

Lernskript

Statistik

nach Bortz „Statistik“ 4.Aufl.

**Prüfungsrelevanter Stoff
aus den Kapiteln 1-10
für das Vordiplom**

(weitgehend ohne Formeln)

**auch weitgehend ohne Methodenlehre :
dafür empfehle ich das Skript von Claudia Evers
(handgeschrieben)**

**Viel Spaß !
Benjamin Zeller**

(zur Überprüfung : 26 Seiten)

Deskriptive Statistik (Kap 1) :

Skalenarten :

- Nominalskala : Gleichheit, Verschiedenheit, Bsp : Krankheiten
- Ordinalskala : größer, kleiner Relationen, Bsp : Windstärken
- Intervallskala : Gleichheit von Differenzen, Bsp : Temperatur
- Verhältnisskala : Gleichheit von Verhältnissen (fester Nullpunkt), Bsp : Längenmessungen
- (Absolutskala : natürliche Maßeinheiten)
- ein empirisches Relativ läßt sich durch Zuordnung von Zahlen in einem numerischen Relativ ausdrücken. Wenn emp. Relativ nur durch bestimmte Zahlen ausgedrückt werden kann und nicht auch durch andere, dann ist es eindeutig / homomorph

Statistische Kennwerte :

Maße der zentralen Tendenz :

- Modalwert : höchster Wert / Kategorie in Verteilung
- Medianwert : mittlerer Wert / Kategorie in Verteilung
- arithmetisches Mittel : Durchschnitt der Werte / Kategorien (\bar{x})

Dispersionsmaße :

- Quadratsumme QS : Summe der quadrierten Abweichungen aller Meßwerte vom arithmetischen Mittel
- Varianz s^2 : Quadratsumme geteilt durch Anzahl der Meßwerte
- Standardabweichung s (auch Standarddiviation SD oder Streuung) : Wurzel aus Varianz

Wahrscheinlichkeitstheorie & - verteilungen (Kap 2) :

Axiome :

- Additionstheorem für nicht disjunkte oder auch vereinbare Ereignisse :

$$P(A \text{ oder } B) = P(A) + P(B) - P(A \text{ und } B)$$
- Bei disjunkten / nicht vereinbaren Ereignissen : $P(A \text{ oder } B) = P(A) + P(B)$
- Bedingte Wahrsch. : $P(B \text{ unter Bed. } A) = P(A \text{ und } B) / P(A)$
- Multiplikationstheorem für abhängige Ereignisse :

$$P(A \text{ und } B) = P(A) * P(B \text{ unter Bedingung } A) = P(B) * P(A \text{ unter Bed. } B), \text{ bei}$$
 Umformung ergibt sich Theorem von Bayes
- Multiplikationstheorem für unabhängige Ereignisse :
 - $P(A \text{ und } B) = P(A) * P(B)$

Regeln :

- 1. Variationsregel : bei Münzwürfen (unabhängiges Ereignis) ergeben sich k^n verschiedene Möglichkeiten
- 2. Variationsregel : bei voneinander unabhängigen Einzelversuchen (z.B. 1 Münzwurf, 1 Würfelwurf, beides jeweils unabh. Ereignisse) ergeben sich $k_1 * k_2 * \dots * k_n$ Möglichkeiten
- Permutationsregel : bei Urnenversuch mit untersch. Ereignis ohne Zurücklegen (abhängiges Ereignis) ergeben sich $n!$ Möglichkeiten
- 1. Kombinationsregel : gleicher Versuch, nur gleich mehrere Kugeln gleichzeitig : ergibt $n! / (n-r)!$ Möglichkeiten
- 2. Kombinationsregel : gleicher Versuch, aber Reihenfolge egal (Lotto) ergibt $n! / (r! * (n-r)!)$ Möglichkeiten (auch n über r genannt)
- 3. Kombinationsregel : Einteilen der Kugeln in Gruppen, sollen als Gruppen gezogen werden, aber Reihenfolge egal (z.B. erst grüne, dann rote Kugeln) : $n! / (n_1! * \dots * n_k!)$ Möglichkeiten

Verteilungen :

- \bar{x} und s^2 einer empirischen Verteilung / Stichprobe werden durch μ (Erwartungswert) und σ^2 bei einer theoretischen Verteilung einer Zufallsvariablen / Population ersetzt, bei diskreter Verteilung ergibt sich $\mu = \sum x_i * p_i$ und $\sigma^2 = \sum p_i * (x_i - \mu)^2$
- Zufallsergebnisse in zwei Alternativen (Kopf/Zahl) verteilen sich nach der Binominalverteilung, ergibt sich mit $p^k * q^{n-k}$ (diskrete Verteilung)

Normalverteilung :

- stetig, unimodal, symmetrisch, glockenförmiger Verlauf, asymptotisch d. x-Achse nähernd
- zwischen den Werten $\bar{x} + s$ und $\bar{x} - s$ liegen 68% aller Werte, im Bereich $\bar{x} \pm 2s$ liegen 95,5%
- Binominalverteilung ist für $n \rightarrow \infty$ gleich der Normalverteilung

Standardnormalverteilung :

