse können realisiert werden -> stetige Dichteverteilung ohne Lücken
 - Man kann Intervallwahrscheinlichkeiten berechnen

$$p(a < X < b) = \int_a^b f(X) dX \quad \text{mit} \quad \int_{-\infty}^{\infty} f(X) dX = 1$$

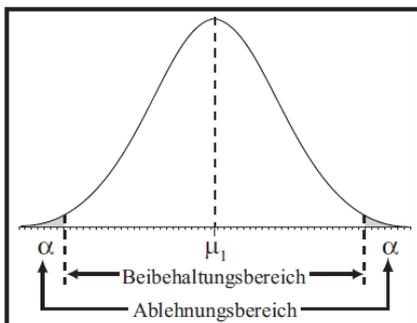
- - f(X): Dichtefunktion
 - a, b: untere bzw. obere Grenze des Intervalls
- **Gauß'sche Normalverteilung:**

$$p(x < a) = \int_{-\infty}^a \frac{1}{\sigma_x \cdot \sqrt{2 \cdot \pi}} \cdot e^{-\frac{(x-\mu)^2}{2 \cdot \sigma_x^2}} dx$$

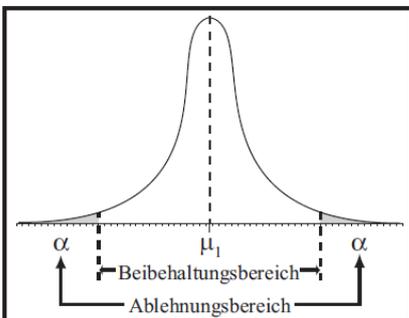
-
- Im Prinzip gibt es davon unendlich viele, die alle durch z-Transformation ineinander überführbar sind.
- In den Lehrbüchern steht nur die **Standardnormalverteilung** mit $\mu=0$ und $\sigma=1$.
 - Basis vieler statistischer Kennwerte
 - Alle anderen statistisch wichtigen Verteilungen lassen sich von ihr ableiten
- Z-Werte (zur Wahrscheinlichkeitsermittlung)
 - 2.57: Fläche 0.005
 - 2.32: Fläche 0.010
 - 1.96: Fläche 0.025
 - 1.64: Fläche 0.050
 - 0: Fläche 0.50
 - 1.64: Fläche 0.950
 - 1.96: Fläche 0.975
 - 2.32: Fläche 0.990
 - 2.57: Fläche 0.995

- Wahrscheinlichkeiten für X-Punkte können abgelesen werden, Wahrscheinlichkeiten für X-Intervalle als Integrale berechnet

- **α -Niveau**



-



- bei großer Stichprobe, weil Annäherung an Normalverteilung mit Varianz gleich (Varianz der Population/N)

- **Freiheitsgrade**

- $F = n - u$
 - N = Anzahl unabhängiger Beobachtungen
 - U = Anzahl schätzbarer Parameter

- F = Freiheitsgrad, d.h. Anzahl „überflüssiger Messungen“, die nicht für Schätzung benötigt würden
- entspricht Anzahl unabhängiger Residuen, d.h. Abweichungen der Messwerte von den Parametern; denn jeder geschätzte Parameter wird zu einer Bedingung für die Residuen
- z.B. nach Schätzung des Mittelwerts muss die Summe der Residuen gleich sein; wenn Mittelwert bekannt, können in Linearkombination die Werte der ersten N-1 Personen frei gewählt werden der Wert der n-ten ist dann aber definiert. N-1 heißt Freiheitsgrad des Kennwerts.
- Ist der Mittelwert der Population aber bekannt und wird nicht aus den Daten geschätzt, dann hat er f=n: Die Werte können in beliebiger Weise von ihm abweichen, ohne Bedingung.

- **Erwartungswerte**

- Über sie kann man schätzen, welche Werte unter Bedingung der H0 zu erwarten sind.
- Beispiel: Einfaktorielle Varianzanalyse

$$E(MS_{between}) = \sigma_e^2 + \frac{\sum_{j=1}^p n_j \cdot \alpha_j^2}{p - 1}$$

$$E(MS_{within}) = \sigma_e^2$$

-
- MS = Mean Squares
- Wenn alle Effekte α gleich 0 sind, wird der Bruch 0 und beide Erwartungswerte sind gleich. Sonst nicht. Daraus folgt:
 - H0: $E(MS_{between}) = E(MS_{within})$
 - H1: $E(MS_{between}) > E(MS_{within})$

- **Teststärke/Power (1- β)**

- = Wahrscheinlichkeit, dass in Population vorhandener Unterschied bei statistischer Testung entdeckt wird
- Gegenläufig zum Beta-Fehler
- Hängt ab von [genau denselben Dingen wie Beta-Fehler]:
 - α -Niveau
 - Ein- oder zweiseitige Testung
 - Homogenität der Merkmalsverteilung (d.h. Streuung σ von x)
 - Stichprobenumfang
 - Größe des statistischen Effekts
 - Abhängige versus unabhängige Stichproben
 - Beta-Fehler
- G-Power

- **Häufige Voraussetzungen**

- Stichprobengröße
- Normalverteilung
- Homoskedastizität
- Linearer Zusammenhang
- **Skalenniveau**
 - Nominalskala: Eindeutigkeit der Messwerte
 - Ordinalskala: Rangordnung der Messwerte
 - Intervallskala: Verhältnisse der Messwerte
 - Verhältnisskala: Verhältnisse der Messwerte, absoluter Nullpunkt
 - Division von Merkmalsausprägungen und Multiplikation mit Zahlen sinnvoll
 - Absolutskala: Einheit nicht willkürlich definiert

Allgemeines Lineares Modell (ALM/GLM)

- Verallgemeinerndes statistisches Modell
- Untersucht, ob Ausprägung einer Person in einer AV durch verschiedene UVs erklärt werden kann
- *Besteht aus drei Teilen*
 - *Gewichtete Linearkombination von Variablen*
 - *Zufallskomponente (abhängig von Verteilung, die der abhängigen Variable zugrundeliegend angenommen wird)*
 - *Z.B. Binomial, Normal, Multinomial*
 - *Link-Funktion, die Zusammenhang zwischen Linearkombination und Zufallskomponente spezifiziert*
 - *Z.B. Log, Logit, Identität*
- **Definition: Linearkombination**

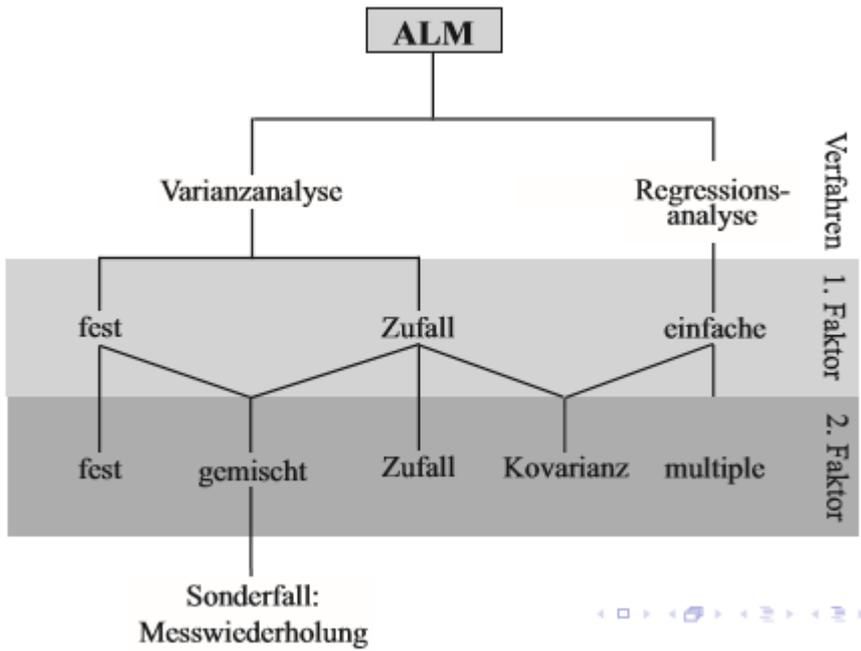
$$y_i = a_0 \cdot x_{i0} + a_1 \cdot x_{i1} + a_2 \cdot x_{i2} + \dots + a_p \cdot x_{ip} + e_i$$

 - y_i : Wert einer Person i in AV y
 - x_{ij} : Werte der Person i in UVs
 - x_{i0} ist normalerweise gleich 1, sodass a_0 den y-Achsen-Abschnitt darstellt
 - a_j : b-Gewichte in der Regression
 - e_i : individueller Fehler
- **Integrativ**
 - Integriert
 - die wichtigsten Verfahren der Elementarstatistik (z.B. t-tests)
 - die Korrelations- und Regressionsrechnung
 - die Varianzanalyse
 - Alle statistischen Berechnungen, denen ein lineares Modell zugrunde liegt (also alle in diesem Skript bis auf LogR) lassen sich in einer ALM-Gleichung ausdrücken.
- **Spezifische Verfahren**
 - Einfache lineare Regressionsanalyse:

$$y_i = a_0 + a_1 \cdot x_i + e_i$$
 - a_0 : Konstante
 - a_1 : b-Gewicht
 - x_i : Wert der Person i auf UV
 - T-Test:

$$y_i = a_0 + a_1 \cdot x_{i1} + e_i$$
 - a_0 : Mittelwert in Kontrollgruppe
 - a_1 : Mittelwertunterschied der beiden Gruppen (= Effekt)
 - x_{i1} : Kodierung der Gruppenzugehörigkeit von Person i (0 = KG, 1 = EG)
- **Parameter für Entscheidung für ein Verfahren**
 - Skalenniveau der UV
 - Ziel der statistischen Analyse (ineinander überführbar)
 - (Primäres Ziel immer
 - Varianzaufklärung
 - Zusammenhang UV -> AV)
 - Vorhersage der AV? -> Regressionsanalyse (eines der diversen Verfahren, je nach Zahl und Eigenschaften der UVs)
 - Ermittlung des Einflusses der UV (=Gruppenzugehörigkeit) auf die mittlere Ausprägung in der AV in der Gruppe? -> Varianzanalyse (t-Test)
 - Nominalskalierte UV: Varianzanalyse für feste Effekte

- Intervallskalierte UV: Varianzanalyse für Zufallseffekte
- Mehrere UVs: Mehrfaktorielle Varianzanalyse
- Eine oder mehrere AVs?
 - Eine AV: univariate Varianz- oder Regressionsanalyse
 - Mehrere AVs: multivariate Varianzanalyse oder Strukturgleichungsmodell

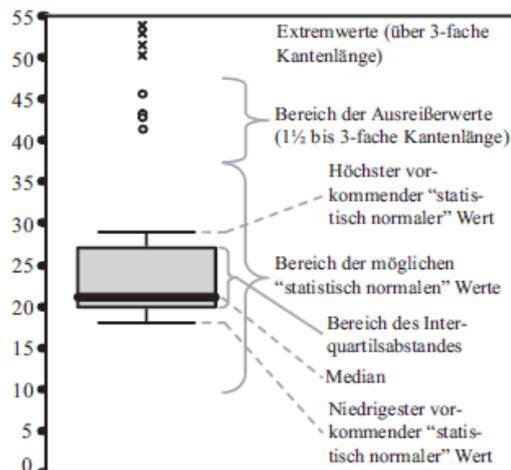


Datenvorbereitung

- **Angeben im Methodenteil**
 - Fehlende Werte
 - Ausreißeranalyse
 - Gründe für Ausschlüsse
- **Datenquellen**
 - Befragungen (persönlich, Internet)
 - Experimente
 - Beobachtungen

Fehlerhafte Daten

- Fehlerquelle: Faktor Mensch
 - Sehr viele Werte
- **Was kann man tun?**
 - Vorher
 - Doppelte Dateneingabe erfordern; mind. 10. Prozent zufällig auswählen und vergleichen
 - Excel: Fordern, dass Eingabe ein bestimmtes Format haben oder einem von x festgelegten Werten entsprechen muss
 - Zeit für Beantwortung erheben
 - Nachher
 - Vor Analysen prüfen (grafisch) auf
 - Korrektheit/Verständnis des Probanden
 - Plausibilität
 - Fehleingaben
 - Doppelte Fälle (kann man in SPSS bei „Daten“ ermitteln lassen)
 - Explorative Datenanalyse
 - Wie groß Veränderungen bei Prä-Post-Messungen?
 - Bleibt Geschlecht „konstant“ über Messzeitpunkte?
 - Sind Daten realistisch und überhaupt möglich?
 - Ausreißer und Extremwerte – manchmal schwer von falscher Eingabe zu unterscheiden
 - Nicht sparen, viele Ausgaben/Ergebnisse vergleichen
- **Ausreißeranalyse**
 - Wie finden?
 - Definition 1: Kriterien nach Tukey (Box-Plot)



- Definition 2
 - Bei Stichproben mit $N < 80$ sind Ausreißer $\pm 2,5$ SD vom Mittelwert
 - Bei größeren Stichproben ± 4 SD vom Mittelwert
- Zusätzlich bivariate Methoden (Scatterplots)
- Problem
 - Gradwanderung zwischen
 - Möglichst große Stichprobe
 - Möglichst repräsentativen Daten
 - Keine generellen Richtlinien
- Was tun?
 - Indexvariablen zur Markierung der jeweiligen Ausreißer pro Item
 - Hinweis auf „kritische Probanden/Subgruppen“
 - Schauen, ob die sich in bestimmten Variablen bedeutsam unterscheiden (z.B. Zweiterkrankung)
 - Ausschluss immer begründen
 - Eventuell Analysen mit und ohne Ausreißer rechnen
 - Sollten sich nicht bedeutsam unterscheiden! Sonst ... sehr gut nachdenken ...
 - **Winsorisieren**
 - Ersetzen, um den Einfluss des Ausreißers zu reduzieren, ohne dass viele Daten verloren gehen
 - 90%-KI um Mittelwert der Stichprobe; alle Werte außerhalb werden ersetzt
 - Alle Werte unterhalb der 5%-Grenze durch Wert der unteren Grenze
 - Alle Werte über der 95%-Grenze durch Wert der oberen Grenze
 - Vorteil: N bleibt erhalten
 - Nachteil: gute Schätzung des wahren Werts unwahrscheinlich

Fehlende Werte

- **Definition:** Wenn Wert einer Person im Datensatz fehlt, obwohl Merkmalsausprägung empirisch vorhanden
 - Achtung: Manche Merkmale existieren nicht bei jedem!
- **Problem**
 - Häufig
 - Behindern Auswertung und Interpretation
 - Verringern Effizienz und Power (vor allem bei multivariaten Analysen und MW-Designs, wo üblicherweise Fälle mit fehlenden Werten komplett ausgeschlossen werden)
 - Man muss mit systematischer Verzerrung rechnen (warum fehlen genau diese Werte?)
 - Zu wenig diskutiert
- Könnten auch als wichtige Information gewertet werden
- **Typen fehlender Werte**
 - **Missing completely at random (MCAR)**
 - Fehlende Werte über alle Beobachtungen hinweg zufällig verteilt, d.h. Fälle mit fehlenden Werten unterscheiden sich nicht von anderen – weder auf der eigenen Variable noch auf den anderen
 - Deshalb geht man davon aus, dass es auch keinen Zusammenhang des Auftretens eines fehlenden Wertes mit der Ausprägung der Person auf dieser Variablen gibt
 - Datensatz mit vollständigen Daten repräsentative Teilstichprobe des Gesamtdatensatzes; von ihr kann auf alle Probanden geschätzt werden
 - „im wahren Leben“ meist zu streng, da bei vielen Variablen eigentlich immer irgendwo ein signifikanter Zusammenhang
 - Inhaltlich auch ziemlich komisch, weil irgendwarum müssen die ja fehlen
 - **Missing at random (MAR)**
 - Auftreten eines fehlenden Werts vollständig durch Ausprägung in anderen Variablen erklärt, d.h. Personen mit unvollständigen Daten unterscheiden sich von solchen mit vollständigen – aber nur auf einer anderen Variable im Datensatz, NICHT auf der eigenen
 - Diese Zusammenhänge können bei Ersetzung berücksichtigt werden
 - Je mehr korrelierende Variablen, desto besser Schätzung
 - Gute moderne Verfahren -> Datensatz, der Relationen in Populationen gut wiedergibt
 - Nur, wenn genügend weitere Variablen erhoben wurden!
 - Schon vorher überlegen, welche fehlenden Werte auftreten könnten und was man gleich miterheben könnte, um das zu klären
 - Demographische Variablen am Anfang -> Drop-Out besser erklärbar ... macht aber Drop-Out groß (und Priming?) -> also lieber doch ans Ende, aber vllt. 1 wichtige Variable gleich am Anfang
 - Ob Ersetzung gut ist, kann nicht am Datensatz getestet werden!
 - Fehlen liegt nicht an Ausprägung in Variable selbst
 - **Not missing at random (NMAR/NRM/non-ignorable)**
 - Auftreten eines Fehlenden Werts nur in Zusammenhang mit der Ausprägung auf dieser Variablen – egal, ob auch noch mit einer anderen -> systematische Verzerrung
 - Keine Variable im Datensatz erlaubt Vorhersage des Fehlers
 - Jede Form der Ersetzung sehr schwierig
- **Missing-Data-Diagnose**
 - Wie viele fehlende Werte?
 - Je Person?
 - Je Variable?
 - Eventuell danach Ausschluss einzelner Personen/Variablen ausreichend
 - Falls mehr als 5% pro Variable/Person: intensivere Diagnostik

- Gruppenunterschiede fehlend/nicht fehlend
 - auf diesem Merkmal zu anderem Zeitpunkt?
 - Demographisch?
 - Kontroll-/Experimentalgruppe
 - Andere Merkmale
- Typ (MCAR/MAR/NMAR)
 - Achtung: keine finale Prüfung
 - Achtung: konservative Bewertung
- **Ersetzung/Ausschluss**
 - **Klassische Verfahren** (alle schlecht, besonders bei vielen fehlenden Werten, da immer Verzerrung)
 - **Listenweiser Ausschluss**
 - komplett aus Analyse ausgeschlossen
 - Problem: evtl. starke Reduktion von N
 - **Paarweiser Ausschluss**
 - für Teilberechnungen ausgeschlossen
 - Problem: evtl. werden unterschiedliche Statistiken im Rahmen der Analyse an unterschiedlichen Substichproben berechnet -> nur bedingt zusammen interpretierbar
 - **Mittelwertersetzung**
 - Mittelwert der Variable zur Ersetzung verwendet
 - Problem: Kovarianzunterschätzung
 - **Regressionsimputation**
 - Vorhersage des fehlenden Wertes
 - Problem: Kovarianzüberschätzung
 - **Moderne Verfahren**
 - **Expectation-Maximization-Algorithmus** (EM-Algorithmus)
 - In Kombination mit multipler Imputation momentan das beste Mittel
 - Kann einzelne Werte ersetzen
 - Nicht alleine benutzen, da dann meist keine Zufallskomponente hinzugenommen
 - *1. Schritt: Schätzen der Mittelwerte und Varianzen aus den vollständigen Fällen; dann Regressionsgleichungen für jede einzelne Variable als Kriterium mit allen anderen Variablen als Prädiktoren; dann Einsetzen der Werte der unvollständigen Fälle in diese Gleichungen -> erste Schätzung der fehlenden Werte*
 - *2. Schritt: Jetzt dasselbe nochmal, mit allen Werten inklusive der gerade geschätzten*
 - ... (typischerweise so 25-mal)
 - **Multiple Imputation**
 - State of the art (mit maximum-likelihood-Methode)
 - Keine einzelnen Werte ersetzt, nur Mittelwerte etc. besser geschätzt
 - Erweiterung einfacher Imputationsmethoden (Mittelwert, Regression)
 - Man ersetzt die fehlenden Werte auf viele verschiedene Weisen durch $m > 1$ plausible Werte, ermittelt jeweils die Statistik und führt dann alle diese Statistiken zusammen (mittelt)
 - SPSS: Variablenniveau muss richtig angegeben sein!
 - Ab 5%, bis 50% fehlende Werte
 - Wenn dadurch deutlich mehr Varianz: schlecht; trotzdem machen, aber transparent sein, Fußnote, dass Datensatz auf Anforderung zur Verfügung steht
- **Kennwerte**
 - **Fraction of missing information** nach Schafer
 - = Anteil der Gesamtvarianz, die auf fehlende, imputierte Werte zurückzuführen ist
 - Misst Unsicherheit bezüglich der Werte, die bei Imputation geschätzt würden, sowie bezüglich der Schätzungen, die damit vorgenommen werden
 - **Relative Zunahme der Varianz durch Imputation**

- Quotient, der Veränderung der Varianz beschreibt
- > 1: Varianzerhöhung
- < 1: Varianzreduktion
- = 1: keine Veränderung
- Keine Grenzwerte
- **Relative Effizienz**
 - Definition: „Relative Effizienz eines erwartungstreuen Schätzers im Vergleich zu einem anderen erwartungstreuen Schätzer für denselben Parameter definieren wir als den Quotienten der Varianz des einen und der Varianz des anderen.“
 - Gleiche Varianzen -> $\eta = 100\%$
 - Quotient, der Varianz des Parameterschätzers bei m Imputationen im Verhältnis zu unendlich vielen Imputationen setzt
 - Notwendig zur Bewertung der Imputation
 - Grenzwert: .9 oder höher
 - In den meisten Fällen reichen m = 5 Imputationen!
- **Anwendung in SPSS**
 - Durchführen Fehlende Werte-Diagnose: „Analysieren“ -> „Analyse fehlender Werte“
 - „Muster“
 - Fälle bei weniger als 1% übergehen
 - Variablen sortieren
 - Fälle mit fehlenden Werten sortiert nach Muster (erst ab Version 22! Davor fehlerhaft!!!!)
 - Variablen sortieren
 - „Deskriptive Statistik“
 - Alles anklicken
 - Z.B. t-Test für Gruppen, die durch Indikatorvariablen gebildet werden
 - Variablen übergehen mit f.W. in weniger als 5% der Fälle
 - Durchführen MI: „Analysieren“ -> „Multiple Imputation“ (erst ab Version 22 benutzen!!) -> „Fehlende Datenwerte imputieren“
 - In neue Datendatei schreiben
 - Dort Daten -> Datei aufteilen; dann Gruppen vergleichen -> Gruppen basierend auf Imputationsnummer
 - Anschließend „normal“ rechnen; SPSS erkennt MI
 - SPSS kennzeichnet Berechnungen, bei denen MI funktioniert
 - Interpretation MI
 - „unterstrichenes x“-Symbol: diese Werte sind schon auf Imputation vorbereitet

Faktorenanalyse

- Nicht ein einzelnes, sondern eine Gruppe statistischer Verfahren
- Ziel: große Anzahl *manifeste* Variablen auf kleinere Anzahl hypothetischer, d.h. *latente* Faktoren reduzieren, wobei möglichst viel Varianz der ursprünglichen Variablen beibehalten werden soll
- Anwendung: Fragebogenkonstruktion
- 2 Typen
 - Explorativ
 - Um Faktoren in Variablenatz zu finden, also Zusammenhänge zwischen den Variablen
 - Anwendung: z.B. Fragebogenerstellung
 - Keine theoretischen Grundlagen zu Anzahl oder Zuordnung (d.h. ich brauche keine, sollte aber eigentlich auch keine haben)
 - Konfirmatorisch
 - Prüft, ob empirisch gefundene Daten zu vorher definiertem theoretischem Modell passen
 - Keine alternativen Zuordnungen testbar
 - Verschiedene Kennwerte für Passung

Exploratorische Faktorenanalyse

- **Ziel:** Faktoren in Variablenatz finden, also Zusammenhänge zwischen den Variablen
- **Anwendung:** z.B. Fragebogenerstellung
- Keine theoretischen Grundlagen zu Anzahl oder Zuordnung nötig! (d.h. ich brauche keine, sollte aber eigentlich auch keine haben)
- *Leonhart macht EFA generell nur noch, um in Datensatz Strukturen zu finden, wenn er dann in zweiter Analyse KFA machen kann (Problem: wie oft hat man schon einen zweiten Datensatz?)*
- **Grundlagen**
 - **Erste Grundgleichung der FA:**
$$z_{ij} = f_{i1} \cdot a_{1j} + f_{i2} \cdot a_{2j} + \dots + f_{ip} \cdot a_{pj}$$
 - - F: Wert der Person auf dem Faktor
 - A: Ladung des Items auf den Faktor; *Determinationskoeffizient*
 - Zeigen Korrelationen zwischen Faktor und Ausgangsvariable
 - Quadrieren -> Anteile der erklärbaren Varianz
 - Quadrate über alle Faktoren für ein Item summieren: Kommunalität
 - Quadrate über alle Items für einen Faktor summieren: Eigenwert
 - D.h. jeder beobachtete Wert jeder (hier standardisierten) Variablen lässt sich als Linearkombination mehrerer hypothetischer Faktoren beschreiben
 - In Matrixschreibweise für alle Personen:
$$Z_{N \times p} = F_{N \times p} \cdot A_{p \times p}^T$$
 - **Zweite Grundgleichung der FA**
 - ergibt sich aus erster; berücksichtigt reduzierte Anzahl der Faktoren -> Fehlerkomponente E, da Gesamtvarianz jetzt nicht mehr ganz erklärbar
 - $$Z_{N \times p} = F_{N \times q} \cdot A_{p \times q}^T + S_{N \times p} \cdot B_{p \times p} + E_{N \times p} \quad (1)$$
$$z_{ij} = f_{i1} \cdot a_{1j} + f_{i2} \cdot a_{2j} + \dots + f_{iq} \cdot a_{qj} + s_{ij} \cdot b_j + e_{ij}$$
 - **Dritte Grundgleichung der FA**

$${}_h R_{p \times p} = A_{p \times q} \cdot A_{p \times q}^T$$

- Beschreibt (unperfekte, da um Uniqueness reduzierte) Reproduktion der Korrelationen der Ausgangsvariablen

- **Ablauf: Im Prinzip**

- Stell dir die Wolke im Raum vor, mit so vielen Dimensionen wie Variablen im Datensatz
 - Keine Korrelationen -> Kugel
 - starke Korrelationen -> „Zeppelin“-bzw.-*Ellipsoid*-Form
- Wir spannen durch diese Punkte wohl Faktoren als Vektoren λ (*Hauptachsen*), die ein Koordinatensystem so aufspannen, dass die größte Varianz aller Variablen erklärt wird
- Winkel zwischen Vektoren ist Maß für Korrelation der zugrundeliegenden Faktoren

$$r_{xy} = \cos \angle \vec{x}, \vec{y}$$

- (d.h. Kosinus des Winkels der Vektoren)
- Erster Vektor λ_1 wird so gelegt, dass er möglichst viel Varianz aufklärt
 - Größte Varianz da, wo Ellipsoid größten Durchmesser, d.h. größter Anteil der Streuung aller Messwerte -> Vektor, der die längste Strecke im Ellipsoid erfasst, klärt maximal viel Varianz auf
- Zweiter Vektor wird so dazugelegt, dass er
 - möglichst viel Residualvarianz (Restvarianz, NICHT Varianz der Residuen oder so) aufklärt
 - und unabhängig (unkorreliert) vom ersten Vektor ist -> senkrecht zu erstem Vektor
- Und so weiter; iteratives Verfahren
 - Mehrere Schleifen (normalerweise 7-12) -> optimale Schätzung notwendiger Parameter
 - Wenn SPSS sagt, nach 25 Iterationen hat es noch nichts gefunden, heißt das, es gibt ein Problem im Datensatz -> einfach Anzahl der Iterationen hochsetzen bringt nichts
- Man kann so viele Faktoren reinlegen, wie man ursprünglich Dimensionen (also Variablen) hatte, aber dann wird die Information nicht reduziert, denn die komplette Varianz ist aufgeklärt

- **Varianzzerlegung eines Items** (d.h. einer ursprünglichen Variablen)

$$\sigma_j^2 = \underbrace{a_{j1}^2 + a_{j2}^2 + \dots + a_{jq}^2}_{\text{Kommunalität}} + \underbrace{b_j^2}_{\text{Spezifität}} + \underbrace{e_j^2}_{\text{Fehler}}$$

$\underbrace{\hspace{10em}}_{\text{Reliabilität}}$

- **Spezifität:** Varianzanteil, der durch einen (theoretischen, nicht tatsächlich gefundenen) spezifischen Faktor erklärt werden kann (könnte)
- **Fehler:** Varianz des Fehlers
- **Reliabilität:** zuverlässig durch die Faktoren erklärte Varianz
- **Uniqueness:** Spezifität + Fehler (weil variablenspezifisch); praktisch sind die beiden auch nicht auseinanderzuhalten

- **Voraussetzungen**

- Intervallskalierte & normalverteilte oder dichotome Variablen
- N sollte mindestens 3x, eigentlich aber 5x so hoch sein wie die Anzahl der Variablen
- Bedeutsame, d.h. *substanzielle* Zusammenhänge
 - Prüfen anhand Korrelationsmatrix
 - **Bartlett-Test**
 - Immer berichten!
 - Weicht $R_{p \times p}$ signifikant von Einheitsmatrix ab? (soll sein)
 - Erfassen die extrahierten Faktoren die Gesamtvarianz hinreichend?
 - **Prüfgröße von Kaiser-Mayer-Olkin**
 - Immer berichten!
 - „measure of sampling adequacy“ (MSA)

$$KMO = \frac{\sum \sum r_{ij}^2}{\sum \sum r_{ij}^2 + \sum \sum r_{ij,z}^2}, i \neq j$$

-
- = Summe der einfachen Determinationskoeffizienten aller Variablenkombinationen (d.h. gemeinsamer Varianzanteil) / gemeinsamer Varianzanteil aller Variablen plus Summe der quadrierten Partialkoeffizienten
- Partialkoeffizienten-Quadrat-Summe ist Indikator für Höhe der Fehlerkorrelationen
 - Wenn klein, wird KMO groß -> Matrix beinhaltet viel gemeinsame Varianz
 - Wenn groß, wird KMO klein -> Matrix beinhaltet viel Spezifität
- Zwischen 0 und 1
- Deutsche Übersetzung:
 - KMO > 0,9 erstaunlich
 - KMO > 0,8 verdienstvoll
 - KMO > 0,7 ziemlich gut
 - KMO > 0,6 mittelmäßig
 - KMO > 0,5 kläglich
 - KMO < 0,5 untragbar
- **Bildung der Inversen**
 - Liegen Elemente außerhalb der Hauptdiagonalen in der inversen Matrix nahe bei 0? (soll sein)
 - Add-on
 - Inverse Matrix = Matrix, die mit der Ausgangsmatrix multipliziert die Einheitsmatrix ergibt

Ablaufschema: Matrizen

	$X_{N \times p}$	Rohdatenmatrix Matrix der Ausgangswerte	Personen in Reihen, Variablen in Spalten
Z-Transformation	$Z_{N \times p}$	z-transformierte Datenmatrix normierte Datenmatrix	Alle Variablen identische Streuung 1 und Mittelwert 0
Transponation	$Z^T_{N \times p}$	Transponierte z- transformierte Datenmatrix	
$1/(N-1) * Z_{N \times p} * Z^T_{N \times p}$ (wenn keine fehlenden Werte) = Produkt-Moment-Korrelation für z-transformierte Daten	$R_{p \times p}$	Korrelationsmatrix	Hauptdiagonale nur 1
Prüfen, ob FA sinnvoll (Bartlett, Inverse, KMO)			
Kommunalitätenproblem -> iterative Schätzung			
In Diagonale keiner 1er mehr, sondern erklärter Varianzanteil	$hR_{p \times p}$	Reduzierte Korrelationsmatrix	
Extraktionsproblem			
	$A_{p \times q}$	Faktorladungsmatrix	

Rotationsproblem -> iterative Schätzung			
Evtl. Rotation	$A'_{p \times q}$	Rotierte Faktorladungsmatrix	
Faktorwertproblem			
	$F_{N \times q}$	Faktorwertematrix	

- **Probleme/Entscheidungen:**

- **Kommunalitätenproblem:** Kommunalitäten (aufgeklärte Korrelationen der Variablen) kleiner als tatsächliche Korrelationen, weil nicht gesamte Kovarianz aufgeklärt werden soll ($q < p$; Vorhersagefehler entsteht) -> lassen sich erst bestimmen, wenn Analyse schon gerechnet ist, aber sind nötig für Analyse -> müssen vorher geschätzt werden -> dafür gibt es mehrere Möglichkeiten (HKA, HAA)
- **Extraktionsproblem:** Wie viele Faktoren sollen extrahiert werden?
- **Rotationsproblem:** Welche Rotationstechnik?
- **Faktorwertproblem:** Welche Faktorwerte sollen berechnet werden? -> auch hierfür gibt es mehrere Techniken mit unterschiedlichen Lösungen
- Viele Möglichkeiten -> vor dem Rechnen müssen all diese Entscheidungen getroffen werden -> verschiedene Forscher kommen am selben Datensatz durch unterschiedliche Entscheidungen zu verschiedenen Ergebnissen
- Versuchung: „Wenn mir das Ergebnis nicht passt, rechne ich halt nochmal ...“

- **Kommunalität h^2**

- = aufsummierte quadrierte Ladungen aller Faktoren auf dieses Item

$$h_j^2 = \sum_{k=1}^p a_{jk}^2$$

- Zwischen 1 (100 % erklärbar) und 0
- SPSS: Zweite Spalte („Extraktion“) der „Kommunalitäten“-Tabelle
- Sinn:
 - Erklärbare Varianz des Items (durch alle Faktoren)
 - Mindestschätzer für Reliabilität
 - Kennwert, der uns hilft, zu bewerten, ob eine Variable gut in unsere Struktur passen kann (z.B. ein Item in den restlichen Datensatz, d.h. den zukünftigen Fragebogen)
- *Leonhart: Kommunalitäten über .9 sind sehr verdächtig, weil man einfach immer nur dieselbe Frage gestellt hat*
- Zwei Arten der Schätzung -> zwei Arten der FA (eigentlich 8 oder 9 Arten, die aber keiner kennt)
 - **Hauptkomponentenanalyse** (HKA, principle component analysis, PCA)
 - Korrelationsmatrix „positiv semidefinit“: man geht von vollständiger Varianzaufklärung aus
 - „wir fangen mal mit 100 % und gehen dann runter“ -> alle Kommunalitäten anfangs auf 1 gesetzt, d.h. man arbeitet mit Korrelationsmatrix
 - Leonhart selber nimmt meistens diese
 - Ziel: möglichst viel Gesamtvarianz erklären, egal, ob gemeinsam oder nicht
 - **Hauptachsenanalyse** (HAA, principle factor analysis, PFA)
 - Kommunalitäten vor erster Iteration geschätzt (Mindestschätzung)
 - Anfangskommunalität R^2 : aus Regressionsanalyse eines Items als Kriterium mit restlichen Items als Prädiktoren
 - „wir machen vorher schon mal zu Beginn eine Schätzung und fangen mit einem auf jeden Fall zu kleinen Wert an“

- Ziel: möglichst viel *gemeinsame* Varianz der Variablen beschreiben; d.h. gemeinsame Ursache für Ausprägungen auf verschiedenen Variablen

- **Eigenwert λ**

- = summierte Quadrierte Ladungen aller Items auf einen Faktor

$$\lambda_k = \sum_{j=1}^p a_{jk}^2$$

-

- **Bedeutung:**

- Eigenwert 1: Faktor erklärt so viel wie 1 ursprüngliche Variable
 - -> das ist bei vielen ursprünglichen Variablen natürlich nicht so viel, bei wenigen schon -> p berücksichtigen
- Anzahl Variablen/Eigenwert: % der Gesamtvarianz, die dieser Faktor erklärt
- WURZEL(λ) = Länge der Hauptachse (Hauptachse als Vektor; Länge entspricht der Strecke, die sie in der Punktwolke verläuft)

- Sinn: Bewertung eines Faktors

- Problem: Man kriegt Faktoren mit sehr sehr hohen Eigenwerten raus (so ein Faktor heißt „*Generalfaktor*“) und welche mit sehr sehr niedrigen (-> Rotation, um das gleicher zu verteilen)

- **Kriterien zur Anzahl der Faktoren**

- *Leonhart: „suchen Sie sich was aus“, spricht nie alles für dieselbe Anzahl an Faktoren*

- **Kaiser-Gutman-Regel**

- Eigenwert > 1
 - Alles, was größer ist, extrahieren wir, alles, was kleiner oder gleich ist, extrahieren wir nicht
- Voraussetzungen
 - $P < 40$
 - $N > 5 * p$
 - Erwartete Anzahl der Faktoren zwischen $p/5$ und $p/3$
- Vorteil: hartes Kriterium, kann nicht im Nachhinein verschoben werden
 - *Leonhart nimmt deshalb dieses Kriterium; er findet, man sollte in Statistik nicht zu viele Optionen haben, damit man nicht versucht ist, zu spielen*
- *Leonhart: In Publikation würde man eher neuere Verfahren nehmen*

- **Kriterium der extrahierten Varianz**

- Vorher definiert, wie viel Gesamtvarianz erklärt werden soll
 - Muss theoriegeleitet begründet sein
- Problem: keine allgemeinen Kriterien für akzeptablen Prozentwert
 - *Leonhart: „Das macht mir noch mehr Bauchweh.“*

- **Scree-Test**

- Grafisch, also gar kein Test, sondern Scree-Plot
- „Knick“ im Eigenwerteverlauf der Größe nach von links nach rechts
- VOR dem Knick abbrechen!
- Gut bei vielen Variablen ($p > 40$)
- Manchmal sehr sauber und schön;
- Manchmal nicht sauber -> dann willkürlich

- **Evaluation der Lösung**

- Mehrere Faktorenlösungen anschauen; welche passt mir inhaltlich am besten/ist plausibel?
- Validierung notwendig! (Kreuzvalidierung)

- **Paralleltest nach Horn**

- Der Paralleltest ist eine Monte-Carlo-Simulation
 - = Verfahren aus der Stochastik, bei dem eine sehr große Zahl gleichartiger Zufallsexperimente die Basis darstellt. Es wird dabei versucht, analytisch nicht oder nur aufwendig lösbare Probleme mit Hilfe der Wahrscheinlichkeitstheorie numerisch zu lösen

- Neues Verfahren
- Hypothesentest, d.h. echter statistischer Test
- Ziel: Könnten meine Eigenwerte auch zufällig sein (d.h. es stecken gar keine Faktoren dahinter?)
- Methode
 - Man simuliert zufällige Datensätze (z.B. 1000) mit demselben N und derselben Itemzahl wie die Stichprobe und rechnet an denen die Faktorenanalyse
 - Dann bildet man die Verteilung der Eigenwerte, die dabei rauskommen
 - Extrahieren aller Faktoren, deren zufällige Eigenwerte größer sind als Vergleichswert
 - Entweder man vergleicht die empirischen Eigenwerte mit dem Mittel der zufälligen Eigenwerte (z.B. den ersten empirischen Faktor mit dem Mittel der zufälligen ersten Faktoren)
 - oder mit dem 95. Perzentil der zufälligen Eigenwerte
 - oder für den ersten Faktor mit dem Perzentil und danach mit dem Mittelwert
 - Tatsächlich weiß man noch nicht, welches Vorgehen zu besseren Ergebnissen führt! Aber in der Klausur nimmt man immer das 95 %-Perzentil.
- Macht SPSS noch nicht selbst; gibt ein Makro dafür
<https://people.ok.ubc.ca/briocconn/nfactors/nfactors.html>
- **Velicer's MAP-Test (minimum average partial)**
 - Neues Verfahren (na ja, von 1976)
 - Kein statistischer Test, eigentlich deskriptiv
 - Methode
 - Schleife (bis mittlere quadrierte Partialkorrelation nicht weiter reduziert ... bzw. erst mal bis so viele Komponenten wie Variablen)
 - Komponente extrahieren (mit HKA)
 - Danach diese Komponente aus den restlichen Korrelationen auspartialisieren -> Partialkorrelationsmatrix (enthält nur Varianz, die noch nicht aufgeklärt ist)
 - Durchschnittliche quadrierte Korrelation der Residualmatrix ermitteln (also, der vorübergehenden Residualmatrix, d.h. der Partialkorrelationsmatrix)
 - Die Partialkorrelationen werden eine Weile lang nach jedem Faktor kleiner und dann plötzlich wieder größer – vor diesem Faktor hört man dann auf (d.h. man wählt die Anzahl von Komponenten mit der gemeinsam niedrigsten quadrierten partiellen Korrelation)
 - Nachfolgende Komponenten erklären keine systematische Varianz
 - Macht SPSS noch nicht selbst; gibt ein Makro dafür
<https://people.ok.ubc.ca/briocconn/nfactors/nfactors.html>
- **Rotation**
 - Ziel
 - Inhaltlich sinnvolle Interpretation
 - Wird erreicht durch möglichst gute Einfachstruktur, d.h. jede Variable lädt nur auf genau einen Faktor hoch und die anderen sehr niedrig
 - Verfahren
 - **Orthogonal**
 - Unabhängigkeit der Faktoren bleibt erhalten
 - **Oblique**
 - Faktoren dürfen korrelieren
 - Das kann manchmal zu besserer Interpretierbarkeit führen