- um zwei Werte aus untersch. Normalverteilungen vergleichen zu können, müssen beide Verteilungen in die Standardnormalvert. überführt werden, indem die Abweichungen vom Mittelwert an der Unterschiedlichkeit aller Werte in der jeweiligen Verteilung relativiert werden. Abweichungen werden also durch die SD geteilt : $z_i = \frac{x_i - \bar{x}}{s}$
- eine Standardnorm.vert. hat Mittelwert 0 & Streuung 1

 χ^2 - Verteilung :

- ist Quadrat einer Standardnorm.vert : $z^2 = \chi^2$
- Summe zweier unabhängiger z^2 -Verteilungen ergibt χ^2_2 - Verteilung, usw.
- in Abhängigkeit von Anzahl der z^2 - Variablen haben χ^2 - Verteilungen untersch. Freiheitsgrade / df (degrees of freedom) n
- nähert sich mit wachsenden Freiheitsgraden (n) der Normalverteilung an
- Variationsbreite von 0 bis ∞

t-Verteilung :

- definiert durch $\frac{z}{\sqrt{\chi_n^2 / n}}$
- unterscheiden sich durch Anzahl der Freiheitsgrade (= Freiheitsgrade der χ^2 -Verteilung)
- nähert sich mit wachsenden Freiheitsgraden (n) der Normalverteilung an

F-Verteilung :

- definiert durch $F_{(n_1, n_2)} = \frac{\chi_{n_1}^2}{\chi_{n_2}^2} * \frac{n_2}{n_1}$
- unterscheiden sich durch Zähler- (n_1) und Nenner- (n_2) Freiheitsgrade
- Variationsbreite von 0 bis ∞

Stichprobe & Grundgesamtheit (Kap 3) :

- Inferenz- oder schließende Statistik ermöglicht Schlüsse von Stichprobe auf Grundgesamtheit / Population & überprüft Hypothesen
- Population sollte 100x so groß wie Stichprobe sein, daß erstere als unendlich gilt
- Zufallsstichprobe : zufällige Auswahl, fast nie möglich
- Klumpenstichprobe / Cluster Samples : leicht vorsortiert, z.B. alle Alkoholiker in Kliniken
- Proportional geschichtete Stichprobe : entsprechend der Merkmalsverteilung in Population zusammengestellt, z.B. 40% Singles, 60% Ehepaare

Verteilung von Stichprobenkennwerten :

- bei zwei Stichproben aus 1 Pop. : je weiter Mittelwerte der Stichpr. voneinander entfernt, desto geringer die vermutl. richtige Schätzung d. Populationsmittelwerts
- wenn unendlich viele Stichproben, dann Stichprobenkennwerteverteilung, je geringer die Streuung (= Standardfehler $\sigma_{\bar{x}}$), desto genauer Schätzung

Der Standardfehler $\sigma_{\bar{x}}$:

- Standardfehler verändert sich proportional zur Streuung in Population
- Stand.fehler verringert sich mit zunehm. Stichprobenumfang
- berechnet sich aus Varianz der Population : $\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}}$
- wenn diese unbekannt, dann Schätzung dieser aus Stichprobendaten, aber nicht als Durchschnitt der Stichprob.varianzen (keine erwartungstreue Schätzung), sondern durch Summe der quadrierten Abweichungen aller Meßwerte vom Mittelwert (Quadratsumme) geteilt durch n-1 (Freiheitsgrade der Varianz)
- jetzt Berechnung des (geschätzten) Stand.fehlers möglich

Zentrales Grenzwerttheorem :

- Stichprobenkennwerteverteilung für Mittelwerte geht mit wachsendem Stichprobenumfang in Normalverteilung über (ab Umfang 30) -> kann man sich gut vorstellen beim Beispiel des Würfels : wenn aus Grundgesamtheit 1-6 (Würfel) immer zwei gezogen werden (Stichproben mit Größe 2), davon der Mittelwert gebildet, ergibt sich Verteilung ähnlich der Normalverteilung
- für Varianzen ist Stichprobenkennw.vert. χ^2 -verteilt

Mittelwert der Stichprobenkennwerteverteilung der Mittelwerte :

- entspricht μ der Population bei hinreich. großen Stichprob.umfängen
- auch hier bei $\mu \pm \sigma_{\bar{x}}$ 68% und bei $\mu \pm 2 \sigma_{\bar{x}}$ 95,5%
mit $\sigma_{\bar{x}}$ = Streuung der Mittelw. der Stichproben / Standardfehler

Intervallschätzung :

\bar{x} -Werte-Bereich :

- es befindet sich also Mittelwert einer beliebigen Stichprobe mit Umfang $n \geq 30$ mit 95,5% Wahrsch. im Bereich $\mu \pm 2 \sigma_{\bar{x}}$ (eigtl. ist das falsch, ein Wert kann nur drin oder draußen liegen, also Wahrsch. entweder 1 oder 0 - ist aber nicht so wichtig).
- Wenn $\mu = 100$ und $\sigma_{\bar{x}} = 5$ ergibt sich Bereich von 90-110, der sog. \bar{x} -Werte-Bereich (in diesem liegen 95,5% aller Stichprobenmittelwerte, die von Stichproben aus dieser Population stammen)
- dies ist alles analog zu $z = \frac{\bar{x}_i - \mu}{\sigma_{\bar{x}}}$, für $z_{(\alpha/2)}$ wird +1 od. +2 eingesetzt, entspricht 68 bzw. 95,5% - natürlich sind auch 90% oder 95%- \bar{x} -Werte-Bereiche möglich
- ($\alpha/2$) weil zweiseitiges Intervall gesucht wird, also z.B. auf jeder Seite 2,25%
- Aufgabentypen hierzu :
 1. bestimmen Sie \bar{x} -Werte-Bereich (wie oben) (bei bekannter Pop)
 2. Es werden versch. Stichproben gezogen, wie hoch ist Wahrsch., daß Mittelwert einer Stichprobe über 105 liegt ? (bei bekannter Pop)
 3. Wie hoch ist Wahrscheinlichkeit, daß *ein* (nicht ganze Stichprobe) aus Pop gezogener Wert > 120 ist ? (für \bar{x} diesen Wert einsetzen) (b. bek. Pop)