- *Leonhart empfiehlt immer erst mal orthogonale Rotation für erste Analyse, und wenn die nicht so gut passt, dann oblique (weil man ja vorher gesagt hat, sie sollen nicht korrelieren, deshalb sind die dann meist auch nach obliquen Rotation nur sehr schwach korreliert)*
 - Durch Rotation: Lage der Vektoren ändert sich -> Zuordnung der Variablen auf den Faktoren ändert sich
 - -> insgesamt aufgeklärte Varianz bleibt gleich
 - -> hin zu optimaler Einfachstruktur, sodass Items eindeutig Faktoren zuordenbar sind
 - -> meist ähnlichere Verteilung der aufgeklärten Varianz über die Faktoren
- **Faktorwerte**
 - = Gewichte der Ausprägungen einer Person auf den Faktoren
 - Eigentlich vor Analyse gebraucht, aber erst währenddessen iterativ geschätzt -> vorher geschätzt -> mehrere Möglichkeiten -> Faktorwerteproblem
 - Eher mathematisches als inhaltliches Problem
- **Anwendung in SPSS**
 - Input
 - SPSS sagt NIE, dass N zu klein ist oder zu viele Werte fehlen! Das muss man selber schnallen!
 - Varimax Standard für orthogonal
 - Umgang mit Fehlenden Werten: Bitte listenweise erstmal so drin lassen
 - Output
 - Immer deskriptive Statistiken anschauen!
 - Korrelation und Kovarianz müssen da sein -> geht nur, wenn man Varianz hat! -> wenn 1 Item weniger Varianz hat als die anderen, oder einen Boden-/Deckeneffekt, macht das die Lösung schlechter
 - Items sollten ähnlich verteilt sein (sonst wird Lösung schlechter wegen weniger möglicher Kovarianz)
 - Immer Voraussetzungen prüfen, immer KMO und Bartlett angeben
 - „Erklärte Gesamtvarianz“-Tabelle
 - „Anfängliche Eigenwerte“ -> „Gesamtsumme“: Eigenwerte
 - „Extrahierte Summen von quadrierten Ladungen“ -> „Kumulativ %“ -> aufgeklärte Gesamtvarianz aller extrahierten Faktoren!
 - Da hätte ich eigentlich gerne > 50 %
 - Kommunalitäten sollten alle > .4 sein, sonst passen sie nicht gut (was nicht bedeutet, dass sie schlechte Items sind! Vor Rauswerfen inhaltlich begründen und auf Validität achten!)

Konfirmatorische Faktorenanalyse

- SEM-Anwendung
- VL 12, #4 und #25

Diskriminanzanalyse

- **Ziel**
 - Optimales Vorhersagen/Erklären von Werten einer nicht-metrischen AV (mit 2 oder mehr Stufen) durch (eine oder) mehrere metrische UV(s)
 - Bestimmung von Funktion und Cutoff-Wert, um Gruppen zu trennen
- **Beziehungen zu anderen Analysen**
 - Logistische Regression: macht dasselbe, aber nur für genau 2 Stufen der AV und metrische UVs; unter diesen Bedingungen effizienter als Diskriminanzanalyse
 - ANOVA
 - erster Schritt der DA (Signifikanztest der Gruppenunterschiede) eigentlich inverse ANOVA: Je mehr Gruppen überlappen, desto schwerer lassen sie sich trennen; das hängt von der Varianz innerhalb relativ von den Mittelwertsunterschieden ab – so wie bei ANOVA
 - Zusätzlich bei DA Cutoff-Wert und Evaluierung der Prädiktionsgüte
- **Voraussetzungen**
 - AV
 - Nominalskaliert
 - Wenn metrisch:
 - Dichotomisieren
 - Extremgruppen vergleichen
 - Eindeutige Gruppenzugehörigkeit
 - UVs
 - Normalverteilt
 - Intervallskaliert
 - Stichprobe
 - Mind. 20 pro Prädiktor
 - Mind. 20 pro Kategorie der AV
 - Ähnlich große Gruppen, da sonst Schätzungen problematisch
 - Außer wenn Gruppen in Population unterschiedlich groß sind
- **Ablauf**
 - Ermittlung der Diskriminanzfunktion(en)
 - Evaluation der Diskriminanzgewichte pro Prädiktor (= Signifikanzprüfung)
 - Evaluation der Diskriminanzladungen (pro Person, eher unwichtig)
 - Bewertung der F-Werte
 - Signifikante Differenzen in den Gruppenunterschieden?
 - Signifikanzprüfung der Diskriminanzfunktion(en) overall
 - Bestimmung des/r Trennwerte/s
 - Validierung an der Trefferquote („praktische Signifikanz“)
- **Diskriminanzfunktionen**
 - $$Z_{jk} = a + W_1 \cdot X_{1k} + W_2 \cdot X_{2k} + \dots + W_p \cdot X_{pk}$$
 - A = Konstante
 - Z = Z-Wert der Diskriminanzfunktion
 - Das ist kein normaler Z-Wert, sondern der Wert auf einer neuen, metrischen Variable, die dann entlang des Trennwerts in die nonmetrische Variable „vorhergesagte Gruppenzugehörigkeit“ zerhackt wird
 - kann standardisiert oder unstandardisiert sein und heißt in jedem Fall Z
 - Aus Z-Wert wird p-Wert berechnet, der ist ein Maß für die Wahrscheinlichkeit der Gruppenzugehörigkeit. Direkte Wahrscheinlichkeitsberechnung nur bei LogR.
 - W = Diskriminanzgewicht

- $X =$ Wert der Variablen i für Person k
 - Es werden [Anzahl der UVs – 1] Diskriminanzfunktionen gebildet
 - Bei mehr als 2 Gruppen sind grafische Analysen (z.B. in Form von Territorial Maps) möglich
 - Jede davon grenzt eine bestimmte Gruppe von den anderen ab und hat genau einen Cut-Off-Wert
 - Sie wird so bestimmt, dass diese Gruppe sich in ihrem Centroid (Mittelpunkt) auf dem dieser Funktion zugehörigen Z-Score von den anderen möglichst stark unterscheidet
 - Dafür werden jeweils relevante Prädiktoren aufgenommen, unwichtige nicht und die Prädiktoren nach Relevanz für diese konkrete Funktion gewichtet
 - Von allen Funktionen wird diejenige Kombination ausgewählt, die die AV am besten vorhersagt
 - Es sei denn, man weiß schon theoretisch/aufbauend auf bestehenden Erkenntnissen, welche Variablen getestet werden sollen
 - Besonderheit der DA unter allen Multivariaten Verfahren: Einzige Methode, die mehrere *Variates* berechnet
 - **Kriterien zur Ermittlung der Funktionen**
 - **Wilks Lambda**
 - Koeffizient für Anteil nicht erklärter Varianz (soll möglichst klein sein)
 - Die UV wird aufgenommen, die Lambda am stärksten verkleinert
 - Prüfung über partiellen F-Wert (modifizierbar)
 - Optimierung über alle Gruppen
 - Häufiges/empfohlenes Verfahren (*das, was wir nehmen sollen*)
 - **Nicht-erklärte Varianz**
 - Es wird die UV aufgenommen, die das Maß am meisten verringert
 - Optimierung anhand der am schlechtesten trennbaren Gruppen
 - **Mahalanobis-Abstand D^2**
 - berücksichtigt quadrierte euklidische Differenz zwischen individuellem Wert und Gruppenmittelwert und Streuung und Kovarianz der Variablen
 - Es wird die UV aufgenommen, die die kleinste D^2 zwischen zwei Gruppen maximiert
 - Optimierung anhand der am schlechtesten trennbaren Gruppen
 - **Kleinster F-Quotient**
 - Abhängig von df
 - Deshalb auch von Anzahl der Prädiktoren
 - Es wird die UV aufgenommen, die den F-Wert der Mahalanobis-Distanz maximiert (also zu den am stärksten signifikanten Gruppenunterschieden der Centroide führt)
 - Optimierung der am schlechtesten trennbaren Gruppen
 - **Rao-V**
 - Maß für Unterschiede zwischen Gruppenmittelwerten
 - Entspricht Hotelling's Spur bei der MANOVA
 - Es wird die UV aufgenommen, durch welche Rao-V am stärksten steigt
 - Optimierung für alle Gruppen
 - **Interpretation**
 - *Diskriminanzgewichte*
 - *Analog zu Regressionskoeffizienten*
 - *Auch standardisiert möglich*
 - *Diskriminanzladungen*
 - *Besser als Gewichte*
 - *Entspricht Korrelation der einzelnen Variable mit jeweils einer Diskriminanzfunktion*
 - *Rotation möglich für bessere inhaltliche Erklärung*
- **Trennwert**
 - Bei gleicher Gruppengröße

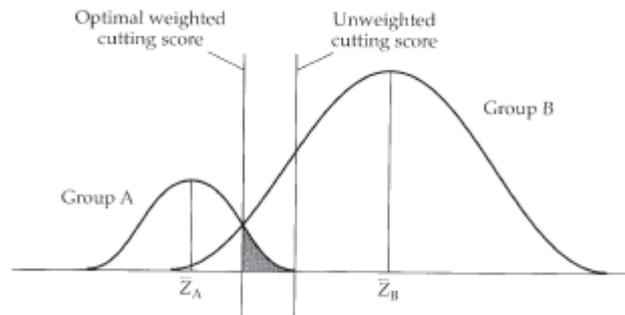
$$Z_{CE} = \frac{Z_A + Z_B}{2}$$

- (Mittel auf Z-Score zwischen Centroiden)
- Z_A = Centroid in Gruppe A
- Z_{CE} = Trennwert

- Bei ungleicher Gruppengröße:

$$Z_{CS} = \frac{N_A \cdot Z_B + N_B \cdot Z_A}{N_A + N_B}$$

- N_A = Anzahl der Beobachtungen in Gruppe A



- Cut-Off verschiebt sich in Richtung der kleineren Gruppe

- **Prädiktionsgüte/Übereinstimmung**

- = relativer Anteil der korrekt vorhergesagten Personen über alle Gruppen hinweg
- Berechnung anhand von Vorhergesagt-Tatsächlich-Matrizen
- Signifikanztests (ineinander überführbar, wenn zwei Gruppen)
 - Zwei Gruppen: T-Test

$$t = \frac{p - .5}{\sqrt{\frac{.5 \cdot (1.0 - .5)}{N}}}$$

- p: relative Häufigkeit der korrekt Klassifizierten
- N: Stichprobenumfang
- Mindestens 2 Gruppen: Press's Q (in SPSS integriert)

$$Press's Q = \frac{(N - (n \cdot k))^2}{N \cdot (K - 1)}$$

- n = Anzahl der richtig Klassifizierten
- N = Gesamtstichprobenumfang
- K = Gruppenanzahl
- Beeinflusst von (d.h. muss bei Interpretation berücksichtigt werden)
 - Anzahl der Gruppen
 - Gruppengrößen

- **Einzelfalldiagnostik**

- Unterscheiden sich richtig und falsch Klassifizierte innerhalb einer Gruppe?

- **Validierung**

- Sehr wichtig, da exploratives Verfahren
- Meist Kreuzvalidierung mit geteiltem Datensatz
- *Leonhardt: Wenn Teilung zu zu kleinen Stichproben führt, lieber nicht teilen und später an anderem Datensatz validieren. Kreuzvalidierung macht Modell nur valider, aber nicht besser im Sinne einer besseren Prädiktion.*

- **Anwendung in SPSS**

- Input
 - Analysieren -> Klassifizierung -> Diskriminanzanalyse
 - Bereich für Gruppenvariable spezifizieren
 - Deskriptiv: „Mittelwert“ (Centroide)
 - Schrittweise Methode: „Zusammenfassung der Schritte“
 - Wichtig, ob Gruppen gleich groß geschätzt oder aus tatsächlicher Gruppengröße berechnet werden sollen; ist Gruppengröße repräsentativ für Gruppengrößen in Population? Dann empirische Werte benutzen!
 - „Klassifikation mit Fallauslassung“: Validierung?
 - „Speichern“:
 - Vorhergesagte Gruppenzugehörigkeit
 - Wahrscheinlichkeiten der Gruppenzugehörigkeit
- Output
 - „Gruppenstatistik“: Centroide
 - „Aufgenommene/Entfernte Variablen“: Übersicht der Schritte bei schrittweiser Auswahl
 - „Variablen in der Analyse“: Toleranzen, Wilks Lambda; diese F-Werte sind NICHT die F-Werte für die Aufnahme der Variablen! Die ist in der „aufgenommene/entfernte Variablen“-Tabelle!
 - „Variablen, die NICHT in der Analyse sind“: anschauen, falls hohe Kollinearität -> dann könnten Variablen nicht aufgenommen werden, die aber durchaus stark mit dem Kriterium assoziiert sind
 - „Eigenwerte“: wie bei FA geben Eigenwerte die Relevanz der Funktionen für die Varianzaufklärung an; im Verhältnis zueinander interpretieren. Ändern sich bei Rotation.

Regressionsanalyse

Einfache (Lineare) Regression

- Gleichung

$$\hat{y}_i = r_{xy} \cdot \frac{s_y}{s_x} \cdot (x_i - \bar{x}) + \bar{y}$$

-
- Problem: Für psychologische Fragestellungen fast immer zu simpel

Multiple (Lineare) Regression

- **Ziel**
 - Ausprägung in Variable Y (Kriterium) mit Variablen X1, X2, X... (Prädiktoren) vorhersagen/erklären
 - Funktioniert (zu einem gewissen Grad), falls Y und X korrelieren
 - Es geht um stochastische (also nicht funktionale) Zusammenhänge; d.h. solche, bei denen ein Zufallselement beteiligt ist
 - Diese werden als Gerade dargestellt
 - Sinn
 - Vorhersage
 - Erklärung
 - Dabei sollen ...
 - Möglichst viel Varianz am Kriterium erklärt werden
 - Und möglichst wenige Prädiktoren verwendet werden
- **Voraussetzungen** (gelten für alle Variablen einzeln UND für die Linearkombination (engl. „*variate*“))
 - Linearer Zusammenhang
 - Wenn nicht:
 - Transformation (log, Wurzel) verantwortlicher Variablen
 - Oder Aufnahme eines Terms in die Gleichung, der die Nichtlinearität ausdrückt
 - Homoskedastizität (d.h. gleiche Varianz auf allen Variablenausprägungen) der Vorhersagefehler
 - Normalverteilung der Vorhersagefehler (mit Mittelwert 0)
 - Unabhängigkeit der Vorhersagefehler von Prädiktorwerten
 - Prädiktor und Kriterium intervallskaliert
 - Wenn Prädiktor nicht: Dummy-Transformation in k-1 Dummies
 - Wenn Kriterium nicht + nur 2 Stufen: logistische Regression
 - Prädiktor und Kriterium normalverteilt
 - Sonst: Transformation
- **Ausreißeranalyse**
 - *Influence = Leverage * Discrepancy; Maß: Cook's Distance*
 - *Leverage = Abweichung vom Mittelwert der x-Achse*
 - *Diskrepanz = Abweichung von Regressionsgerade auf y-Achse*
- **Methode**
 - Auswählen derjenigen Regressionskoeffizienten *b*, die zu den aufsummiert kleinsten quadrierten Vorhersagefehlern führen (**Least Squares-Methode**)
 - **Regressionsgleichung**

$$\hat{y}_i = b_1 \cdot x_{i1} + b_2 \cdot x_{i2} + \dots + b_k \cdot x_{ik} + a_{1.23\dots k}$$

-
- Standardisiert:

$$\hat{z}_{yi} = \beta_1 \cdot z_{i1} + \beta_2 \cdot z_{i2} + \dots + \beta_k \cdot z_{ik}$$

- Statistische Signifikanztests
 - F-Tests
 - Der Regressionsgleichung(en) (=> Omnibustest)
 - Quadratsummen sind Maß für Variabilität
 - Quadratsummenzerlegung: $SS_{total} = SS_{reg} + SS_{res}$
 - SS_{reg} = erklärbare Varianz
 - SS_{res} = nicht erklärbare Varianz
 - Signifikanzprüfung
 - $F = (SS_{reg} : df1) : (SS_{res} : df2)$
 - Frage: Besser als Vorhersage mit Mittelwert des Kriteriums?
 - T-Tests
 - Der Regressionskoeffizienten
 - KI um Punktschätzung der Koeffizienten mit Standardfehler; signifikant, wenn 0 nicht enthalten
 - Der Korrelationskoeffizienten r_{xy} und $r_{\wedge yy}$
- **Strategien zur Modellspezifikation** (d.h. Auswahl der Prädiktoren)
 - **A priori: Researcher Specified**
 - Theorie- und evidenzgeleitet (vorherige Untersuchungen); nur eine Analyse
 - ☺ Vorteil
 - ☺ Keine Capitalization of Chance
 - ☹ Nachteil
 - ☹ Specification Error durch Inklusion irrelevanter (oder multikollinear) oder Auslassung relevanter Variablen
 - **A posteriori: Sample Based**
 - Mehrere Regressionsanalysen mit verschiedenen Prädiktorkombinationen; iterative Auswahl der besten Kombination
 - ☹ Nachteile
 - ☹ Führen zu α -Fehler-Kumulierung
 - ☹ Führen zu Capitalization of Chance
 - Bitte immer Kreuzvalidierung!
 - Beispiele
 - Sequentiell
 - **Schrittweise Regression** („Stepwise“)
 - Ablauf
 - Kombination aus Vorwärtsselektion und Rückwärtselimination
 - Vorwärtsselektion, aber nach jeder Aufnahme ein Rückwärtseliminations-Durchgang, um zu prüfen, ob eine bereits aufgenommene Variable jetzt nicht mehr signifikant ist -> dann wird sie rausgeworfen
 - Stoppt, sobald in einer Runde weder Aufnahme noch Ausschluss erfolgt
 - ☺ Vorteile
 - ☺ Minimum an Prädiktoren, da Ausschluss der nicht-signifikanten

- ☺ Trotzdem Berücksichtigung primär der stark mit Kriterium korrelierten Prädiktoren
 - ☹ Nachteile
 - ☹ Viel Capitalization of Change -> Koeffizienten in Kreuzvalidierung oft nicht bestätigbar
 - **Vorwärtsselektion** („Forward Addition“)
 - Ablauf
 - Zuerst Prädiktor mit höchster Korrelation mit Kriterium
 - Dann iterativ immer derjenige Prädiktor mit der höchsten inkrementellen Varianzaufklärung
 - Stoppt, wenn kein Prädiktor mehr inkrementelle Validität besitzt
 - ☺ Vorteile
 - ☺ Sehr ökonomisch: maximal so viele Analysen wie mögliche Prädiktoren
 - ☺ Signifikanztest (F-Test) vor Inklusion
 - ☹ Nachteil
 - ☹ keine nachträgliche Exklusion möglich, obwohl öfter mal die inkrementelle Validität der ersten Prädiktoren durch später hinzugefügte bis hin zur Nicht-Signifikanz reduziert wird
 - ⇒ Nur in Ausnahmefällen optimale Lösung
 - v.a. bei hoher Multikollinearität
 - **Rückwärtselimination** („Backward Elimination“)
 - Dasselbe wie Vorwärtsselektion, nur andersrum
 - Meist gleiches Ergebnis wie VS, außer bei hoher Multikollinearität – dann besser
 - Kombinatorisch
 - **Alle möglichen Untermengen** („All Possible Subsets“)
 - Ablauf
 - Jede mögliche Prädiktorkombination analysiert
 - Modell mit dem größten R^2 ausgewählt
 - Modelle mit ähnlich großen R^2 ebenfalls präsentiert
 - ☺ Vorteil: Findet tatsächlich optimale Kombination
 - ☹ Nachteil: Viel Rechenzeit
- **Kennwerte**
 - **Varianzaufklärung**
 - **(Multipler) Determinationskoeffizient R^2**
 - = Anteil der Kriteriumsvarianz, der durch alle Prädiktoren zusammen vorhergesagt werden kann
 - = $r_{Y,X1} + r_{Y(X2*X1)} = r_{Y,X2} + r_{Y(X1*X2)}$
 - = SS_{reg}/SS_{total}
 - **Adjusted R^2**
 - Korrigiert dafür, dass selbst ohne Effekt zufallsbedingt jeder Prädiktor ungleich 0 mit dem Kriterium korreliert, was sich in R^2 ansammelt
 - Heranziehen, wenn Modelle verglichen werden sollen, die ... (beides führt sonst zu Unterschieden in R^2 , die nicht auf tatsächlichen Fit zurückgehen)
 - Unterschiedliche viele Prädiktoren
 - Und/oder unterschiedlich große N haben
 - **Korrelationskoeffizient R**
 - = $r_{\hat{y}y}$ = Korrelation der vorhergesagten mit den tatsächlichen Werten
 - **Regressionskoeffizienten b**