Konfidenzintervall :

- genau umgekehrt : nicht welche Stichprobe gehört zur Population, sondern welche Populationen gehören zu Stichprobe
- K.-Intervalle sind Bereiche, in denen sich Pop-Parameter befinden, die als Erzeuger einer Stichprobe (mit xx% Wahrsch.) in Frage kommen
- hier nimmt man 95% oder 99% Wahrsch. (Konfidenzkoeffizient)
- der einzige Unterschied zw. \bar{x} -Werte-Bereich & K-Intervall ist also, daß bei ersterem die Pop gegeben ist & geprüft wird, welche Stichproben dazu passen (mögl. \bar{x} -Werte werden errechnet) & bei letzterem die Stichprobe bekannt ist & dazu passende Pops gesucht werden (mögl. μ -Werte gesucht)
- in Standardnorm.vert. liegt 95% der Fläche zwischen $z = \pm 1,96$ und 99% zwischen $z = \pm 2,58$
- gesucht sind nun die Grenzen des Konfidenzintervalls
- also überführen wir Stichprobenkennwertevert. in Standardnorm.vert. mit $z = \frac{\bar{x}_i - \mu}{\sigma_{\bar{x}}}$, wobei \bar{x}_i die Grenze des K.Interv. ist, die man mit Einsetzen der z-Werte für die gesuchte Wahrsch. leicht errechnen kann, Konfidenzintervall Δ_{crit} ist also $\mu \pm z_{(\alpha/2)} * \sigma_{\bar{x}}$ - da keine Pop-Parameter bekannt, schätzen wir μ durch \bar{x} und $\sigma_{\bar{x}}$ durch $\hat{\sigma}_{\bar{x}}$ (welcher sich ja über $\hat{\sigma}$ aus Stichprobendaten geschätzt wird)
- bei kleinen Stichproben (unter 30) sind aber die „z-Werte“ nicht standardnorm.verteilt, sondern t-verteilt (wenn Pop normalverteilt), also t-Werte aus Tabelle einsetzen, Freiheitsgrade = $n-1$ (da Summe der (unquadrierten) Abweichungen von Mittelwert 0 ergibt, können nur $n-1$ frei variieren)
- Aufgabentypen :
 1. Bestimmen Sie das 95%-Konfidenzintervall (bei bekannter Stichprobe)

Kriterien der Parameterschätzung :

- wenn wir Pop-Parameter aus Stichprobenkennwerten schätzen (z.B. den Erwartungswert μ der Pop aus dem Mittelwert \bar{x} der Stichprobe) müssen wir bestimmte Kriterien anlegen, damit diese Schätzung zulässig ist ...

Erwartungstreue :

- Ein statistischer Kennwert schätzt einen Pop-Parameter erwartungstreu, wenn das arithmetische Mittel der Kennwerteverteilung (wohlgemerkt : Kennwerte sind nicht nur Mittelwerte) bzw. deren Erwartungswert dem Pop-Parameter entspricht

Konsistenz :

- Von einem konsistenten Schätzwert sprechen wir, wenn sich ein statistischer Kennwert mit wachsendem Stichprobenumfang dem Parameter, den er schätzen soll, nähert

Effizienz :

- Je größer die Varianz der Stichprobenkennwerteverteilung, desto geringer die Effizienz des entsprechenden Schätzwertes

Exhaustivität :

- Ein Schätzwert ist exhaustiv oder erschöpfend, wenn er alle in den Daten einer Stichprobe enthaltenen Informationen berücksichtigt, so daß durch Berechnung eines weiteren statistischen Kennwertes keine zusätzlichen Informationen über den zu schätzenden Parameter gewonnen werden kann

Formulierung & Überprüfung von Hypothesen (Kap 4) :

- inwieweit können postulierte Eigenschaften der Pop (Theorie) durch Stichproben (Empirie) bestätigt werden ?

Unterscheidung inhaltliche Hypothesen & Erstellung statistische Hypothesen :

- Unterschiedshypothesen : Häufigkeits-, Mittelwertsvergleiche
- Zusammenhangshypos : Korrelationsrechnung

weitere Unterscheidung :

- gerichtete Hypos : z.B. besser als ..., hier noch Unterteilung in
 - spezifische Hypos : um (mind.) 10 Punkte besser als
 - unspezifische Hypos : irgendwie besser als
- ungerichtete Hypos : lediglich anders als ...
- statistische Hypos heißen dann H_1 : Korrelation > 0 oder $H_1 : \mu_0 > \mu_1$
- Nullhypothese : ist Negativhypothese, die die zur Alternativhypo komplementäre Aussage behauptet, z.B. Korrelation < 0 oder $\mu_0 < \mu_1$
- Nur wenn Empirie mit 90, 95 oder 99% nicht mehr mit Nullhypo erklärt werden kann, darf sie zugunsten der Alternativhypo verworfen werden

Signifikanz :

- nicht-signifikantes Ergebnis heißt nicht, daß H_0 richtig, z.B. bei Verfehlung des 5%-Niveaus um 1% ist H_0 nur mit Wahrsch. von 6% richtig !!
- jede Alternativhypothese wird signifikant, wenn Stichprobe genügend groß, da die Nullhypo $\mu_0 = \mu_1$ eine Fiktion ist, in der Realität werden zwei Pops immer voneinander abweichen & je größer die Stichprobe, desto realistischer wird sie die Pop mit μ_1 wiedergeben (bei $n \rightarrow \infty$ ist sie identisch), sich also immer deutlicher von der Pop mit μ_0 unterscheiden
- deswegen Einführung der Effektgröße ε : Differenz $\mu_1 - \mu_0$ relativiert an σ (geteilt durch σ), man muß sich also vorher genau überlegen, wie groß μ_1 sein soll
- Effektgrößen verschiedener Untersuchungen lassen sich dann vergleichen & die Aussagekräftigkeit untersuchen. Effektgröße gibt also an, wie weit - in einem spez. Test - μ_1 von μ_0 entfernt sein sollte, um von einem praktisch bedeutsamen Effekt sprechen zu können

Alpha- & Beta-Fehler :

- α -Fehler : H_0 gilt, ich entscheide aber H_1
- β -Fehler : H_1 gilt, ich entscheide aber H_0
- sind beides bedingte Wahrsch., also Wahrsch. unter Bed. H_0 oder H_1

α -Fehler :

- Mittelwerte der Stichproben verteilen sich nach zentr. Grenzwerttheorem normal mit Mittelwert μ_0 , der wahren Pop-Erwartungswert entspricht
- also läßt sich über z-Transformation z.B. ein 95% oder 99% Intervall erstellen, es ergibt sich Signifik.niv. von 5%(sign.) o.1%(sehr sign.), mittels Formel

$$z = \frac{\bar{x}_i - \mu_0}{\sigma_{\bar{x}}} \text{ (oder mit geschätztem Standardfehler ...)}$$

- Irrtumswahrsch. α ist immer vorhanden, verringert sich bei ...
 1. größer werdender Diskrepanz $\bar{x} - \mu_0$
 2. Vergrößerung des Stichprobenumfangs : wenn so groß wie Pop, dann sichere Entscheidung
 3. kleiner werdender Popstreuung $\hat{\sigma}$

β -Fehler :

- Ermittlung des β -Fehlers durch Verteilung der Mittelwerte von Stichproben aus Pop μ_1
- wenn jedoch Hypo unspezifisch, ist μ_1 nicht bekannt & β -Fehler kann nicht bestimmt werden
- wenn μ_1 jedoch festgelegt, läßt sich dann mit z-Transformation $z = \frac{\bar{x}_i - \mu_1}{\sigma_{\bar{x}}}$ (oder mit geschätztem Standardfehler) der β -Fehler ermitteln
- β -Fehler sollte mind. 20% betragen, wenn auch keine genauen Konventionen, bei Bestätigung Nullhypo umgekehrt : α mind. 20%, β 5% oder 1%

Power :

- α und β verändern sich gegenläufig, wenn α klein, ist β groß und damit die Power / Teststärke / Trennschärfe ($1-\beta$) gering. Die Power gibt an, mit welcher Wahrsch. ein Signifikanztest zugunsten einer spezifischen Alternativhypo entscheidet
- sie vergrößert sich mit
 1. wachsender Differenz von μ_0 und μ_1 (also auch mit wachsend. Effektgröße)
 2. wachsendem Stichprobenumfang (da dann der Standardfehler kleiner ist)
 3. sinkender Pop-Streuung
- bei gerichteter Hypo hat einseitiger Test größere Power als zweiseitiger

Verfahren zur Untersuchung von Unterschiedshypothesen (Kap 5) :

- hier werden normalerweise zwei Stichproben miteinander verglichen, nicht eine Stichprobe mit einer Population, wie bisher
- es gibt aber auch den Vergleich eines Stichprobenmittelwerts mit einem Populationsmittelwert, man überprüft hier, ob Stichprobe zu bekannter Pop gehört

Verfahren für Intervalldaten :*Vergleich eines Stichprobenmittelwerts mit einem Populationsmittelwert (s.o.) :*

- $H_0 : \mu_0 = \mu_1$ $H_1 : \mu_0 \neq \mu_1$
- hierzu vorgehen wie bei Berechnung von xx%-Intervallen, identische Formel :

$$z = \frac{\bar{x}_i - \mu_0}{\sigma_{\bar{x}}} \quad (\text{od. m. gesch. Stand.fehl.}) \quad \text{Hier sind jetzt allerdings alle Größen außer}$$

z bekannt, also $z_{\text{empirisch}}$ ausrechnen, vergleichen mit z_{kritisch} bei z.B. 95% Wahrscheinlichkeit (+-1,96 - bei zweiseitigem Test), wenn $z_{\text{emp}} > z_{\text{krit}}$, dann H_0 verwerfen zugunsten H_1 : Stichprobe gehört nicht zu Population mit μ_0 , bzw. die Pop 1 aus der die Stichprobe stammt ist nicht mit Pop 0 identisch ...