- = Veränderung in Kriterium, wenn sich dieser Prädiktor um 1 ändert
- Signifikanztest nur am unstandardisierten b möglich
- Berechnung

$$\beta_1 + r_{x_1x_2} \cdot \beta_2 = r_{x_1y}$$

- Lösen des Gleichungssystems $r_{x_2x_1} \cdot \beta_1 + \beta_2 = r_{x_2y}$ durch Einsetzen der Korrelationswerte

- Nicht eindeutig lösbar -> Least Squares für eindeutige Lösung

- Interpretation

- Problem: Schlecht interpretierbar, wenn unterschiedliche
 - Skalenniveaus
 - Und/oder Varianzen der Prädiktoren
- **β -Koeffizienten** (= Standardisierte Regressionskoeffizienten)
 - Wie b , nur besser interpretierbar, da Varianz und Einheit konstant
- Beide Formen können nur innerhalb dieses Modells im Kontext mit den anderen Prädiktoren interpretiert werden, da andere Mit-Prädiktoren den Wert verändern würden

- **Prädiktionsgüte**

- **Residuum/Vorhersagefehler**

- $Y_{res(i)} = Y_i - Y_{reg(i)}$
 - D.h. Abweichung der vorhergesagten von den tatsächlichen Y-Werten

- **Standardschätzfehler $S_{y.x}$**

- = Streuung der tatsächlichen y -Werte um die Regressionsgerade
 - = insgesamter Fehler, der bei Vorhersage entsteht

$$S_{y.x} = s_y \cdot \sqrt{1 - r_{xy}^2}$$

- **Validierung** (Validität: Gültigkeit des Modells; Bezug zu tatsächlichen Kausalzusammenhängen)

- *Übereinstimmung mit Theorie*
- *Übereinstimmung mit bisherigen Ergebnissen*
- Modellüberprüfung mit Hilfe von Strukturgleichungsmodellen
- Kreuzvalidierung
 - (= Test der Generalisierbarkeit der Ergebnisse auf eine andere Stichprobe)
 - In jedem Fall: Verfahren nochmal „über Kreuz“ in andere Richtung
- **Geteilte Stichproben**
 - Systematisch oder zufällig geteilt
 - Berechnung einer Regressionsgleichung in einer Hälfte
 - Anwendung der Gleichung zur Vorhersage in zweiter Hälfte
 - Vergleich der vorhergesagten mit den tatsächlichen Werten in zweiter Hälfte
 - je ähnlicher, desto valider Modell; Signifikanztests der Unterschiede der beiden Modelle (z.B. mit AMOS)
 - Unterschiede der Regressionskoeffizienten?
 - Unterschiede der Standardfehler der Regressionskoeffizienten?
 - Unterschiede zwischen Korrelationskoeffizienten r_{xy} (in 1. Stichprobe) und $r_{\hat{y}y}$ (in 2. Stichprobe) (z.B. mit AMOS)
 - $r_{\hat{y}y}$ kann eigentlich nur maximal so groß werden wie r_{xy} und muss vor diesem Hintergrund interpretiert werden (in derselben Stichprobe sind die beiden Korrelationskoeffizienten identisch)
 - Meist $r_{xy} \geq r_{\hat{y}y}$, da Vorhersage in zweiter Stichprobe nicht besser sein wird als in erster (das wäre ein sehr krasser Zufall)
 - Validere Gleichung bestimmen

- Neue Stichprobe
 - [Wie bei geteilten Stichproben]
 - Vorhersagen
 - PRESS-Statistik
- **Capitalization of Chance**
 - Nur bei a posteriori-Modellspezifikation
 - Bei Schätzung einer Populationskorrelation durch Bestimmung einer Stichprobenkorrelation wird immer Populationskorrelation überschätzt (**biased estimate**), weil die Standardfehler bei Multikollinearität unterschätzt werden; deshalb werden eventuell auch Prädiktoren signifikant, die eigentlich nicht viel beitragen. Und natürlich werden gerade die Prädiktoren, die in der Stichprobe die höchste Korrelation mit dem Kriterium haben (also die überschätztesten), in der Regression bevorzugt.
 - Außerdem kommt es durch die vielen Tests ohne eingebaute Korrektur des Niveaus zu einer α -Fehler-Inflation.
 - Dieser Effekt ist umso größer, je ...
 - Größer die Multikollinearität (weil der Anteil, um den sich die Prädiktoren überschneiden, zweimal eingeht)
 - Kleiner die Stichprobe
 - Größer die Zahl der Prädiktoren („**exponentiell**“)
 - Gleichung von Gauß: Anzahl der berechneten Korrelationskoeffizienten bei k Prädiktoren

$$\sum_{i=1}^k i = \frac{k \cdot (k + 1)}{2}$$

=> Anzahl zu schätzender Korrelationskoeffizienten nimmt mit Anzahl der Prädiktoren quadratisch zu, deshalb auch Anzahl der möglicherweise fehlerhaften Schätzungen

=> „Rumprobieren“ mit vielen Variablen pro Versuchsperson führt immer zu viel Fehler
 - Lösungsansätze
 - Ausreichendes N
 - Unkorrelierte Prädiktoren
 - Nur relevante Prädiktoren
 - Kreuzvalidierung
- **Multikollinearität**
 - *Problematische Effekte*
 - *Maximal mögliches R^2 verringert*
 - *Konfundiert Schätzung der Regressionskoeffizienten sehr stark*
 - *Macht Signifikanztestung schwerer (d.h. größerer Beta-Fehler), weil Standardfehler vergrößert*
 - *Kann Werte drastisch verändern, sogar bis hin zu Vorzeichenumkehrung*
 - Ausnahme: Das kann gut sein, wenn dadurch ein **Suppression Effect** aufgedeckt wird
 - d.h. Prädiktor korreliert nicht mit Kriterium, leistet aber wertvollen Beitrag, nachdem anderer, mit diesem korrelierter Prädiktor hinzugefügt und dem einzigartigen Einfluss dieses Prädiktors gegenläufiger geteilter Einfluss herausgefiltert wurde

- Bei empirisch basierten Verfahren zur Selektion des besten Modells wird oft nur eine der Variablen mitaufgenommen -> ausgeschlossene Variablen trotzdem mit Kriterium korrelieren und nicht einfach folgern, dass sie keine Vorhersagerelevanz haben! (praktische Signifikanz)
 - Umgangsmöglichkeiten
 - Faktorwerte
 - Nur eine nehmen
 - Problem: Specification Error
 - Nur R^2 interpretieren; statt Regressionskoeffizienten bivariate Korrelationen interpretieren
 - Spezialfall **Singularität**: Eine Variable perfekt aus den anderen Vorhersagbar
 - Dummy-Transformation ohne Weglassen einer Kategorie als Referenzkategorie
 - Faktorwerte zusammen mit den Originalvariablen
 - Macht Ermittlung der Kennwerte unmöglich, weil 0 im Nenner
 - Kennwerte
 - **Variance Inflation Factor**
 - **Toleranz**
 - Sollte mindestens 0.2 sein
- Anwendung mit SPSS
 - „Lineare Regression“ genauso wie bei ELR
 - R-Quadrat = R^2
 - Korrigiertes R-Quadrat = Adjusted R^2
 - „Beta“: Beta-Koeffizient
 - Tabelle „ANOVA“: Signifikanztest des Gesamtmodells durch Varianzanalyse (erklärbare vs. nicht erklärbare Varianz)
 - Methoden:
 - Einschluss = Researcher Specified
 - Schrittweise = Stepwise
 - Entfernen = vom Wissenschaftler spezifizierte Variablenblöcke können entfernt werden
 - Rückwärts, Vorwärts wie oben
 - Toleranz und partielle Korrelation der ausgeschlossenen Variablen automatisch ausgegeben

Logistische Regression

- **Ziel**
 - Binäres Merkmal mittels eines oder mehrerer intervallskaliert(er) (oder nichtmetrischer) Prädiktoren vorhersagen
 - Gute Prädiktion: möglichst wenige falsche pro richtigen Zuordnungen
 - **Anpassungsmöglichkeiten**
 - Es gibt auch logistische Verfahren für mehr als zwei Gruppen
 - Bei vielen Gruppen (5,6,7) manchmal schwieriger als linear, weil sie für jede Gruppe bis auf eine ein Modell „Zugehörigkeit zu Gruppe X – Ja/nein“ sucht -> viele Modelle + relativ kleine Anteile der jeweiligen Gruppen
 - Dann eher Diskriminanzanalyse
 - Man kann auch Interaktionseffekte prüfen, aber nicht mit SPSS (müssen da von Hand modelliert werden)
 - **Beziehung zu anderen Analysen**
 - **Unterschiede zur Diskriminanzanalyse**
 - Kann in Standardversion nur exakt 2 Gruppen der AV
 - Kann nicht-metrische Prädiktoren durch Dummy-Kodierung aufnehmen
 - braucht größere Stichprobe (mind. 400, dabei mind. 20 pro Prädiktor und pro Untergruppe)
 - effizienter als DA, wenn Voraussetzungen der DA verletzt, da robuster gegen
 - Nicht-Normalverteilte UVs
 - Ungleiche Varianz-Kovarianz-Matrizen
 - Berechnet direkt die Wahrscheinlichkeit der Gruppenzugehörigkeit als metrische Variable
 - **Unterschiede zur (multiplen) linearen Regression**
 - Logistischer, kein linearer, Zusammenhang
 - Je steiler, desto besser Prädiktion im Übergangsbereich
 - Funktion erreicht weder 1 noch 0, nur approximative Annäherung
 - Verwandtschaft -> verwandte Konzepte und Interpretation, aber andere Umsetzung
- | Multiple Regression | Logistische Regression |
|--------------------------|-----------------------------------|
| SS_{total} | -2LL des Nullmodells |
| $SS_{residuum}$ | -2LL des ermittelten Modells |
| $SS_{regression}$ | Differenz im -2LL beider Modelle |
| F-Test für Modellpassung | χ^2 -Test für -2LL-Differenz |
| R^2 | pseudo R^2 |
- Lineare Regression: Least Squares; Logistische Regression; Maximum Likelihood
 - Nullmodell
 - In ELR: Mittelwert des Kriteriums als Prädiktor
 - In LR: mehrere Kriterien
 - **Maximum Chance:** Zuordnung aller Fälle zu größter Gruppe
 - **Proportional Chance:** Versucht, alle Gruppen richtig zu treffen
 - Berechnet direkt Wahrscheinlichkeit als metrische Variable
 - Interpretation
 - Je nach Lage auf Skala der UV hat eine UV-Änderung desselben Umfangs unterschiedlich große Wirkungen auf Änderung der Wahrscheinlichkeitsfunktion
- **Wann gute Prädiktion mithilfe eines Prädiktors?**
 - Große Mittelwertunterschiede
 - Kleine Streuung innerhalb der Gruppen
 - *Modell immer nur angepasst an Daten im UV-Bereich, wo viele Daten liegen – Funktion verläuft vielleicht auch außerhalb dieses Bereichs, wo gar keine Daten liegen können, aber das macht nichts (wie lineare Regression auch)*

- Gleichung mit einem bzw. mit mehreren Prädiktoren

$$\pi_j = \frac{e^{b_0 + b \cdot x_j}}{1 + e^{b_0 + b \cdot x_j}}$$

$$\pi_j = \frac{e^{b_0 + b_1 \cdot x_{1j} + b_2 \cdot x_{2j} + \dots}}{1 + e^{b_0 + b_1 \cdot x_{1j} + b_2 \cdot x_{2j} + \dots}}$$

-
- π_j : Wahrscheinlichkeit der Zugehörigkeit zu der Gruppe mit Kodierung 1
- D.h. wir versuchen, die Kurve so reinzulegen, dass der Vorhersagefehler minimal/die Trefferquote maximal ist
- D.h. sodass an ihren Enden jeweils nur noch Personen sind, die in der jeweils zugehörigen Gruppe sind (in der Mitte gibt es immer einen Bereich, in dem man manche Leute falsch zuordnet)

- Voraussetzungen

- AV
 - Nominalskaliert
 - (zwei Kategorien)
 - Eindeutige Gruppenzugehörigkeit
- UVs
 - Intervallskaliert
 - Muss nicht unbedingt, aber man braucht auf jeden Fall auch intervallskalierte; nominalskaliert nur zur Unterstützung.
 - normalverteilt
- $N > 400$
 - Da iterative maximum likelihood Schätzung als Methode -> braucht viele Fälle, damit Iteration nicht in lokales Minimum läuft

- Vorgehen

- Auswahl der Prädiktoren
 - Theoriegeleitet
 - Schrittweise nach Bedeutung
- Anpassung einer Kurve an empirische Werte
- Transformationen
 - 1. Wahrscheinlichkeiten zu Verhältnissen (odds)
 - p zwischen 0 und 1 (einschließlich)

$$Odds_j = \left(\frac{prob_{event}}{1 - prob_{event}} \right)$$

$$= e^{b_0 + b_1 \cdot x_{1j} + b_2 \cdot x_{2j} + \dots}$$

- - $p = 0.5$ -> odds = 1
 - $p = 1$ -> ERROR
 - $p = 0$ -> ERROR
- Odds zwischen 0 und 1 (ausschließlich)

- 2. Odds zu log odds logarithmieren

$$Logit_j = \ln \left(\frac{prob_{event}}{1 - prob_{event}} \right)$$

$$= b_0 + b_1 \cdot x_{1j} + b_2 \cdot x_{2j} + \dots$$

-
- Odds < 1 -> log odds < 0
- Odds > 1 -> log odds > 0
- Odds = 1 -> log odds = 0

- Jetzt suchen wir nur noch eine lineare Funktion, für die wir eine Steigung und eine Konstante schätzen

- **Schätzung der Parameter**
 - Maximum Likelihood Schätzung
 - = iterative Ermittlung der Parameter (b₀ bis b_p), welche Wahrscheinlichkeit für korrekte Vorhersage maximieren, über Modellvergleich
 - 1. Null-Modell: Wie gut Vorhersage ohne Prädiktor?
 - 2. Vorgeschlagenes Modell: Wie gut Vorhersage beim ermittelten Modell (mit Prädiktor)?
 - 3. Modellbewertung über -2LL-Differenz zwischen diesen beiden Modellen
 - **Kennwerte für Bewertung**
 - **Qualität der Prädiktoren**
 - **-2 * log likelihood**
 - Kleiner Wert = guter Wert (Minimum 0)
 - Können wir direkt gar nicht beurteilen, da nicht normiert, muss nur kleiner werden als beim Nullmodell
 - **Erklärbare Varianzanteile**
 - Im Allgemeinen alle Kennwerte angeben, aber AUF JEDEN FALL IMMER dazuschreiben, welches R² man angibt! Das macht einen Unterschied!
 - Sollen groß werden
 - Benutzt man, wenn man angeben will, „wie viel“ besser das Modell geworden ist (normiert, damit kann man was anfangen)
 - **Kennwerte**
 - „pseudo“ R²
 - Kann 1 erreichen
 - Methode der Wahl
- **Validierung an der Trefferquote**
- **Anwendung in SPSS**
 - Input
 - Regression -> binär logistisch
 - Methode: Vorwärts bedingt
 - „Kovariaten“ = UVs/Prädiktoren (*Übersetzung aus dem Englischen in SPSS nicht immer optimal, so auch hier*)
 - Codierung: Die Gruppe, die vorhergesagt werden soll, bekommt die 1
 - SPSS beendet iterative Schätzung von selbst, wenn keine substanziellen Änderungen mehr
 - Output
 - „Klassifizierungstabelle“ enthält Trefferquoten; je besser Trefferquote, desto besser Modell
 - Auch restliche Hauptdiagonale anschauen, falls nur eine Gruppe gut vorhergesagt
 - „Variablen in der Gleichung“: Signifikanzprüfung des Prädiktors
 - $\text{Exp}(B) = e^b$
 - Entspricht odds ratio bei binären Prädiktoren, bei intervallskalierten nicht; das ist relevant, weil Mediziner mit odds ratio umgehen können, weshalb man es berichtet, wenn man kann
 - „Modellieren, wenn Term entfernt“ = „Wie ändert sich die Signifikanz des Modells, wenn ich auf diesen Prädiktor verzichte?“

$$R_{\text{logit}}^2 = \frac{-2LL_{\text{null}} - (-2LL_{\text{model}})}{-2LL_{\text{null}}}$$

- „Bedingt vorwärts“: Schrittweise je 1 Prädiktor mehr hinzugenommen
 - Reines Vorwärtsverfahren, kein Rauswurf möglich
 - Deshalb prüfen: Wenn ein Prädiktor im Laufe des Fortschreitens des Modells an Wert verliert (was bei (Multi-)Kollinearität passieren kann) ... manuell rauswerfen, falls nicht mehr signifikant
 - Erfahrungswert: frühestens bei 6, 7 Prädiktoren -> 1. Oder 2. („dickste Fische“) später nicht mehr gebraucht
- Warnung: Die Daten in diesem Beispiel sind zu gut, so was hat man normalerweise nicht

Multiple Varianzanalyse (MANOVA)

- **Ziel**

- Varianzanalyse zur Entdeckung von Mittelwertunterschieden zwischen vorgegebenen Gruppen
- Einfluss eines oder mehrerer Faktoren
- Interaktion zwischen Faktoren
- Ermittlung der Effektgrößen

- **Beziehung zu anderen Verfahren**

- **ANOVA**

- Weitestgehend Analog
- Erweiterung für mehrere AVs
- Besonders sinnvoll, wenn Multikollinearität der AVs
- Kontrolliert Alpha-Fehler-Inflation
- Besonders sinnvoll, wenn Hypothese, dass Unterschiede bei einer Kombination der AVs auftreten (die bei univariaten Analysen nicht entdeckt werden)
- **Gleichung**

ANOVA

$$Y_1 = X_1 + X_2 + \dots + X_n$$

MANOVA

$$Y_1 + Y_2 + Y_3 + \dots + Y_n = X_1 + X_2 + \dots + X_n$$

-

- X sind die unabhängigen, Y die abhängigen Variablen

- **Nullhypothese**

ANOVA

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$$

MANOVA

$$H_0 : \begin{pmatrix} \mu_{11} \\ \mu_{21} \\ \vdots \\ \mu_{p1} \end{pmatrix} = \begin{pmatrix} \mu_{12} \\ \mu_{22} \\ \vdots \\ \mu_{p2} \end{pmatrix} = \dots = \begin{pmatrix} \mu_{1k} \\ \mu_{2k} \\ \vdots \\ \mu_{pk} \end{pmatrix}$$

-

- D.h. (MANOVA): Vektoren der Mittelwerte in allen Gruppen gleich

- **Hotelling's T²**

- Zwei Gruppen: t-Test bei 1 AV -> Hotelling's T² bei 2+ AVs

- Kennwertberechnung:

$$C = W_1 \cdot X_1 + W_2 \cdot X_2 + \dots + W_n \cdot X_n$$

-

- Gewichte W so gewählt, dass Mittelwertunterschied zwischen Gruppen maximal ist

- = Diskriminanzfunktion wie bei DA

- X1: Mittelwerte der Gruppen

- Einmalige Hypothesenprüfung: T² muss über T²_{krit} liegen -> nur ein Test, gegen α -Fehler-Inflation

$$T_{krit}^2 = \frac{p \cdot (N_1 + N_2 - 2)}{N_1 + N_2 - p - 1} \cdot F_{krit}$$