- beachten, ob ein (α)- oder zweiseitig ($\alpha/2$) & b. kleinen Stichpr. t-verteilt mit $n-1$ df

Vergleich zweier Stichprobenmittelwerte aus unabhängigen Stichproben (t-test) :

- 2 unabh. Stichproben aus 2 Populationen
- t-test für unabh. Stichproben überprüft Nullhypo, daß die beiden Stichproben aus Pops stammen, deren Erwartungswerte μ_1 und μ_2 gleich sind : $H_0 : \mu_1 - \mu_2 = 0$
- ungerichtete $H_1 : \mu_1 - \mu_2 \neq 0$
- bei unendlich vielen Stichproben aus d. 2 Pops erhalten wir Verteil. d. Differenzen d. Stichprobenmittelwerte, bei Gültigkeit von H_0 hat diese Erwartungswert von 0
- zu untersuchende Frage : Wie weit weichen d. Stichprobendifferenzen (der beiden gegebenen \bar{x} -Werte) von d. Pop-Diff. ab (μ -Werte, nicht gegeben) ?
- deswegen t-Verteilung definiert durch Differenz d. beiden \bar{x} -Werte (Stichproben) minus der Differenz der beiden Erwartungswerte μ (Population) durch geschätzte Streuung / Standardfehler der Mittelwertedifferenzen (letzteres wg. Relativierung, dieser Standardfehler wird mittels Rechnung aus Stichprobenwerten geschätzt)
- da Differenz der Erwartungswerte nach Nullhypo = 0 ist, fällt sie weg, nachschauen in t-Verteilung, für größere Stichproben auch in Normalverteilung
- bei $n \geq 30$ der Stichproben : Kennwerteverteilungen normal, also auch deren Differenzverteilung, bei kleineren Stichpr. $n_1+n_2 < 50$) : t-verteilt mit $n_1 + n_2 - 2$ df
- Voraussetzungen :
 4. bei kleinen Stichproben : Populationen normalverteilt (goodness-of-fit test)
 5. Varianzen in Pops homogen (F-Test)
 6. Stichproben voneinander unabhängig

Vergleich zweier Stichprobenmittelwerte aus abhängigen Stichproben (t-test) :

- wenn Stichproben paarweise zugeordnet sind (z.B. Partner-Befragung), sind sie abhängig und auch, wenn mit Meßwiederholung gearbeitet wird
- $H_0 : \mu_1 - \mu_2 = 0$, ungerichtete $H_1 : \mu_1 - \mu_2 \neq 0$
- es wird berücksichtigt, daß Varianz der 1. Stichprobe die Varianz der 2. beeinflusst (und/oder umgekehrt)
- wenn wir Standardfehler der Differenzen wie vorher aus Differenzen der Standardfehler der beiden Mittelwerte schätzen, so werden Unterschiede zwischen Personen, die z.B. vor & nach Schulung (Meßwiederholung) bestehen, doppelt berücksichtigt, da sie beide Standardfehler beeinflussen
- verhindern läßt sich dies über Betrachtung lediglich der zusammengehörenden Meßwertpaare, für jedes Meßwertpaar bilden wir also Differenz d_i
- und bilden dann arithmetisches Mittel aller d_i -Werte
- es interessieren also nicht die Verteilung der Differenzen der Mittelwerte (unabh.t-test), sondern die Verteilung der Mittelwerte der Differenzen
- jetzt Standardfehler der Verteilung der Mittelwerte von Differenzen schätzen
- t-Verteilung ergibt sich mit Mittel aller d_i - Werte minus μ_d (nach Nullhypo = 0, fällt also weg), durch Streuung / Standardfehler der Verteilung der Mittelwerte der Differenzen
- Frage also : Unterscheidet sich Mittelwert der gegebenen Stichprobendifferenzen von (eigtl. nicht gegebener) Popdifferenz ($\mu_1 - \mu_2 = \mu_d$)?
- Ablesen in t-Tabelle, Freiheitsgrad : $n-1$, wobei n = Anzahl Meßwertpaare, bei größeren Stichproben auch Normalverteilung
- Voraussetzungen :
 1. bei kleinen Stichproben : Grundgesamtheiten normalverteilt

Vergleich zweier Stichprobenvarianzen aus unabhäng. Stichproben (F-Test) :

- Nullhypo sagt, daß sich die Varianzen der zugrundeliegenden Pops nicht unterscheiden, wird anhand der Stichprobenvarianzen (nur die sind ja gegeben) überprüft
- $H_0 : \sigma^2_1 - \sigma^2_2 = 0$ $H_1 : \sigma^2_1 - \sigma^2_2 \neq 0$
- wir nehmen aber - da keine Pop-Parameter bekannt - die geschätzten σ -Werte, es ergibt sich also $F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}$ mit n_1-1 Zähler- und n_2-1 Nennerfreiheitsgraden
- bei einseitigem Test größere Varianz im Zähler - wird fast immer einseitig verwendet, in Tabelle auch nur einseitige Werte
- wenn H_0 gültig, ist $F = 1$ (Erwartungswert), kann nicht negativ werden, da Varianzen quadriert
- wird meist benutzt bei Varianzanalyse & bei Überprüfung von Varianzhomogenität
- Voraussetzung :
 1. Grundgesamtheit normalverteilt
 2. Stichproben unabhängig