-

- Mehr als zwei Gruppen: ANOVA bei 1 AV -> MANOVA bei 2+ AVs
 - Wie Hotelling's T² mit mehreren Diskriminanzfunktionen

- Hypothesenprüfung der Funktionen (mehrere Möglichkeiten)
 - **Diskriminanzanalyse**
 - Diskriminanzanalyse:
 - AV
 - Nominalskaliert
 - eindimensional (d.h. keine Systematik der Gruppen)
 - UV intervallskaliert
 - Mehrere Diskriminanzfunktionen
 - MANOVA
 - AV intervallskaliert
 - UV
 - Feste Effekte: nominalskaliert
 - Zufällige Effekte: intervallskaliert möglich
 - Mehrere Diskriminanzfunktionen
 - Kann Interaktionseffekte, da Gruppen nicht selbst die UV, sondern entstehen durch definierte andere UV(s) (d.h. Gruppen können „geordnet“ werden)
 - Kann Kovariaten -> gut für Power
 - Kann Messwiederholungs-Designs -> dabei werden Personeneigenschaften innerhalb der Personen konstant gehalten -> teststärker als ohne MW
- **Voraussetzungen (F-Test allerdings recht robust, besonders bei ähnlich großen Gruppen)**
 - Multivariate Normalverteilung
 - Unabhängige Messungen
 - Homogene Varianz-Kovarianz-Matrizen in Gruppen (Box' M-Test), analog zur Varianzhomogenität beim t-Test
 - Lineare Zusammenhänge
 - Keine Ausreißer
 - Mindestens N = 20 pro Zelle
 - Abhängig von Anzahl AVs
 - Mit G-Power gegenrechnen
 - Anzahl der möglichen Kovariaten auch von N abhängig bzw. andersrum
 - Ähnlich/gleich stark besetzte Zellen (dann Abweichungen von Multivariater Normalverteilung nicht schlimm)
- **Signifikanzprüfung**
 - Power und Signifikanzniveau immer abhängig von
 - N
 - Anzahl Gruppen
 - Effektgrößen
 - Anzahl AVs
 - Je mehr, desto schlechtere Power
 - **F-Test in ANOVA**
 - Quadratsummenzerlegung: $SS_{\text{total}} = SS_{\text{between}} + SS_{\text{within}}$
 - SS_{between} = erklärbare Varianz
 - $(MS_{\text{between}}/MS_{\text{within}})$ wird auf Signifikanz geprüft
 - SS_{within} = nicht erklärbare Varianz
 - **Signifikanztests für MANOVA** (immer nur einen Kennwert angeben; bei 2 Gruppen alle Kennwerte identisch, je mehr Gruppen, desto größere Unterschiede)
 - **Größte charakteristische Wurzel nach Roy (gcw; engl. gcr)**

$$Roy's\ gcw = \frac{(k - 1) \cdot F_{krit}}{N - k}$$

- Empirisches gcw muss über kritischem gcw liegen
- Vorteil: Bessere Power als andere Tests, wenn alle Voraussetzungen streng eingehalten
- Nachteil: Berücksichtigt nur erste Diskriminanzfunktion -> nur wirklich sinnvoll, wenn entweder
 - Nur 2 Gruppen
 - Nur Unterschiede auf einer Dimension relevant
- **Wilks's Lambda (U-Statistik)**
 - Berücksichtigt alle Diskriminanzfunktionen
 - Gut, wenn alle Voraussetzungen streng eingehalten
 - Selten verwendet
- **Pillai's trace**
 - Bester, am häufigsten eingesetzter Kennwert: stabil gegenüber Voraussetzungsverletzung
 - Berücksichtigt alle Diskriminanzfunktionen
- **Hotelling's T²**
- **Effekte**
 - **Effektgröße** (d.h. Maße für erklärbare Varianzanteile): **Partielles η^2**
 - wichtig: NICHT η^2 ausgeben!
 - Partielles η^2 und η^2 identisch bei nur einem Faktor; bei mehr als einem η^2 nicht sinnvoll
 - Partielles η^2 immer höher als η^2 , weil
 - Partiiell: $SS_{\text{Faktor}} / (SS_{\text{Faktor}} + SS_{\text{error}})$
 - η^2 : $SS_{\text{Faktor}} / SS_{\text{total}}$
 - $SS_{\text{total}} = SS_A + SS_{\text{ErrorA}} + SS_B + SS_{\text{ErrorB}}$
 - **Feste vs. zufällige Effekte**
 - **Entscheidungs**
 - Effekttyp im Vorfeld für das Gesamtmodell festgelegt
 - Modell I: Alle UVs zufällige Faktoren
 - Modell II: Alle UVs feste Faktoren
 - Modell III: Gemischt
 - Mit Restriktion
 - Ohne Restriktion
 - Entscheidung hängt am theoretischen Modell: Sind UVs intervall- oder nominalskaliert?
 - Im Zweifel einfach feste Effekte, da
 - Noch keine Einigung, ob bei gemischten Effekten Restriktionen für Interaktionseffekte gelten müssen
 - Modelle mit zufälligen und gemischten Effekten noch kontrovers
 - Bei zufälligen Effekten Kontrolle über Auswahl aus der Hand gibt
 - Evtl. bekommt man keine Normalverteilung, was dann?
 - *Leonhart findet es nicht gut, dann einfach doch mit festen Effekten zu rechnen, weil bei künstlichem Split (z.B. Median) Information verloren geht/weil die Trennung nicht gut replizierbar ist*
 - Ausnahme: Wenn klare Höcker-Verteilung; dann Split sinnvoll
 - ANOVA mit zufälligen Effekten gibt dieselben Ergebnisse wie Regressionsanalyse mit Interaktionseffekten
 - **Prüfvarianzen bei zwei Faktoren (ANOVA?)**

	Faktor A	Faktor B	Faktor AxB
A fest, B fest (Modell I)	$F = \frac{MS_{FaktorA}}{MS_{within}}$	$F = \frac{MS_{FaktorB}}{MS_{within}}$	$F = \frac{MS_{FaktorAxB}}{MS_{within}}$
A zufällig, B zufällig (Modell II)	$F = \frac{MS_{FaktorA}}{MS_{FaktorAxB}}$	$F = \frac{MS_{FaktorB}}{MS_{FaktorAxB}}$	$F = \frac{MS_{FaktorAxB}}{MS_{within}}$
A fest, B zufällig, mit der üblichen Restriktion (Modell III)	$F = \frac{MS_{FaktorA}}{MS_{FaktorAxB}}$	$F = \frac{MS_{FaktorB}}{MS_{within}}$	$F = \frac{MS_{FaktorAxB}}{MS_{within}}$
A fest, B zufällig, ohne die übliche Restriktion (Modell III)	$F = \frac{MS_{FaktorA}}{MS_{FaktorAxB}}$	$F = \frac{MS_{FaktorB}}{MS_{FaktorAxB}}$	$F = \frac{MS_{FaktorAxB}}{MS_{within}}$

	Feste Effekte	Zufällige Effekte
Wann?	<ul style="list-style-type: none"> Alle möglichen Ausprägungen der UV sind als Stufen im Design realisiert ODER es wird kein Schluss auf nicht realisierte Ausprägungen (z.B. Zwischenstufen) gezogen 	<ul style="list-style-type: none"> UV hat theoretisch unendlich viele Abstufungen UND Realisation einzelner Abstufungen nicht von Interesse -> zufällige Auswahl einiger Faktorstufen Gut, wenn man z.B. a priori nicht weiß, was eine „hohe“ und was eine „niedrige“ Ausprägung ist, um Gruppen vorzudefinieren
Schluss	Mittelwertunterschiede in Stichprobe -> Mittelwertunterschiede in Population	Varianz der Effekte in Stichprobe -> Varianz der Effekte in Population
Mögliche Aussagen	<ul style="list-style-type: none"> Nur über realisierte Ausprägungen Kein „je ... desto ...“ 	<ul style="list-style-type: none"> „je ... desto ...“ Aussagen generalisierbar auf alle möglichen Ausprägungen
Strukturgleichung	$y_{ij} = \mu_y + \alpha_j + \epsilon_{ij}$ mit α_j als Effekt der Gruppenzugehörigkeit zur festen Faktorstufe j	$y_{ij} = \mu_y + \alpha_j + \epsilon_{ij}$ mit α_j als Populationseffekt der Faktorstufe j des zufälligen Effekts
Nullhypothese	Nullhypothese: $H_0 : \alpha_j = 0$ Die Effekte der Faktorstufen sind null.	Nullhypothese: $H_0 : \sigma_A^2 = \sigma_\alpha^2 = 0$ Die Varianz der zufälligen Effekte ist null. Es gibt keine Streuung der Effekte.
Erwartungswerte der mittleren Quadratsummen	$E(MS_{Faktor A}) = \sigma_e^2 + \frac{q \cdot n \cdot \sum_{j=1}^p \alpha_j^2}{p-1}$ $E(MS_{Faktor B}) = \sigma_e^2 + \frac{p \cdot n \cdot \sum_{k=1}^q \beta_k^2}{q-1}$ $E(MS_{Faktor AxB}) = \sigma_e^2 + \frac{n \cdot \sum_{j=1}^p \sum_{k=1}^q (\alpha\beta)_{jk}^2}{(p-1) \cdot (q-1)}$ $E(MS_{within}) = \sigma_e^2$	<ul style="list-style-type: none"> Faktor A: $E(MS_{FaktorA}) = \sigma_e^2 + n \cdot \sigma_{AxB}^2 + q \cdot n \cdot \sigma_A^2$ <ul style="list-style-type: none"> Problem: Da ist die Interaktionsvarianz auch mit drin! D.h. der F-Test wird auch signifikant, wenn Interaktion signifikanter Einfluss Deshalb macht man <u>bei zufälligen Effekten andere Signifikanztests</u>

		<ul style="list-style-type: none"> ○ Faktor B: $E(MS_{\text{FaktorB}}) = \sigma_e^2 + n \cdot \sigma_{A \times B}^2 + p \cdot n \cdot \sigma_B^2$ ○ Interaktion: $E(MS_{\text{FaktorAxB}}) = \sigma_e^2 + n \cdot \sigma_{A \times B}^2$ ○ Fehlervarianz: $E(MS_{\text{within}}) = \sigma_e^2$
Annahmen	<ul style="list-style-type: none"> ○ Zellen des ersten Faktors gleich besetzt (z.B. bei Messwiederholung) Restriktion ○ Summe der Effekte der Stufen eines Faktors = 0 <ul style="list-style-type: none"> ○ Deshalb Erwartungswert der Interaktionseffekte = 0, weil die durch Multiplizieren der Effekte der Faktoren entstehen ○ Deshalb enthält der Haupteffekt keinen Anteil des Interaktionseffekts -> einfach interpretierbar 	<ul style="list-style-type: none"> ○ Bei den theoretisch unendlich vielen Realisierungen des Treatments wird davon ausgegangen, dass in der Population eine Verteilung von α_j-Werten (d.h. Effekten) mit einem Mittelwert $\mu_\alpha = 0$ vorliegt ○ Aber hier kann nicht davon ausgegangen werden, dass in der Stichprobe die Summe der Effekte 0 ist (in der Population schon!), weil sich der Stichprobenmittelwert durch die zufällige Ziehung vom Populationsmittelwert unterscheidet
Voraussetzungen		<ul style="list-style-type: none"> • Mindestens Intervallskalenniveau der AV • Normalverteilung der AV • Zufallseffekte voneinander unabhängig • Zufallseffekte identisch verteilt • Mögliche Effekte a_j werden durch Zufallsvariable mit Mittelwert 0 und Varianz σ_A^2 realisiert

- **Restriktion bezüglich Interaktion**

- Bei festen Effekten: Immer Restriktion

- Summen der Interaktionseffekte zeilen- und spaltenweise immer = 0

- D.h. $\sum ab_{ij} = ab_{.j} = 0$

- für alle j

- Deshalb wird der Term für den Haupteffekt eines Faktors, der sich auf die Interaktion bezieht, gleich 0, und damit sind die Haupteffekte unabhängig von der Interaktion interpretierbar

- Daraus folgen Abhängigkeiten für die Interaktionen (#35; aber die Interaktionseffekt-Formel ist doch immer dieselbe, vgl. #34; was sich ändert ist der Interaktions-Term in der Formel für die Haupteffekte? -> Antwort Leonhart: *Durch die unterschiedlichen Voraussetzungen ergeben sich für die Erwartungswerte der verschiedenen Quadratsummen (Faktor A, Faktor B und Interaktionseffekt) unterschiedliche Werte. Diese führen dann zu unterschiedlichen F-Tests (mehr bei Werner, 1997 oder Leonhart, 2013).*)

- Bei gemischten Effekten: Ausschuen

- Mit Restriktion: Wie bei festen Effekten

- Restriktion gilt für den zufälligen Effekt (d.h. Haupteffekt des zufälligen Effekts hat keinen Interaktionsanteil)

- Alternatives Modell ohne Restriktionen

- Löst Abhängigkeiten auf
 - Andere Erwartungswerte, Prüfsummen und Hypothesen
 - Mit Modellspezifikation (MW vs. kein MW; feste bzw. zufällige Effekte) festgelegt, ob mit oder ohne Restriktion
 - Ich glaube, das liegt daran, dass man für Restriktionen gleiche Zellenbesetzungen auf den Faktorstufen des ersten (d.h. des festen) Faktors braucht und die meist nur bei MW-Designs hat.
- SPSS
 - MW -> mit Restriktion
 - Keine MW -> keine Restriktion
- Anwendung in SPSS
 - „ALM“->“Multivariat“
 - „Feste Faktoren“: SPSS kann keine zufälligen oder gemischten Effekte rechnen (-> andere Programme, z.B. Stata)
 - #46: Hypothese $df = df_{\text{between}}$ und Fehler $df = df_{\text{within}}$?
 - Aber das müssten doch dann 4 Gruppen sein, um von $N = 217$ auf $df = N - p = 213$ und von 2 Gruppen auf $df = p - 1 = 3$ zu kommen ... aber hier haben wir nur 2 Gruppen (Kontroll- und Experimentalgruppe)?
 - Deskriptiv
 - Leonhart schaut „nach Gefühl“ ob Standardabweichungen ähnlich sind
 - „Multivariate Tests“: Omnibustest
 - Zweite große Zeile! Nicht „konstanter Term“!
 - Erstmal nur „es gibt einen Gruppenunterschied“, wenn man alle AVs berücksichtigt
 - „Tests der Zwischensubjekteffekte“ braucht man eigentlich nur „korrigierte Gesamtvariation“ und „Fehler“ und „gruppe“
 - Sagt mir, auf welcher AV die Gruppenunterschiede liegen
 - Dieser ganze Output: Gruppenunterschiede zu t0, das heißt, wir wollten auch keine signifikanten Ergebnisse -> yay, Randomisierung hat funktioniert
 - Wenn hier schon signifikanter Mittelwertunterschied auf einer AV → erster Messzeitpunkt als Kovariate (siehe MANCOVA)
- Leonhart: Es passiert nicht, dass bei Overall-Test signifikantes Ergebnis, aber keine signifikanten Unterergebnisse (oder umgekehrt).

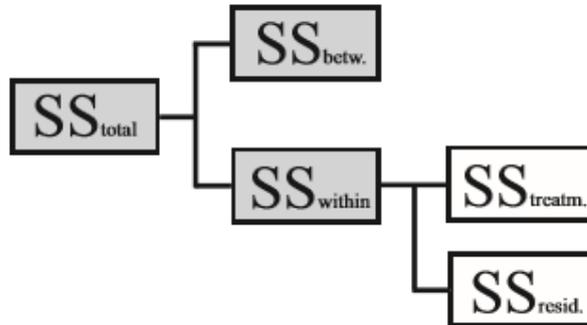
MANOVA mit Messwiederholung

- Immer multivariat, da mehrere Messungen der AV pro Person
 - Geht ein- und multifaktoriell
 - Einfaktoriell (4) mit MW: eigentlich 2 Faktoren
 - Personenfaktor: Unterschiede zwischen Personen, aus methodischer und inhaltlicher Sicht irrelevant
 - Messwiederholungsfaktor: Unterschiede zwischen Messzeitpunkten/Experimentalbedingungen
- Overall-Test wie bei MANOVA
- **Voraussetzungen**
 - Intervallskalierte AV
 - Normalverteilte AV
 - Keine fehlenden Werte
 - Unabhängigkeit der Fehlervarianzen (wie bei MANOVA nötig) nicht möglich -> stattdessen Sphärizität vorausgesetzt
 - Sphärizität (Zirkularität): Kovarianzen müssen zu verschiedenen Messzeitpunkten ähnlich sein
- **Probleme**
 - 1. Keine Unabhängigkeit der Fehlervarianzen der einzelnen Faktorstufen (d.h. wir haben korrelierte Fehler) -> Korrelationen zwischen Messwerten verschiedener Messzeitpunkte -> *erhöht alpha-Fehler-Wahrscheinlichkeit durch Signifikanzüberschätzung im F-Test*
 - Also wird Sphärizität vorausgesetzt
 - Wenn nicht erfüllt: Korrektur nach Greenhouse und Geisser
 - 2. Einfluss des Messzeitpunkts als Lern- oder Ermüdungseffekte (carry-over-Effekte) je nach Motivation (hohe Motivation: lernen) -> gefährden interne Validität
 - 3. Je mehr Messzeitpunkte, desto wahrscheinlicher fehlende Werte, aber bei ANOVA mit MW NUR Subjekte mit vollständigen Daten aufgenommen -> Multiple Imputation oder keine fehlenden Werte haben
- **Sphärizität**
 - Definition
 - Varianz der Differenzwerte (d.h. der Abweichungen der tatsächlichen von den vorhergesagten Werten) innerhalb des Messwiederholungsdesigns über alle Gruppen hinweg gleich
 - D.h. fehlerbedingte Kovarianz zwischen t1 und t2 gleich fehlerbedingter Kovarianz zwischen t2 und t3 usw. (Problem: Das passiert bestimmt super oft, z.B. bei Messwiederholung mit unterschiedlich weit auseinanderliegenden Messzeitpunkten?)
 - -> sonst erhöhte alpha-Fehler-Gefahr durch Signifikanzüberschätzung
 - Test: Mauchly-Test
 - χ^2 -Test auf bedeutsamen Unterschied der Fehlerkovarianzmatrix von der Einheitsmatrix
 - Problem
 - Bei kleinen Stichproben werden Verletzungen nicht entdeckt (klar, da wird nichts signifikant)
 - Bei großen Stichproben ergibt die Prüfung auch bei nur marginalen Verletzungen ein signifikantes Erlebnis
 - Berichten, aber „nicht mehr so ernst nehmen“
 - *Korrektur hier auch nicht mehr so krass*
 - Korrektur
 - Lower-Bound
 - Strengste Korrektur
 - Huynh-Feldt

- Schwächste Korrektur
- Greenhouse-Geisser
 - Bevorzugt verwenden, v.a. weil das die Reviewer kennen

- **Quadratsummenzerlegung**

- SS = Quadratsumme
- MS = Mittlere Quadratsumme (= Quadratsumme/df)
- Einfaktorielle ANOVA ohne MW



-
- Signifikanztest: $MS_{\text{between}}/MS_{\text{within}}$
- Einfaktorielle ANOVA mit MW
 - Wir machen dasselbe wie bei ANOVA, nur in Stufen
 - Wir können SS_{within} – bisher Fehlerquadratsumme – auf tatsächlichen Fehler und Treatment-Effekt (vorher vs. nachher) teilen (was ganz schön ist, weil dadurch die Fehlerquadratsumme schrumpft, wodurch höhere Power, weil F-Test $MS_{\text{treatment}}/MS_{\text{residual}}$ testet)

$$SS_{\text{between subjects}} = p \cdot \sum_{i=1}^n (\bar{p}_i - \bar{y})^2$$

$$SS_{\text{within subjects}} = \sum_{j=1}^p \sum_{i=1}^n (y_{ij} - \bar{p}_i)^2$$

$$SS_{\text{treatment}} = \sum_{j=1}^p n \cdot (\bar{y}_j - \bar{y})^2$$

$$SS_{\text{residual}} = \sum_{j=1}^p \sum_{i=1}^n (y_{ij} - \bar{p}_i - \bar{y}_j + \bar{y})^2$$

- - P = Messzeitpunkte; j = Index der Messzeitpunkte
 - N = Versuchspersonen; i = Index der Versuchspersonen

- Signifikanztest für Wiederholungsfaktor
 - $F = MS_{\text{treatment}}/MS_{\text{residual}}$
 - $df_{\text{treatment}} = p - 1$
 - $df_{\text{residual}} = (p - 1) \cdot (n - 1)$
 - wird bei Sphärizität direkt angewandt; sonst Korrektur

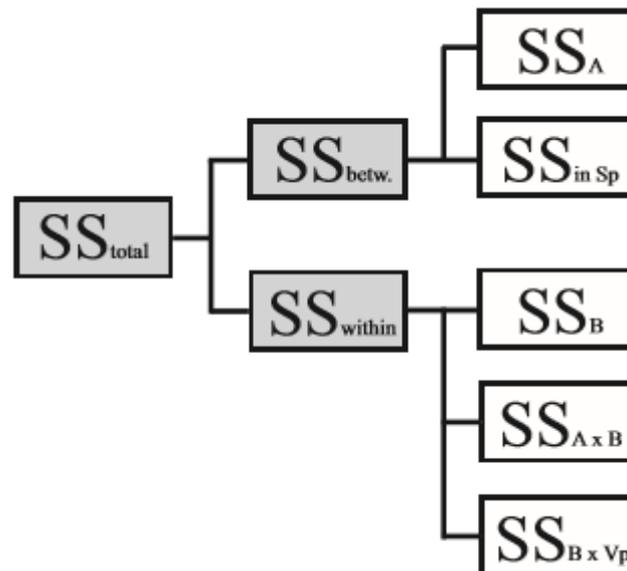
- Erklärte Varianzanteile
 - Effektgröße für MW-Faktor: partielles η^2

$$\eta_{P, \text{treatment}}^2 = \frac{SS_{\text{treatment}}}{SS_{\text{treatment}} + SS_{\text{residual}}}$$

-
- Hier benutzt man, im Gegensatz zur F-Statistik, Quadratsummen statt mittleren Quadratsummen, da MS durch die df die Stichprobengröße berücksichtigen; die Signifikanz hängt von der Stichprobengröße ab, aber die Effektgröße nicht.

- Zweifaktorielle ANOVA mit MW

- Anmerkung: Freiheitsgrade müssen wir nicht für die Klausur auswendig wissen
- Wir sollten wissen, dass N sowie Faktorstufen miteingehen -> wenn Freiheitsgrade schwanken zwischen Analysen an angeblich gleichen Stichproben, wird man skeptisch; vermutlich haben die für manche Analysen Leute ausgeschlossen und sonst nicht, was sie nicht angegeben haben
- Unvollständige MW
 - Ein Faktor mit, ein Faktor ohne MW



$$\begin{aligned}
 SS_{total} &= SS_{between\ subjects} + SS_{within\ subjects} \\
 &= SS_{in\ Sp} + SS_{Faktor\ A} \\
 &\quad + SS_{Faktor\ B} + SS_{Faktor\ AxB} + SS_{Faktor\ BxVpn}
 \end{aligned}$$

- Zwei Fehlerquadratsummen, zwei Prüfgrößen -> höhere Power
- **Signifikanztests**
 - Für Faktor A
 - $F = MS_{FaktorA} / MS_{inSp}$
 - $Df1 = p - 1$
 - $Df2 = p * (n - 1)$
 - Für Faktor B (hier MW-Faktor)
 - $MS_{FaktorB} / MS_{Faktor\ BxVpn}$
 - $Df1 = q - 1$
 - $Df2 = p * (q - 1) * (n - 1)$
 - **$P = \text{Stufen von Faktor A (!!)}$**
 - $Q = \text{Stufen von Faktor B}$
 - Interaktionsfaktor
 - $F = MS_{FaktorAxB} / MS_{Faktor\ BxVpn}$
 - $Df1 = (p - 1) * (q - 1)$
 - $Df2 = p * (q - 1) * (n - 1)$
- **Effekgröße: partielles η^2**
 - Achtung: partielle etas für beide Faktoren plus Interaktion jeweils einzeln; alle angeben, nicht addieren!

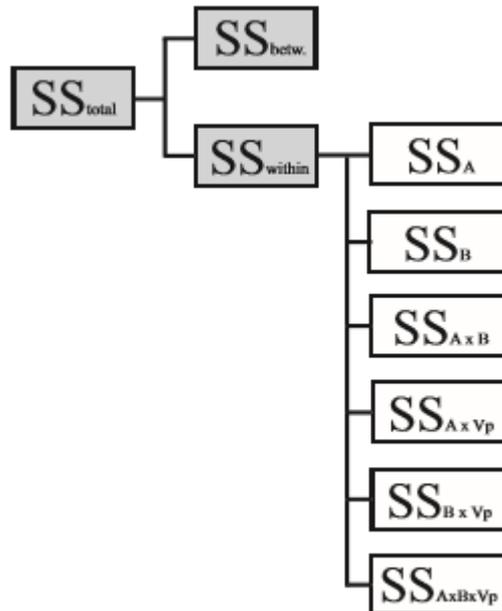
$$\eta_{P, \text{Faktor A}}^2 = \frac{SS_{\text{Faktor A}}}{SS_{\text{Faktor A}} + SS_{\text{in Sp}}}$$

$$\eta_{P, \text{Faktor B}}^2 = \frac{SS_{\text{Faktor B}}}{SS_{\text{Faktor B}} + SS_{\text{Faktor BxVpn}}}$$

$$\eta_{P, \text{Faktor Ax B}}^2 = \frac{SS_{\text{Faktor Ax B}}}{SS_{\text{Faktor Ax B}} + SS_{\text{Faktor BxVpn}}}$$

-
- Vollständige MW

- Beide Faktoren mit MW
- Noch weitere Reduktion der Fehlerquadratsumme (-> teststärkstes Verfahren); außerdem haben wir jetzt drei richtige Fehler:
 - $SS_{A \times Vp}$
 - $SS_{B \times Vp}$
 - $SS_{A \times B \times Vp}$



-
- **Signifikanztests**
 - Faktor A
 - $F = MS_A / MS_{A \times Vp}$
 - $df_A = p - 1$
 - $df_{A \times Vp} = (p - 1)(n - 1)$
 - Faktor B
 - $F = MS_B / MS_{B \times Vp}$
 - $Df_B = q - 1$
 - $Df_{B \times Vp} = (q - 1)(n - 1)$
 - Interaktionseffekt
 - $F = MS_{A \times B} / MS_{A \times B \times Vp}$
 - $Df_{A \times B} = (p - 1)(q - 1)$
 - $Df_{A \times B \times Vp} = (p - 1)(q - 1)(n - 1)$
- **Effektgrößen: partielles η^2**
 - Auch hier immer getrennt angeben

$$\eta_{P, \text{Faktor A}}^2 = \frac{SS_{\text{Faktor A}}}{SS_{\text{Faktor A}} + SS_{\text{Faktor AxVpn}}}$$
$$\eta_{P, \text{Faktor B}}^2 = \frac{SS_{\text{Faktor B}}}{SS_{\text{Faktor B}} + SS_{\text{Faktor BxVpn}}}$$
$$\eta_{P, \text{Faktor Ax B}}^2 = \frac{SS_{\text{Faktor Ax B}}}{SS_{\text{Faktor Ax B}} + SS_{\text{Faktor Ax BxVpn}}}$$

○

Kovarianzanalyse (MANCOVA)

- = MANOVA mit Kovariaten
- = Gleichzeitige Analyse von Varianz und Kovarianz (Kombination aus ANOVA und Regressionsanalyse)
- **Ziel**
 - Könnten Gruppenunterschiede noch auf andere Einflüsse zurückzuführen sein als UV?
 - Bräuchte man bei perfekter Randomisierung/Versuchsplanung nicht, wenn man alle Verunreinigungen vorhersehen und verhindern würde; hat man aber in Realität nie
 - Mögliche Konfundierung?
 - -> perfekt randomisieren oder matchen oder sonst wie rausheben
 - -> wenn das nicht geht: erheben und als Kovariate heranziehen
- **Beziehung zu anderen Analysen**
 - Stepdown-MANOVA
 - Sequenzielles Betrachten der UVs (z.B. bei Messwiederholung)
 - Ab dem 2. Schritt ist das dasselbe wie eine MANCOVA mit den bereits abgehakten UVs als Kovariaten
 - MANOVA mit Kovariate als zusätzlicher Faktor
 - Wir verlieren nur einen Freiheitsgrad durch Hinzunahme der Kovariate, d.h. wir verlieren wenig Power – damit ist das „statistisch gesehen relativ billig“
 - Wenn ich die Kovariate als zusätzlichen Faktor hinzunehme, verliere ich mehr Freiheitsgrade
- **Kovariaten („Confounder-Variablen“, d.h. bekannte & erhobene Störvariablen)**
 - Kovariate = Variable, die Varianz in AV(s) aufklärt, aber nicht im Design enthalten ist (z.B. weil es keine theoretische Basis dafür gibt, nur eine Verunreinigung der Daten vorliegt etc.)
 - Gute Kovariate korreliert mit AV(s), aber nicht mit UV(s)
 - Sie sollte signifikant sein. Nicht-signifikante Kovariaten werden rausgeworfen, es sei denn in dem sehr mysteriösen Fall, dass sie trotzdem die MANOVA unsignifikant machen, wenn man sie mit reinnimmt – dann beides berichten.
- **Vorgehen**
 - Impact beurteilen
 - Einmal mit und einmal ohne Kovariate rechnen
 - Können jeweils einzeln wieder weggelassen werden, wenn sie nicht signifikant beitragen
 - Am Anfang einer MANOVA: Überprüfen, ob es im Datensatz Unterschiede bei potenziellen Kovariaten gibt -> das macht Randomisierungs-Probleme -> als Kovariate dazunehmen
 - Bei Baseline-Unterschieden den ersten Messzeitpunkt als Kovariate nehmen, dann ANOVA ohne Messwiederholung nur mit zweitem Zeitpunkt rechnen
 - Erlaubt Ermittlung des Treatmenteffekts mit gleichzeitiger Bereinigung der Post-Werte um den Effekt des Prä-Werts
 - Problem: Wenn man die Kovariate z-transformiert, bekommt man andere Signifikanzen, weil das Modell regressionsanalytische Anteile enthält und wir hier mit der Konstante spielen
 - Anzahl der Kovariaten sollte nicht mehr als 10 sein
 - Problem mit nicht signifikanten Kovariaten: Auch die reduzieren nicht erklärbare Varianz, deshalb machen sie manchmal den Haupteffekt signifikant -> aufpassen! Fälschungspotenzial (ähnlich Interpretation von R^2 bei Regressionsanalyse mit nicht signifikanten Prädiktoren)
- **Voraussetzungen**
 - (dazu die normalen Voraussetzungen der (M)ANOVA)
 - Notwendig
 - Summe der Fehler und der mittleren Fehler = 0
 - d.h. außer der Kovariate gibt es keine anderen systematischen Messfehler
 - d.h. ich kann ein Restriktionsmodell anlegen
 - Fehler in den einzelnen Gruppen korrelieren nicht miteinander
 - Fehler in den einzelnen Gruppen sind normalverteilt

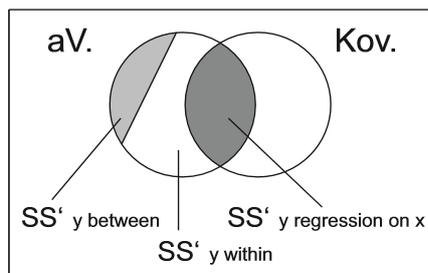
- Nicht zwingend notwendig
 - Regressionsgeraden gleiche Steigungskoeffizienten β_j
 - d.h. identischer Einfluss der Kovariaten innerhalb der Gruppen -> Geraden mit gleicher Steigung
 - SPSS kann das nur so, andere Statistikprogramme kommen auch mit unterschiedlichen Einflüssen klar
 - Regressionsgeraden zwischen den Gruppen gleich (also auch Intercept)
 - Kovariate messfehlerfrei erhoben
 - Messobjekte randomisiert erhoben

- **Signifikanztest**

- F-Test für Kovariate
 - $F = MS_{y \text{ regression on } x} / MS_{y \text{ residuals}}$
 - $Df_1 = 1$
 - $Df_2 = n - 2$
- Bei signifikanter Kovariate: bereinigte Mittlere Quadratsummen für Faktor
 - $F = MS'_{\text{between}} / MS'_{\text{within}}$
 - $Df_{\text{between}} = p - 1$
 - $Df_{\text{within}} = N - p - 1$
 - (das ' bedeutet nur, dass Kovariate auspartialisiert)

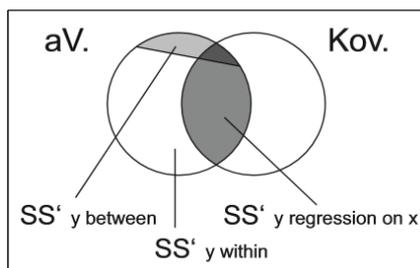
- **Mögliche Outcomes (Quadratsummenzerlegung bildlich dargestellt)**

- Schön: Kovariate korreliert mit AV, aber nicht mit UV



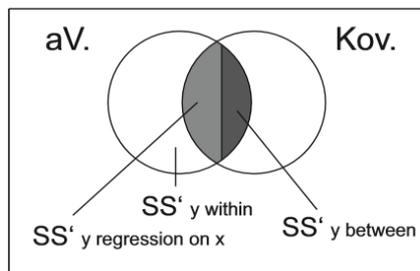
- Hier reduziert die Kovariate die nicht erklärbare Varianz um die dunkle Fläche, d.h. sie wird aus der abhängigen Variablen herauspartialisiert
- ANOVA wird danach sozusagen nur noch an den linken zwei Flächen gerechnet und wird dann natürlich viel eher signifikant

- Blöd: Kovariate korreliert mit AV und ein bisschen mit UV



- Kovariate reduziert sowohl Fehler- als auch durch Faktor erklärte Varianz, sodass sich an Signifikanz kaum was ändert
- Wenn das passiert, ist es das kein inhaltliches Argument für keinen Einfluss der Kovariate (bloß vielleicht trotzdem ein Grund, sie aus der Analyse wegzulassen)

- Blöd: Kovariate korreliert mit AV und hoch mit UV



- Kovariate erklärt die gesamte Varianz, die der Faktor erklären soll -> Faktor nicht mehr signifikant
- Könnte Hinweis auf Mediation sein, dann ist es nicht so, als hätte der Faktor keinen Einfluss - > da sollten wir dann aber Strukturgleichungsmodelle rechnen, bevor wir eine Mediation behaupten

- **Effektgröße: partielles η^2**

$$\eta_{P,treatment}^2 = \frac{SS'_y \text{ between}}{SS'_y \text{ between} + SS'_y \text{ within}}$$

-
- „ ‘ “ bedeutet wieder auspartialisierte Kovariate

- **Anwendung in SPSS**

- Input
 - „ALM“ -> „Messwiederholung“
 - „Innersubjektfaktor“ = MV
- Output
 - Innersubjekteffekte: Faktor * Kovariate: Interaktion (soll nicht existieren)
 - Durch unterschiedliche df durch Korrektur zwar gleiche F-Werte, aber unterschiedliche Signifikanz
 - Wenig df -> höherer kritischer Wert -> weniger Teststärke
 - „quadratischer Trend“: Parabel
 - Findet dieses Verfahren
 - „kubischer Trend“: steilere Parabel
 - Maximum, das SPSS kann
 - Solange ich alles in ein Modell unterbringe, ist alpha-Fehler-Kumulierung kein Problem

Hierarchische Clusteranalyse

- **Ziel:** Gruppenbildung basierend auf Ähnlichkeit
- **Anwendung**
 - Multivariate Suche nach homogenen Untergruppen im Datensatz
 - Multivariate Suche nach Ausreißern im Datensatz
 - Kann Fälle ermitteln, die man univariat nicht findet
- **Voraussetzungen**
 - Repräsentative Stichprobe
 - Entweder alle Variablen nominal- oder alle intervallskaliert (nicht hierarchische Analysen schaffen auch Mischungen)
 - Nominalskalenniveau
 - Z.B. Geschlecht, Indikation, Gruppenzugehörigkeit -> lassen sich alle als Dummyvariablen binär darstellen (X und nicht X)
 - Immer eine Stufe weglassen
 - Ordinal zu Dummy: lieber Median Split oder Quartile, nicht viele unabhängige Dummies für alle Stufen -> zu viel Multikollinearität und falsche Repräsentation der Verhältnisse zueinander (so als wären 6 und 7 gleich unabhängig wie 1 und 7)
 - Achtung: Oft impliziert „nicht X“ nicht wirklich Ähnlichkeit, z.B. bei Gruppenzugehörigkeit zu 4 Gruppe
 - Starke theoretische Basis; Begründung für Variablen
 - Falls Multikollinearität: kontrollieren!
 - CA kommt mit Faktoren und FA mit Clustern schlecht klar (dort Verstoß gegen Voraussetzung unkorrelierter Messfehler)
- **Vorher entscheiden**
 - Proximitätsmaß (Maß für Ähnlichkeit von Punkten)
 - Fusionierungsalgorithmus (Maß für Ähnlichkeit von Clustern)
- **Verfahren bei hierarchischer Clusteranalyse**
 - Nullter Schritt: n Cluster der Größe 1
 - Erster Schritt: Welche zwei Personen sind sich am ähnlichsten?
 - alle paarweisen Distanzen ermitteln
 - Folgende Schritte:
 - Entweder ein weiteres Zweier-Cluster bilden
 - Oder ein drittes Objekt zum ersten Cluster zuordnen
 - Letzter Schritt: 1 Cluster der Größe n
 - Personen, die als letztes zugeordnet werden, nochmal anschauen – könnten Ausreißer sein!
- **Ähnlichkeitsfunktion**
 - **Nominalskalenniveau**
 - Proximitätsmaße basieren auf Ähnlichkeitsfunktion:
$$s_{ij} = \frac{a + \delta \cdot d}{a + \delta \cdot d + \lambda \cdot (b + c)}$$
 - s_{ij} : Ähnlichkeit zwischen i und j
 - δ, λ : mögliche Gewichtungsfaktoren der Ähnlichkeitsmaße s_{ij}
 - Liegt allen Proximitätsmaßen zugrunde
 - **Intervallskalenniveau**
 - Proximitätsmaße basieren auf L-Normen:

$$d_{kl} = \left[\sum_{r=1}^R |x_{kr} - x_{lr}|^c \right]^{\frac{1}{c}}$$

d_{kl} : Distanz der Objekte k und l
 x_{kr}, x_{lr} : Koordinaten der Objekte k und l auf der r-ten Dimension ($r=1, 2, \dots, R$)
 $c \geq 1$

- Mit $c=1$: City-Block-Metrik
- Mit $c=2$: euklidische Metrik

- Proximitätsmaße

- Nominalskalenniveau

Name des Koeffizienten	Gewichtungsfaktor		Definition
	δ	λ	S_{ij}
Tanimoto (Jaccard)	0	1	$\frac{a}{a+b+c}$
Simple Matching (M)	1	1	$\frac{a+d}{m}$
Russel & Rao (RR)	-	-	$\frac{a}{m}$
Dice	0	$\frac{1}{2}$	$\frac{2a}{2a+(b+c)}$
Kulczynski	-	-	$\frac{a}{b+c}$

- m ist die Anzahl der Vergleich, welche zwischen zwei Objekten durchgeführt wurden.
 $m = a + b + c + d$

- Intervallskalenniveau: „Distanzmaße“ („Distanz“ erfordert Intervallskala)

- Minkowski-Metriken (L-Normen)

- **L₁-Norm (City-Block-Metrik; Mannheimer Metrik; Manhattan-Metrik)**

$$d_{kl} = \sum_{r=1}^R |x_{kr} - x_{lr}|$$

-
- Summe der Kanten („abgefahren“ wie im Stadtverkehr)