Verfahren für Ordinaldaten / Non-parametrische / verteilungsfreie Verfahren :

- sind „verteilungsfrei“, vergleichen nur Relationen / Ordinaldaten = Rangdaten

Unabhängige Stichproben : U-Test :

- vergleicht Stichproben hinsichtlich ihrer zentralen Tendenz (analog zum Mittelwert)
- geht problemlos auch mit ungleich großen Stichproben
- wir bilden für jede der zwei Stichproben die Summe der Rangplätze (T_1 bzw. T_2) & deren Mittel (R_1 & R_2)
- wenn die beiden Mittelwerte sich unterscheiden, ist dies Hinweis auf Gruppenunterscheidung
- jetzt Prüfgröße U : errechnet aus Anzahl der numerisch größeren Rangplätze in der einen Gruppe im Vergleich zur anderen : also Rangplatz 1 der G.1 vergleichen mit allen Rangplätzen der G.2, dann Rangplatz 2 der G.1 vergl. usw
- die Größe U' ist analog dazu : sie gibt Anzahl der numerischen Rangplatzunterschreitungen (also besseren Ränge) wieder
- unter H_0 (keine Unterscheidung der Pops) erwarten wir U-Wert gleich $\frac{U_1 * U_2}{2}$
- alle denkbaren U-Werte sind um diesen μ_U symmetrisch verteilt, es ergibt sich also auch eine Streuung σ_U der U-Werte-Verteilung
- bei n_1 oder n_2 größer 10 sind die U-Werte normalverteilt, mit μ_U und σ_U läßt sich nun ein z-Wert errechnen mit $z = \frac{U - \mu_u}{\sigma_u}$, in Standardnormalverteilungstabelle mit kritischen Wert bei xx%-Niveau auf Signifikanz prüfen

Der H-Test :

- erweitert den U-Test bei mehr als zwei Stichproben

Abhängige Stichproben : Wilcoxon-Test :

- nur so zur Info, ist halt der Test für abh. Stichproben bei Ordinalskalenniveau

Verfahren für Nominaldaten (χ^2 - Verfahren) :

- auch verteilungsfrei / non-parametrisch, da Häufigkeiten
- nicht nur für Nominaldaten (Häufigkeitsunterschiede), aber wenn für höhere Skalenniveaus angewandt, gehen Infos verloren

Vergleich der Häufigkeiten eines zweifach gestuften Merkmals (eindimension. χ^2 -T.) :

- z.B. Verhältnis Männer / Frauen
- Aufstellen der erwarteten Häufigkeiten f_e durch Mitteln der beiden beobachteten Häufigkeiten, wenn H_0 , daß 50 zu 50-Verhältnis gilt. Wenn andere H_0 , werden die dann verschiedenen f_e 's über $n * p_i$ berechnet

- Verteilung ergibt sich mit $\sum_{j=1}^2 \frac{(f_{b(j)} - f_{e(j)})^2}{f_{e(j)}}$

- dieser Ausdruck = 0, wenn $f_e = f_b$, wenn also H_0 zutrifft
- Vergleich mit kritischem χ^2 - Wert, wenn $\chi^2_{emp} > \chi^2_{krit}$, dann signifikanter Unterschied, H_0 verwerfen (df = 1, da Summe der beobachteten wie auch erwarteten Häufigkeiten jeweils n ergeben muß)
- generelle df-Berechnung : Anzahl der Summanden abzügl. der Bestimmungsstücke für Berechnung der erwarteten Häufigkeiten, hier 1

Bestimmungsstück, nämlich Stichprobenumfang n (Summen der erwart. & beobacht. Häufigkeiten ergibt jeweils n)

- alle χ^2 - Verfahren vergleichen erwartete & beobachtete Häufigkeiten, wobei die erwartete Häufigkeit von Nullhypo abhängt
- alle χ^2 - Verfahren testen einseitig & normalerweise ungerichtet
- bei gerichteten Hypos mit doppeltem Niveau, dies nur möglich wenn ein df vorhanden, auch durch Standardnorm.vert. möglich, da $z^2 = \chi^2$, höhere Teststärke als ungerichtet
- Der χ^2 - Wert läßt sich auch über die Binominalverteilung errechnen (wenn erwartete Häufigkeiten kleiner als 10, ist dies nötig)

Vergleich der Häufigkeiten eines k -fach gestuften Merkmals (eindimension. χ^2 -T.):

- fast wie bei zweifach, einfach einmal mehr f_e 's und f_b 's errechnen, gleiche Formel
- Summe der erwarteten Häufigkeiten muß der der beobachteten entsprechen, deswegen $k-1$ df's
- bei der Frage, ob sich Variable 1 von den anderen unterscheidet, wird von den Variablen 2 bis k der Durchschnitt gebildet (fungiert als f_e) & mit Variable 1 (als f_b) über den χ^2 - Test verglichen

χ^2 - Test auf andere Verteilungsformen (eindimension. χ^2 -T.):