- **L₂-Norm (euklidische Metrik)**

$$d_{kl} = \sqrt{\sum_{r=1}^R |x_{kr} - x_{lr}|^2}$$

-
- Direkte Verbindung zwischen 2 Punkten (errechnet über Satz des Pythagoras)
- Gibt's auch als quadrierte euklidische Distanz (SPSS-Standard)
- Wird bei ähnlichen Distanzen auf mehreren Variablen kleiner als City-Block

- **Mahalanobis-Distanz**

$$d_{kl} = \prod_{r=1}^R |x_{kr} - x_{lr}|$$

-
- Produkte der Distanzen auf verschiedenen Merkmalen
- Wird 0, sobald zwei Objekte in mindestens einem Merkmal identisch sind
- Leonhart: *Ich kenn kaum Artikel, die das verwenden (und er findet es auch komisch)*

- **Q-Korrelationskoeffizient**

- Leonhart: *„ganz abgespacet, kenne eigentlich keine Publikation damit“*
- Definiert nicht Differenz, nur Ähnlichkeit

$$r_{k,l} = \frac{\sum_{j=1}^J (x_{jk} - \bar{x}_k) \cdot (x_{jl} - \bar{x}_l)}{\sqrt{\sum_{j=1}^J (x_{jk} - \bar{x}_k)^2 \cdot \sum_{j=1}^J (x_{jl} - \bar{x}_l)^2}}$$

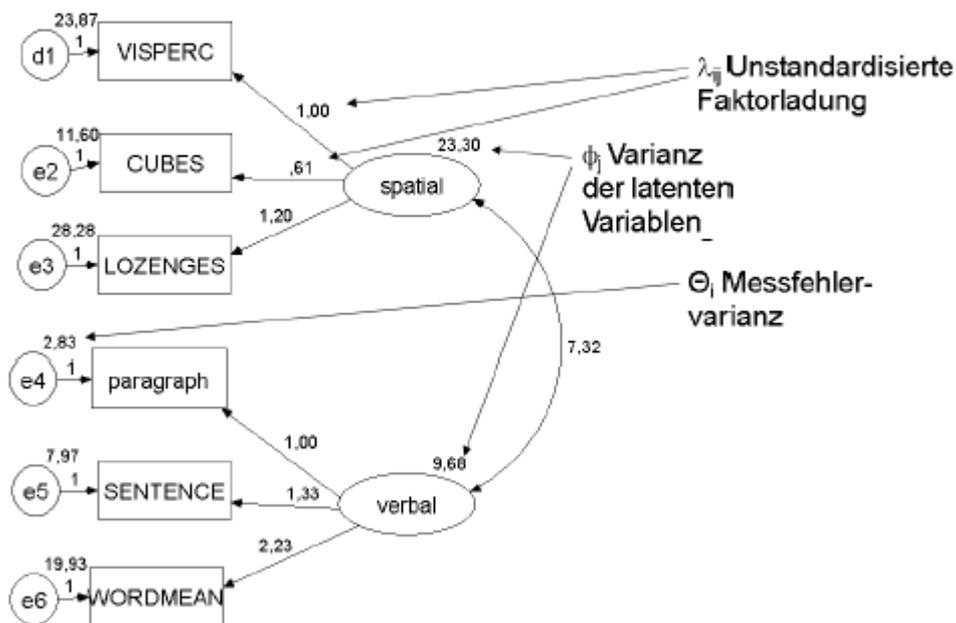
- - X: Ausprägung der Eigenschaft
 - Xstrich: Durchschnittswert aller Eigenschaften eines Objekts bzw. Clusters
- **Fusionierungsalgorithmen**
 - Für alle Proximitätsmaße
 - **Average-Linkage**
 - Verwendet durchschnittliche Distanzen zwischen Punktepaaren
 - Zwei Subtypen (Unterschied?)
 - **Linkage zwischen den Gruppen**
 - Durchschnittliche Distanz aller Paare zwischen den (noch zu fusionierenden) Gruppen
 - Distanzen innerhalb bereits existierender Cluster nicht berücksichtigt
 - **Linkage innerhalb der Gruppen**
 - Durchschnittliche Distanzen aller Paare innerhalb der (potenziellen neuen) Gruppe
 - Inklusive solcher Paare, die sich bereits im selben Cluster befinden
 - **Single-Linkage (Nearest-Neighbour-Verfahren)**
 - Verwendet ursprüngliche Distanzen
 - Nur eine Distanz für Fusionierung von zwei Gruppen: kleinste Differenz zwischen Elementen beider Gruppen
 - Nachteil: potenzieller Wurstcharakter (evtl. sogar gekringelt um einen Punkt, der eigentlich dazugehören müsste)
 - **Complete-Linkage (Farthest-Neighbour-Verfahren)**
 - Wie Single-Linkage, aber Maximaler Abstand verwendet
 - Vorteil: evtl. kompaktere Cluster
 - Bildet eher kleine (und daher viele) Gruppen
 - Nur für Distanzmaße
 - **Centroid**
 - Nicht mehr ursprüngliche Distanzen, sondern für jedes Cluster neue Clusterkennwerte (Gruppenmittelwerte aller Variablen) und davon ausgehend neue Distanzen innerhalb der und zwischen den Gruppen ermittelt (also nach jeder Fusionierung!)
 - Dann werden die Gruppen fusioniert, deren Mittelwerte die geringste Distanz voneinander haben
 - Problem: Mittelwert kann durch Ausreißer und schiefe Verteilungen leicht verzerrt werden
 - **Median**
 - Wie Centroid, nur mit Median, um Verzerrungen durch Ausreißer und schiefe Verteilungen zu vermeiden
 - Je größer Cluster werden, desto geringer wird der Einfluss von Ausreißern
 - **Ward**
 - *Leonhart: Heute Standard, Rest eher historisch (früher Rechenzeitproblem!)*
 - *Das Buch sagt da was Anderes; demzufolge werden auch Average Linkage und Centroid häufig eingesetzt*
 - Es werden jene Objekte/Gruppen fusioniert, welche eine möglichst große Clusterhomogenität beinhalten -> minimale Varianzen innerhalb potenzieller neuer Cluster -> d.h. quadrierte Differenzen zwischen einzelnen Objekten und Gruppenmittelwert gehen mit ein

- Besonderheiten
 - Bildet eher gleich große Gruppen
 - Nimmt nur Leute mit vollständigen Daten mit rein
- Besonders empfehlenswert, wenn
 - Inhaltlich sinnvolles Distanzmaß verwendet werden kann
 - *Leonhart: Gegeben, falls die Form der Ermittlung der Distanz für die inhaltliche Fragestellung und z.B. die Bedeutung des Nicht-Besitzens eines Merkmals passt. Bei der euklidischen Distanz gehen beispielsweise alle Variablen gleich stark in die Berechnung des Maßes ein, bei der quadrierten euklidischen Distanzen werden Werte kleiner eins (durch Quadratur) schwächer gewichtet als Werte größer eins, bei der Mahalanobis-Distanz hingegen wird der Differenz von Null ein stärkeres Gewicht gegeben.*
 - Intervallskalierte, unkorrelierte Variablen ohne Ausreißer
 - Dann auch immer euklidische Distanz nehmen
 - *Leonhart empfiehlt Ward und euklidische Distanz generell für intervallskalierte Daten*
 - Von gleich großen Gruppen ausgegangen werden kann
- **Anwendung in SPSS**
 - Input: Klassifikation -> „Hierarchische Clusteranalyse“
 - Dendrogramm geben lassen
 - Und zwar vertikal
 - Zuordnungsübersicht
 - Output
 - Zuordnungsübersicht
 - „Nächster Schritt“: wann diese Person/dieses Cluster das nächste Mal wieder verwendet wird
 - Nervig: Es gibt keinen Kennwert dafür, ab wann ein Cluster „zu“ inhomogen ist -> grafische Analyse
 - Eiszapfendiagramm (sollte eigentlich Stalagmiten-Diagramm heißen)
 - Einfach grafische Darstellung der Zuordnungsübersicht, auf Individuenebene
 - Dendrogramm
 - Darstellung der Zuordnung, bloß dass man noch sieht, wer zu wem zugeordnet wird -> besser
 - Achtung: sieht so aus als wären die ersten paar Personen alle gleichzeitig geclustert worden und gleich weit voneinander weg
 - Wenn man, wie hier, quadrierte euklidische Distanz nimmt, hat man einen großen Sprung; bei nicht-quadrierter werden kleine Distanzen im Verhältnis zu großen größer und das erschwert die Einteilung (*eher schlecht als gut, obwohl evtl. Überschätzung der letzten Distanzen*)
 - Man kann sich die Clusterzugehörigkeit als neue Variable ausgeben lassen
 - „Messniveau“: Hier Distanzmaß einstellen

Strukturgleichungsmodelle

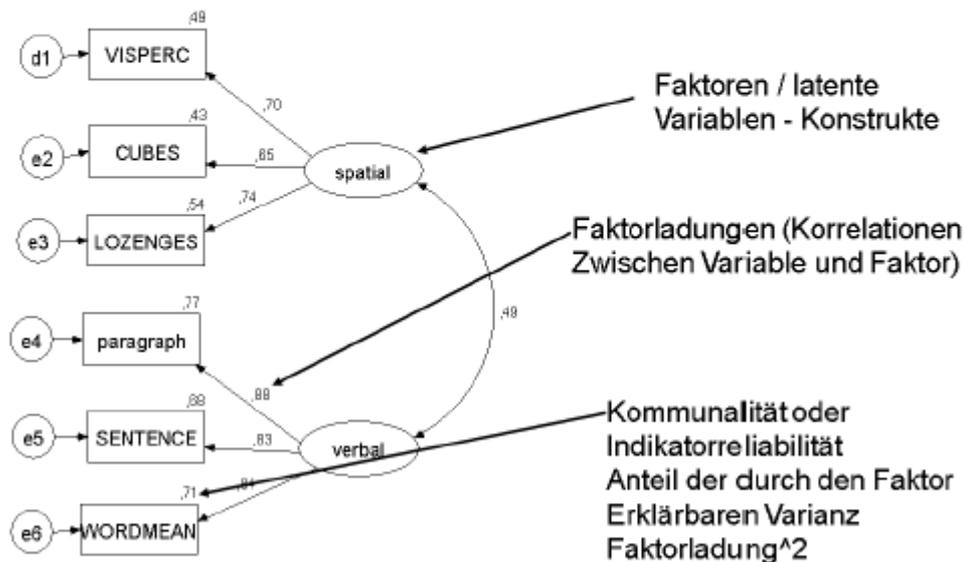
- **Ziel**
 - Mehrere Beziehungen gleichzeitig geschätzt, nur eine Modelltestung -> keine Alpha-Fehler-Inflation
 - Können gestapelt sein: Selbe Variable in einer Beziehung UV, in anderer AV
 - Latente Variablen können integriert werden
 - Messfehler explizit modellieren
 - Eigentlich nicht die MESSfehler, sondern die Vorhersagefehler bei Schätzung einer Variable mit einer anderen
 - -> Reliabilitätsbereinigung für über mehrere manifeste Variablen erhobene Konstrukte, bessere Messmodelle -> höhere Effektgrößen (ca. 12% besser als bei einzelnen Tests)
 - Modelle definieren und überprüfen an empirischen Daten
 - Vielzahl von Beziehungen übersichtlicher darstellen
- **Eigenschaften**
 - Konfirmatorisch
 - Forscher muss alle Beziehungen theoriegeleitet spezifizieren (also, ob sie da sind oder nicht)
 - Kann keine Kausalität finden! Nur wiederlegen
 - Berechnungen auf Ebene der Kovarianz- und Korrelationsmatrix
- **Standardisierte vs. Unstandardisierte Lösung**
 - Eigentlich arbeitet man die ganze Zeit in der unstandardisierten Version, und dann zum Interpretieren wandelt man das in die standardisierte um, weil die besser zu interpretieren ist

Unstandardisierte Lösung



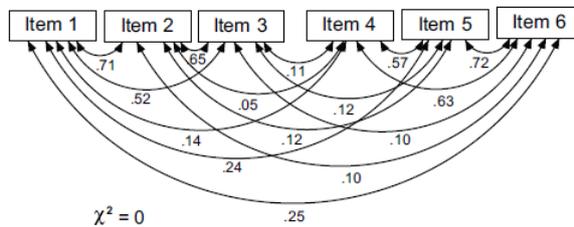
○

Standardisierte Lösung



-
- In der standardisierten Lösung sind alle Konstruktvarianzen = 1.
- **Ablauf**
 - Konstrukte festlegen
 - Messmodell (Items zu Konstrukten) spezifizieren
 - Grafisches Modell (AMOS)
 - Gleichungssystem
 - Daten erheben
 - Messmodell validieren
 - Goodness-of-Fit
 - Konstruktvalidität
 - Prüfen der einzelnen Parameter (Item-zu-Konstrukt-Beziehungen)
 - Strukturmodell (Konstrukte untereinander) spezifizieren
 - Grafisches Modell (AMOS)
 - Gleichungssystem
 - Schätzung der Parameter (Beziehungen)
 - Gleichungssystem wird gelöst
 - Iterative Schätzalgorithmen, z.B. ML, ULS, GLS, ADF, ...
 - Meistens Maximum-Likelihood-Schätzung
 - Strukturmodell validieren
 - GOF
 - Signifikanz, Richtung und Größe der Beziehungsparameter
 - Prüfen der einzelnen Parameter (Konstrukt-Beziehungen)
 - Interpretation
- **Voraussetzungen**
 - Unabhängigkeit der Beobachtungen
 - Ausreichend große Stichproben (mind. 100, je nach Schätzalgorithmus)
 - Endogene Variablen
 - Metrisch
 - Normalverteilt
 - Multivariate Normalverteilung
 - Identifizierbarkeit

- D.h. es gibt wirklich eine Lösung für das Modell
 - D.h. zum Beispiel immer eine 1 als Pfad pro Konstrukt, kein Konstrukt mit nur einem Prädiktor, etc.
- Linearität der Zusammenhänge
- Außerdem überprüfen
 - Fehlende Daten
 - Weil AMOS auch mit N=1 rechnet, ohne sich zu beschweren
 - Möglichst wenige fehlende Werte
 - Kein NRM
 - Ersetzung möglich bei MCAR, MAR
 - Bester Standard nach FIML (Full Information Maximum Likelihood: d.h. nicht die konkreten Personenwerte werden ersetzt, sondern nur die relevanten Werte, also Varianz und Kovarianz, berechnet) und Multiple Imputation
 - Multikollinearität
 - Wenn $r > 0,85$, kann es Probleme bei Lösungsermittlung geben
 - Toleranz (= $1 - R^2$ mit allen anderen Variablen als Prädiktoren) sollte $> 0,1$ sein
 - Redundante Variablen eliminieren oder zu einer Skala zusammenfassen
 - Ausreißeranalyse
 - #48 ergänzen
 - Evtl. Transformation?
 - Homoskedastizität
- Verletzungen umso schlimmer, je ...
 - Komplexer das Modell (weil Folgen schwerer abzuschätzen)
 - Kleiner die Stichprobe
 - Weniger robust der Schätzalgorithmus (z.B. ADF besser als ML)
- **Spezifikationen/Entscheidungen**
 - **Variablentypen**
 - **Manifeste Variablen**
 - Wirklich erhoben
 - Pfeile deuten von LVs zu MVs
 - **Latente Variablen**
 - Nicht direkt messbar
 - Geschätzt über mehrere (SEM braucht mehr als 2, damit's sinnvoll ist, eher 3-7) manifeste Variablen
 - *Leonhardt findet tatsächlich, Fragebögen sollten immer 3-7 Items pro LV haben, weil künstliche Cronbachs-Alpha-Erhöhung oder mehrere Subskalen*
 - **Exogene Variablen**
 - Werden durch keine andere Variable im Modell vorhergesagt
 - Kein Pfeil endet hier
 - **Endogene Variablen**
 - Werden durch mindestens eine andere Variable prognostiziert (aber nicht unbedingt perfekt)
 - Mind. 1 Pfeil endet hier
 - **Inputmöglichkeiten**
 - Rohdatensatz
 - Kovarianz- oder Korrelationsmatrix
 - *Kovarianz besser als Korrelationen, da Mittelwertvergleiche möglich und standardisierte Outputs für leichtere Interpretation trotzdem anforderbar*
 - **Pfade definieren**
 - Empirisch korrelieren ja praktisch alle Variablen irgendwie miteinander



$$\chi^2 = 0$$

- der Forscher legt aber bestimmte Beziehungen als 0 fest, sodass die Korrelationen über die anderen Beziehungen erklärt werden müssen, soweit möglich; diese anderen Beziehungen werden dann berechnet
 - Deshalb ist Chi = 0 und Fit am besten, wenn alle Beziehungen berechnet werden, aber das Modell ist dann nicht toll (Overfit)
- Auf 0 setzen
 - Alle Beziehungen, die theoretisch nicht existieren sollen
- Auf 1 setzen
 - *Leonhart: Pfade auf 1 setzten wir am Anfang, wenn wir ein Modell definieren, also in der unstandardisierten Lösung; in der standardisierten ändern sich dann die Pfadgewichte, sodass nicht mehr erkennbar ist, welcher auf 1 gesetzt wurde*
 - Pro Konstrukt (also pro Messmodell) wird eine Item-zu-LV-Beziehung auf 1 gesetzt (d.h. dieses Item korreliert hypothetisch zu 1 mit dem Konstrukt)
 - Man braucht für ein lösbares Gleichungssystem mehr bekannte als unbekannte Variablen
 - Um die Wahrscheinlichkeit für ein lösbares Gleichungssystem zu erhöhen, legen wir unbekannte Parameter fest
 - Man könnte stattdessen auch Varianzen festlegen; das sollte eigentlich äquivalent sein, aber man hat in empirischen Studien festgestellt, dass damit die Parameter schlechter geschätzt werden
 - *In Hair steht auch, dass das Setzen von Faktorladungen auf 1 in der KFA dafür sorgt, dass das betreffende Konstrukt auf eine Skala festgelegt wird, und dass die tatsächliche Ladung später neu geschätzt wird und am Ende nicht mehr 1 ist.*
 - *(Wenn man 5x so viele Äpfel wie Birnen kaufen soll, kann man das nicht lösen, wenn man nicht erst mal für einen der Parameter einen Wert einsetzt.)*
- **Modellbewertung**
 - Drei Modelle, die verglichen werden können
 - **Default model** (zu testendes Modell)
 - **Saturated model** (gesättigtes Modell)
 - alle Beziehungen, df = 0
 - **Independence model** (globale Nullhypothese)
 - **Passung des Gesamtmodells**
 - **Goodness-of-Fit** (Vergleich der empirischen Kovarianzmatrix und der vom Modell vorhergesagten Kovarianzmatrix, also saturated vs. default)
 - **χ^2 -Wert** (einziger Signifikanztest)
 - Wenn = 0, dann perfekte Passung -> Overfit
 - Soll möglichst klein sein, sodass p möglichst groß (Signifikanz bedeutet, dass signifikante Abweichungen zwischen Matrizen, d.h. Modell schlecht)
 - Erwartungswert = df
 - D.h. je mehr Freiheitsgrade, desto höher darf χ^2 sein
 - Wird immer eher signifikant, je größer N
 - -> bei großen Stichproben und/oder vielen Variablen/Beziehungen kulanter sein
 - -> Kritik, deshalb weitere Kennwerte Entwickelt
 - Wird immer noch berichtet (bzw. χ^2/df)