- wenn H_0 nicht behauptet, daß gleichverteilt (z.B. bei 4-fach gestuftem Merkmal mit $P(1) = 0,25$; $P(2) = 0,25$... $P(4)$), ergibt sich bei steigender Variablenanzahl eine theoretische Verteilung
- Vergleich einer empirischen mit einer theoretischen Verteilung heißt Goodness-of-fit-test
- man kann so z.B. prüfen, ob sich eine empirische Verteilung normalverteilt oder z.B. nach Poisson-Verteilung
- dient also z.B. der Überprüfung der Voraussetzung, daß Grundgesamtheit normalverteilt (H_0 : normalverteilt, H_1 : nicht normalverteilt)
- wenn χ^2 -Wert auf 5%-Niveau signifikant ist, heißt das, daß Wahrsch. für Normalverteilung kleiner als 5%, wenn nicht signifikant, dann lediglich Aussage, daß mit mehr als 5% Wahrsch. normalverteilt - mehr nicht- wir wollen aber H_0 „beweisen“ (sonst will man normalerweise ja die H_1 „beweisen“), deshalb
- beta-Fehler mögl. klein halten, nicht alpha, da aber Hypo in diesem Fall nicht spezifisch, kann beta-Fehler nicht bestimmt werden, also halt einfach alpha-Fehler vergrößern, z.B. auf 25% & damit dann testen (auf 25%-Niveau also)
- wenn die H_1 dann immer noch nicht signifikant ist, haben wir H_0 gesichert

Vergleich der Häufigkeiten von zwei alternativen oder auch $k \cdot l$ -gestuften Merkmalen (4-Felder- χ^2 oder zweidimensionaler χ^2 -Test) :

- wenn n Beobachtungen nicht nur eine Alternative (m/w) zugeordnet wird, sondern zwei (Brille/nicht-Brille), haben wir 4-Felder-Kontingenz-Tafel
- wenn nach Nullhypo Wahrsch. vorgegeben ist,
 - lassen sich f_e 's wie normal berechnen
 - dann normale χ^2 -Formel, nur über zwei Summenzeichen durchlaufend
 - nur ein Bestimmungsstück (Summe n), also $df = 2 \cdot 2 - 1 = 3$, allg. $k \cdot l - 1$
- wenn Wahrsch. nicht durch H_0 bestimmt
 - können sie errechnet werden, mit Annahme der stochastischen Unabhängigkeit der jeweiligen Merkmalsalternativen, über Multiplikationstheorem für unabh. Ereignisse :
 - die allgemeine Formel dafür : (Zeilensumme * Spaltensumme) / n
 - da hier die Randsummen der beobachteten Werte mit den Randsummen der erwarteten Werte übereinstimmen müssen, kann nur 1 Wert frei variieren (1 df), allg. $(k-1) \cdot (l-1)$
- 4-Felder- χ^2 (oder $k \cdot l$ - χ^2) läßt sich auch mit Prozentwerten durchführen
- *mehrdimensionaler χ^2 -Test* (mit m alternativen (also 2fach gestuften) oder mehrfach gestuften Merkmalen) läßt sich auch durchführen...

Voraussetzungen für alle aufgeführten χ^2 -Tests :

1. jede untersuchte Einheit eindeutig einer Kategorie zugeordnet
2. einzelne Beobachtungen voneinander unabhängig
3. nicht mehr als 20% der erwarteten Häufigkeiten in den Kategorien kleiner 5

Verfahren zur Untersuchung von Zusammenhangshypothesen (Kap 6) :

Regressionsgleichung :

- durch Korrelationen Vorhersagen möglich, allerdings nur stochastische Zusammenhänge, keine funktionalen, Korrelationskoeffizient r zwischen $+1$
- Gleichung zur Vorhersage = Regressionsgleichung
- nicht im Sinne von Kausalbeziehungen interpretierbar, Kausalität läßt sich korrelationsstatistisch nur widerlegen, nicht beweisen, lediglich notwendige, aber keine hinreichende Bedingung für Kausalität
- in meisten Regressionsgleichungen wird mit leicht meßbarer Variable (Prädiktorvariable) eine schwer meßbare vorhergesagt (Kriteriumsvariable), ähnlich unabhängiger & abhängiger Variable, aber ohne Kausalbeziehung
- einfachste Beziehung ist linear : $y = bx + a$
- Gerade durch 2 Bestimmungsstücke (Steigung & Höhenlage oder 2 Punkte) eindeutig festgelegt
- Ergebnis einer empirischen Untersuchung ist Punkteschwarm, Regressionsgleichung ist die, die diesen am Besten wiedergibt
- gesucht sind die Koeffizienten a & b
- zwischen durch Gerade vorhergesagten (\hat{y}_1) & tatsächlichem Wert (y_1) besteht Diskrepanz $y_1 - \hat{y}_1$ (Regressionsresiduum)
- diese gibt Vorhersagefehler für jede VP_i an, gesucht also Gerade, die die Summe der *quadrierten* (da sonst evtl. zu Null addiert) Diskrepanzen minimiert

- wichtig : nicht Lote zur Geraden minimieren, sondern Abstand in y-Richtung, damit für Vorhersage x auf y optimiert, deswegen für y auf x auch andere Koeffizienten a & b
- Koeffizienten lassen sich durch Differentialgleichung errechnen
- die beiden Regressionsgraden (x auf y, y auf x) schneiden sich, dort liegen Mittelwerte \bar{x} und \bar{y}
- daraus folgt : bei z-transformierten Variablen ($\bar{x} = \bar{y} = 0$, $s_x = s_y = 1$) liegt Schnittpunkt im Ursprung des Koordinatensystems
- es gibt natürlich auch non-lineare Regressionsgleichungen (also keine Geraden)