- Andere Fit-Indices, hier nach Kategorien:
 - **Absoluter Fit**
 - **Vergleich zu testendes vs. saturiertes Modell**
 - Gewichten Passung höher als Einfachheit
 - D.h. tendieren eher Richtung independence model
 - Kennwerte
 - χ^2 -Freiheitsgrad-Verhältnis (CMIN/df)
 - Sollte < 1,5; 2,5; 3 (je nach Quelle) sein
 - Abhängig von df, aber auch N (und das natürlich immer noch zugunsten kleiner Stichproben)
 - **Noncentrality Parameter (NCP)**
 - $\chi^2 - df$
 - **Standardisierter Noncentrality Parameter (SNCP)**
 - $(\chi^2 - df)/N$
 - **GoF (GFI)**
 - $1 - F/F_{baseline}$
 - **Root Mean Square Error (RMSEA)**
 - Wurzel(F/df)
 - **Inkrementeller Fit**
 - **Vergleich zu testendes vs. Nullmodell**
 - **Normed Fit Index (NFI)**
 - **Tucker-Lewis-Koeffizient (TLI)**
 - **Adjusted Goodness of Fit Index (AGFI)**
 - **Parsimony Fit**
 - **Vergleich zu testendes vs. Nullmodell**
 - Gewichten Einfachheit/Generalisierbarkeit höher als Passung
 - D.h. tendieren eher Richtung saturated model
 - **PRATIO**
 - **PNFI**
 - **PCFI**
 - **AIC-Wert nach Akaikef**
- **Anforderungen**
 - CMIN nicht signifikant bei $100 < N < 300$
 - CMIN/df < 1,5, 2, 3, 5
 - Incremental Fit Indizes > 0,9 bzw. 0,95
 - RMSEA und RMSR < 0,08 bzw. 0,05
 - Bei Modellvergleich Parsimony-Maße
 - Bei Beurteilung der Güte eines Modells alle drei Typen berücksichtigen
 - Kline sagt, wir sollen berichten (und wenn das nicht berichtet wird, mit AMOS-Fit-Calculator nachrechnen):
 - cmin
 - df
 - p
 - GFI
 - NFI
 - CFI
 - TLI
 - RMSEA
- **Lokale Passung (Bewertung einzelner Pfade)**
 - **Konvergente Validität**
 - **Indikatorreliabilität** (= Itemreliabilität, Kommunalität)

- SMC-Wert (squared multiple correlation) sollte > 0,4 sein
- **Signifikanz der Faktorladung**
 - Soll signifikant werden
 - Korreliert hoch mit Indikatorreliabilität
 - Nur für die nicht auf 1 gesetzten Pfade berechnet
 - Deshalb: Wenn man ganz genau sein will, sorgt man dafür, dass derjenige Indikator mit der kleinsten Varianzaufklärung nicht der ist, der auf 1 gesetzt wurde, und ändert das vor Publikation gegebenenfalls nochmal
- **Faktorreliabilität**
 - Soll > 0,6 sein
$$FR = \frac{(\sum \lambda)^2 \cdot \phi}{(\sum \lambda)^2 \cdot \phi + \sum \theta}$$
 - Braucht
 - Messfehlervarianz θ
 - Unstandardisierte Ladung λ
 - Varianz des Faktors ϕ
 - Bezieht sich jeweils nur auf ein Messmodell -> muss für jedes Messmodell einzeln gemacht werden
- **Durchschnittlich erfasste Varianz**
 - Soll > 0,5 sein
$$DEV = \frac{\sum \lambda^2 \cdot \phi}{\sum \lambda^2 \cdot \phi + \sum \theta}$$
 - Bezieht sich jeweils nur auf ein Messmodell -> muss für jedes Messmodell einzeln gemacht werden
- **Diskriminante Validität**
 - **Fornell-Larcker-Kriterium**

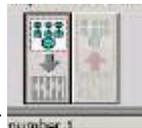
$$DEV > \max r^2$$

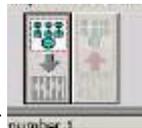
$$\frac{\max r^2}{DEV} < 1$$
 - max r^2 = höchste Korrelation des betrachteten Faktors mit einem anderen Faktor
 - **Cmin-Differenztest**
 - **Signifikante Reduktion der Modellpassung, wenn die Korrelation zweier Faktoren auf 1 gesetzt wird? (Ich gehe davon aus, dass das passieren soll?)**
 - Trick in AMOS: Varianzen der Variablen auf 1 setzen -> dann kann durch Standardisierung auch r gleich 1 gesetzt werden
 - *Leonhart: Sollte häufiger verwendet werden.*
- Excel-File confirm_calculator.xls von Markus Wirtz (Runterladen! Das berechnet alle Kennwerte)
- **Freiheitsgrade**
 - Df = Stichprobenmomente – Schätzmomente
 - **Stichprobenmomente**
 - Berechnet man in Kovarianzmatrix
 - Konzeptuell: Varianzen der MVs + Kovarianzen zwischen MVs
 - Abzählen: Anzahl der MVs + Anzahl der Korrelationen zwischen den MVs

$$\text{Stichprobenmomente} = \frac{p \cdot (p + 1)}{2}$$

- Berechnen: Bei p Variablen:
 - #24: 21
 - **Schätzmomente** (Parameter, die geschätzt werden müssen)
 - Berechnet man im Modell
 - Anzahl der zu schätzenden Parameter im Modell
 - Abzählen: Anzahl der exogenen Variablen + Anzahl der Pfeile – Anzahl der Pfeile, die auf 1 gesetzt wurden (also Anzahl der Konstrukte; konzeptuell: Deren Varianzen)
 - #27: 13
- => Df abhängig von Komplexität des Modells (Variablen & Parameter), aber NICHT von N!
 => Je sparsamer ein Modell, desto mehr Freiheitsgrade und desto bessere Generalisierbarkeit
 -> Kompromisse zwischen Sparsamkeit und Passung
 Bei Anfangslösung (alle Beziehungen geschätzt) sind df=0
- **Modelle vergleichen**
 - **Nested Models**
 - Verknüpfte Modelle = Modelle, die sich auseinander ergeben, indem man die Beziehungen ändert, aber gleich viele Variablen haben
 - Kann man über Chi²-Wert vergleichen, da Chi²-Differenz ebenfalls Chi²-verteilt
 - Sparsamkeit vs. Passungsverlust
 - Pfade dazunehmen? Pfade dazunehmen und 0 setzen?
 - **Non-nested Models**
 - Vergleich unabhängiger Modelle nur über AIC-Wert möglich
 - Problem mit SEM generell: nie klar, ob Optimum gefunden
 - **Pfadanalysen: Berechnung der Beziehungen**
 - **Ziel:** Quantifizierung des Einflusses (eigentlich der Korrelation beider miteinander) einer Variable auf eine andere
 - **Grundregeln**
 - Korrelation zweier Variablen = Summe aller Pfade, die die beiden Variablen verbinden; ein Pfad berechnet sich wiederum als Produkt aller Korrelationen entlang seiner Einzelteile
 - Wenn man auf einem Pfad vorwärts gegangen ist, darf man nicht mehr zurück (aber man darf erst zurück und dann vorwärts)
 - Dieselbe Variable darf nicht zweimal durchlaufen werden
 - Pfad darf nur einen Korrelationspfad (gebogenen Pfeil) enthalten
 - **Annahmen**
 - Alle theoretisch vorhandenen Kausalbeziehungen sind im Modell enthalten
 - Geringste theoretisch gerechtfertigte Anzahl an Beziehungen in Modell aufgenommen
 - Beziehungen zwischen Variablen linear
 - **Bewertung der Pfade**
 - Effektgrößen: .1 ist schwach, .3 ist mittel, .5 ist stark
 - **Voraussetzungen für Kausalität**
 - Genügend starke Assoziation
 - Zeitliche Abfolge: Ursache -> Wirkung
 - Fehlen alternativer kausal wirksamer Variablen
 - Theoretische Basis für die kausale Beziehung
 - **Erklärung zu Leonhardts Darstellungen**
 - Pfeile gehen immer von LV zu MV
 - Messfehler sind Kreise mit e drin
 - Pfeile von Fehlern zu MVs immer 1 (wird so angenommen), deshalb nicht beschriftet
 - Schraffiert: endogen
 - Ovale: LVs, Rechtecke: MVs

- Gebogene Pfeile: Korrelationen, d.h. keine Kausalrichtung angenommen! (nur zwischen exogenen Konstrukten möglich)
 - Das Gepunktete auf #27 = Menge der Varianz, die Konstrukt 1 an Konstrukt 2 aufklärt
 - = Pfad von K1 zu K2 zum Quadrat
 - #27: kleine Zahlen über den Items: quadrierte Korrelation dieses Items mit der LV (hat nichts mit dem e zu tun!)
- **Häufige Fehler**
 - Spezifizierung
 - Falsche Datenbasis
 - **Kreuzvalidierung**
 - 0. Gruppen definieren innerhalb der Datei (auch verschiedene Dateien möglich)
 - 1. Modell wird in jeder Gruppe ermittelt, aber Kennwerte für das Gesamtmodell
 - 2. Parameter in einer Gruppe gleichsetzen mit denen, die für andere Gruppe geschätzt wurden
 - Pfade müssen benannt werden
 - Dadurch mehr Sparsamkeit (weil ja nur Varianzen geschätzt werden), aber schlechtere Passung
 - 3. Vergleich vorher-nachher: signifikante Reduktion der Modellpassung durch Gleichsetzung?
 - Wenn ja: Kreuzvalidierung gescheitert
 - Wenn nein: beide Gruppen entspringen gleicher Population
 - **Anwendung in AMOS** (SPSS kann das alleine nicht)
 - Freier Download Studierendenversion: <http://www.amosdevelopment.com/>
 - Vertrieb von SPSS übernommen
 - Grundeinstellungen minimal, deshalb muss man wissen, was man tut
 - Standardized estimates
 - Squared multiple correlations
 - Estimate means and intercepts bei fehlenden Werten
 - Normalerweise immer maximum likelihood als Voreinstellung
 - AMOS prüft nicht, wie viele Werte fehlen – ersetzt auch fast alle Werte, wenn ich ihm das sage
 - Gruppenauswahl braucht man v.a. für Kreuzvalidierung, sonst einfach Gruppe=1



- Grafisches Zeichnen des Modells bei dieser  Einstellung, dann berechnen lassen mit rechter Schaltfläche
 - Berechnet immer erstmal perfekte Passung; das heißt aber nicht hohe Varianzaufklärung (denn vielleicht gibt es einfach nur wenig systematische Varianz)
 - Chi kann auch Null sein, wenn R^2 ganz klein, so wie hier
 - (denn es kommt auch auf die lokale Passung an)
- „CMIN“ = eigentlich kein Chi-Quadrat-Wert, sondern ein CMIN-Wert, der anhand einer Chi-Quadrat-Verteilung geprüft werden kann

Hierarchische Lineare Modelle (Mehrebenenmodelle)

- **Ziel**
 - Regression: Varianzaufklärung in AV durch mehrere UVs
 - Aber: Daten hierarchisch strukturiert
 - Z.B. Klumpen in der Stichprobe, die alle dieselbe Ausprägung auf einer UV haben
 - -> d.h. keine Unabhängigkeit
 - Aussagen wie „bei gleicher xx gilt: je yy desto zz“
 - D.h. für jeden Klumpen eine eigene Regressionsgerade mit eigener Steigung sowie Konstante
- **Grundlegende Idee**
 - Ermittlung in mehreren Ebenen, z.B.
 - Ebene 1: innerhalb jedes Klumpens einzeln
 - Hier viele Regressionsanalysen
 - Level 1 ist immer dasjenige, in dem es Klumpen gibt
 - Ebene 2: für alle Klumpen
 - Hier nur eine Regressionsanalyse
- **Vorteile**
 - Keine Unabhängigkeitsannahme
 - Mittelwert pro Gruppe berücksichtigt
 - Level-1- und Level-2-Prädiktoren möglich
 - Level-3 ebenfalls integrierbar; meist macht man nicht mehr als 3, aber es gibt theoretisch keine Grenze
 - Anzahl der abhängigen Elemente auf Level-1 muss nicht gleich sein
 - Fehlende Werte auf Level-1 werden ignoriert
- **Voraussetzungen**
 - Kriterium intervallskaliert
 - Prädiktoren
 - haben Varianz
 - Keine Multikollinearität
 - Sonst Probleme bei ML-Schätzung
 - korrelieren nicht mit im Modell unberücksichtigten Variablen
 - Level-1-Residuen normalverteilt mit Mittelwert von 0
 - Stichprobengröße
 - 30/30-Regel nur für 1. Level gut
 - Wichtig: Anzahl der Gruppen
 - 100+ Gruppen für gute Schätzung der Zufallseffekte auf Level 2 nötig!
 -
- **Vorgehen**
 - **Nullmodell** (Level-1 als Zufallseffekt, mindestens eine Konstante im Modell)
 - **Weitere Modelle** durch Hinzunahme von Prädiktoren
 - **Level-1-Modell**
 - $$y_i = \beta_0 + \beta_1 \cdot x_i + r$$
 - Aber wie Konstante und Steigung vorhersagen?
 - Beispiel: Im Allgemeinen errechnet sich die Statistik-Punktzahl aus dem Gesamtmittelwert von $\beta_0 = 50$ + einem Betrag, der linear davon abhängt, wie lange man gelernt hat ($x = \text{Lernzeit in Stunden}$)
 - **Level-2-Modell**

$$\beta_0 = \gamma_{00} + \gamma_{01} \cdot p_1 + \gamma_{02} \cdot p_2 + u_0$$

$$\beta_1 = \gamma_{10} + \gamma_{11} \cdot p_1 + \gamma_{12} \cdot p_2 + u_1$$

-
- D.h. Vorhersage der Konstanten und Steigung auf Level 1 durch Level-2-Prädiktoren (p_1 und p_2)
- Beispiel:
 - Der Gruppenmittelwert der Statistiknoten β_0 ist von Jahrgang zu Jahrgang unterschiedlich; er setzt sich wiederum zusammen aus einem mittleren Mittelwert γ_{00} sowie einem Betrag, der negativ linear davon abhängt, wie groß der Statistikkurs in diesem Jahrgang war (p_1 = Größe des Statistikkurses), sowie einem Betrag, der linear davon abhängt, wie gut in diesem Jahr die Lehre war (p_2 = Maß für Qualität der Lehre, z.B. Ergebnis der Evaluation am Ende des Semesters)
 - Auch das Ausmaß, in dem eine Stunde mehr Lernen zu mehr Statistikpunkten führt, ist nicht bei allen Student*innen gleich groß. Der „Wechselkurs“ hängt wiederum von einem mittleren Wechselkurs γ_{10} ab sowie von einem Betrag, der linear davon abhängt, wie gut im entsprechenden Jahr die Lehre war (p_2 = Qualität der Lehre) (z.B. weil besonders gute Dozent*innen Übungsblätter zur Verfügung stellen, die das Lernen effizienter machen) sowie von einem weiteren Betrag, der linear davon abhängt, wie groß der Statistikkurs im entsprechenden Jahrgang war (p_1 = Größe des Statistikkurses) (z.B. weil bei großen Kursen die Student*innen besser Lerngruppen bilden können)
 - So kann Einfluss der Prädiktoren auf beiden Levels ermittelt und verglichen werden
- **Vergleichung der Passung der Modelle** über Deviance (Log-Likelihood-Wert)
 - **Log Likelihood Ratio**
 - Unterschied in **Deviance** zwischen Modellen
 - **Log Likelihood Ratio** χ^2 (df) = **deviance(Referenzmodell/Nullmodell) – deviance(Alternativmodell)**
 - Df = df(Alternativmodell) – df(Referenzmodell)
 - Signifikant geringere Deviance bedeutet höhere Wahrscheinlichkeit für die richtige Vorhersage
- **Modellvarianten**
 - Varianz kann auf mehrere Quellen zurückgeführt werden
 - Level-1-Zufallseinflüsse (immer im Modell)
 - Level-2-spezifische Zufallseffekte (eigentlich immer im Modell)
 - Systematische Effekte von Level-1-Prädiktoren
 - Systematische Gruppeneffekte von Level-2-Prädiktoren
 - Interaktionen zwischen Level-1- und Level-2-Prädiktoren
 - Nullmodell („intercept-only-model“; „baseline model“)
 - Nur Regressionskonstante und Residuen auf Level 1 und 2
 - Dient zur Berechnung der Intraclass-Korrelation
 - ρ = (Varianz zwischen Gruppen)/Gesamtvarianz
 - Hohes ρ -> Mehrebenenanalyse lohnt sich
 - D.h. HLM macht natürlich nur Sinn, wenn Klumpen-UVs wirklich mit Unterschieden auf AV zusammenhängen
- **Parameterschätzung**
 - Normalerweise mit kleinster quadratischer Abweichung
 - Voraussetzungen
 - Unabhängigkeit
 - Normalverteilung

- Wenn verletzt: Maximum Likelihood
- **Zentrieren der Prädiktoren**
 - Was?
 - In-Verhältnis-Setzen zum Mittelwert
 - Warum?
 - In psychologischer Forschung oft kein sinnvoller Nullpunkt -> Konstante in Regressionsanalyse (und Varianz der Konstanten in den Level-1-Analysen) nicht sinnvoll interpretierbar -> nach Zentrierung schon
 - Formen der Zentrierung
 - **Grand-mean-centering:** $(X_{ij} - \bar{X}_{..})$
 - D.h. Zentrieren am Gesamtmittelwert
 - Intercept und Varianz des Intercepts können geschätzt werden
 - Werte durch lineare Transformation wieder in ursprüngliche (nicht zentrierte) Werte überführbar
 - Interpretation
 - **Intercept:** Diesen Wert hat eine Person, die im Prädiktor den Gesamtmittelwert hat
 - **Varianz des Intercepts:** Variabilität der Personen zwischen den Gruppen
 - Wann?
 - Sinnvoll, wenn primär Level-2-Variablen interessant
 - Führt zu besseren Modellen
 - Komplexer; liefert nur gute Modelle, falls Gruppenmittelwerte von Level 1 als Prädiktor auf Level 2 einbezogen werden (und keine random slopes modelliert werden)
 - **Group-mean-centering:** $(X_{ij} - \bar{X}_{.j})$
 - D.h. Zentrieren am Gruppenmittelwert
 - Abweichung einer Person vom Gruppenmittelwert kann als Prädiktor verwendet werden
 - Aber: erzeugt nicht-äquivalente Modelle, d.h. Rücktransformation auf ursprüngliche Werte nicht möglich
 - Interpretation
 - **Intercept:** Diesen Wert hat eine Person, die im Prädiktor den Gruppenmittelwert hat
 - **Varianz des Intercepts:** Variabilität der durchschnittlichen Personen in der jeweiligen Gruppe zwischen den Gruppen
 - Wann?
 - Sinnvoll, wenn mich primär die Variablen für Level 1 interessieren
 - Achtung. Bloß nichts dichotom codiertes zentrieren!
- **Programmvergleiche**
 - Ergebnisse leicht unterschiedlich!
 - Deshalb muss man angeben, welches Programm man benutzt hat
 - **HLM6**
 - Schön, falls
 - Nur intervallskalierte Variablen
 - Nur Mehrebenenmodell rechnen
 - Einfache Bedienung ... wenn man seine Daten erfolgreich eingelesen hat
 - **Mplus**
 - Kann auch kategoriale/ordinalskalierte Variablen
 - Kann SEM und mehr
 - **SPSS**

- katastrophal
- R (lme4)
- Stata
 - schön
- SAS

• Anwendung mit MPlus (nicht klausurrelevant)

	Level 1	Level 2	
	1.220794	0.340346	-2.558013 -1.368350 1
	-0.605589	-0.796987	-2.558013 -1.368350 1
	-0.289682	0.013798	-2.558013 -1.368350 1
	-1.605372	0.429240	-2.558013 -1.368350 1
	0.786728	0.812824	-2.558013 -1.368350 1
	2.391486	-2.649131	0.970545 0.042646 2
	2.175327	-0.646735	0.970545 0.042646 2

- Gruppierungsvariable nicht unbedingt Prädiktor!
 - Kann ja z.B. sein, dass man mehrere Hauptschulen hat, die dann auf manchen UVs gleich sind, auf manchen verschieden
- Within -> Level 1
- Between -> Level 2
- Mini-Pfeil (#40): Fehler
- Man kann dem Programm sagen, ob es allen Level-2-Einheiten dieselbe Steigung geben soll oder ob auch unterschiedliche Steigungen okay sind („x und w dürfen s beeinflussen“).
 - Erst in zweiterem Fall gibt es überhaupt eine Variable „Steigung“, sonst ist es ja nur ein einzelner Wert.
 - MPlus modelliert den Zusammenhang von Prädiktor, Steigung und Kriterium als medierten Zusammenhang mit M als Mediator
 - Im Zweifel immer mehrere Steigungen zulassen; das führt zu besser Anpassung des Modells an die Daten
- Mehrebenen-Faktorenanalyse: Therapieevaluation durch Therapeuten und Patienten (ab 1-#63)
 - Einschätzungen der Patienten schwankt stärker intra- als interpersonell