Multiple Regression (passt hier zwar nicht so ganz rein, fragt aber zumindest Rietz gerne laut Protokoll - obwohl erst Kap. 13 - haben wir auch nie behandelt):

- mehrere Prädiktoren, ein Kriterium
- mehrere b-Gewichte : negative b's bedeuten eine Abnahme des Kriteriumswertes, wenn der zum b dazugehörige Prädiktorwert steigt, positive b's bedeuten das Gegenteil (Zunahme Kriterium bei Zunahme Prädiktor)
- die b-Gewichte kann man nur dann interpretieren, wenn die Prädiktoren statistisch unabh. sind, d.h. nicht miteinander korrelieren, was sehr selten vorkommt
- Korrelationskoeffizient R nimmt nur positive Werte zw. 0 und 1 an
- mehr Infos im genannten Methodenskript

bei nominalskalierten Daten ... (Dummyscodierung)

- wird die nominalskalierte (od. auch kategorisierte) Variable x (z.B. Geschlecht : m/w) mit 0 & 1 kodiert, heißt dann Indikatorvariable
- Dummyscodierung :
- bei mehrstufigen nom.sk. Variablen werden bei z.B. 4 Parteipräferenzen 3 Indikatorvariablen erstellt, Leute die die Partei 1 präferieren, kriegen auf den 3 Indikatorvariablen $x_1 - x_3$ die Werte 1-0-0, Partei 2-Leute die Werte 0-1-0 usw.
- Leute mit Partei 4 haben automatisch 0-0-0, die 4.Indikatorv. erübrigt sich also
- Regressionsgerade mit $b_1 * x_{1m} + b_2 * x_{2m} \dots + a$
- für Partei4-Leute ergibt sich also als Vorhersage-Wert a (da 0-0-0 eingesetzt)
- beste Vorhersage für diese Leute wäre aber Mittelwert von y_4 also ist $a = \bar{y}_4$
- für Partei1 ergibt sich Vorhersage \bar{y}_1 , dies ist aber gleich $b_1 + a$ (da 1-0-0 eingesetzt), da $a = \bar{y}_4$ ergibt sich $b_1 = \bar{y}_1 - \bar{y}_4$
- für die anderen Gruppen analog
- ein b_i -Gewicht errechnet sich also als Differenz der Mittelwerte der Gruppe i und der Referenzgruppe (immer die 0-0-0-codierte)
- Dummyscodierung läßt sich auch bei Kovarianzanalyse mit kategorisierter (nicht intervallskalierter) Kovariable einsetzen

Kovarianz :

- der Zähler des Regressionskoeffizienten b dividiert durch n ergibt Kovarianz, bezeichnet Grad des Zusammen-Variierens von x und y , ist durch Mittelwert aller Produkte von jew. zugehörigen Abweichungen gekennzeichnet : $\frac{(x_i - \bar{x}) * (y_i - \bar{y})}{n}$
- wenn bei einem Werte-Paar beide Werte hoch od. niedrig über/unter Durchschnitt liegen ergibt sich hohe pos. Kovarianz, wenn einer niedrig, einer hoch, ergibt sich hohe negative Kovarianz, wenn keine Systematik, dann keine Kovarianz (=0)
- bei Kovarianz von Null erhalten wir die Regressionsgraden $\hat{y} = \bar{x}$ und $\hat{x} = \bar{y}$, also Geraden parallel zur x - bzw. y -Achse, die deshalb senkrecht aufeinander stehen (90°), stochastisch unabhängig
- die maximale Kovarianz beträgt $s_x * s_y$, dort fallen beide Geraden zusammen (0°), je kleiner der Winkel also, desto größer Kovarianz
- Kovarianz ist ja durch Anzahl & Größe der gleichgerichteten Abweichungen bestimmt, also kann ein hohes negatives Abweichprodukt durch mehrere kleine positive kompensiert werden
- a und b streuen natürlich, je nach Stichprobe, um die wahren Pop-Parameter α und β
- man kann aber zeigen, daß a und b erwartungstreue Schätzungen für die Pop-Parameter (α und β) sind. Wenn man die Stichprobenkennwerteverteilung für a und b berechnen will, müssen sich die beiden untersuchten Merkmale in der Pop allerdings bivariat normalverteilen (sieht aus wie eine Art 3D-Normalverteilung)
- jede Menge weitere Vorauss., aber Monte-Carlo-Studien zeigen, daß sehr robust
- Der Standardschätzfehler : kennzeichnet Streuung der y -Werte um die Reg-gerade & ist damit Maßstab für Genauigkeit der Reg-vorhersagen, berechnet sich über die einzelnen Regressionsresiduen

Korrelation (Produkt-Moment-Korrelation oder Bravais-Pearson-Korrelation) :

- Kovarianz hat Nachteil : abhängig von Maßstab der Variablen bzw. von deren Varianz (max. Kovarianz beträgt ja auch $s_x * s_y$), wenn Werte verdoppelt, verdoppelt sich auch Abweichung & Varianz
- also nur sinnvoll, wenn fester Maßstab, z.B. Meter
- deshalb relativiertes Maß : Korrelationskoeffizient $r : r = \frac{\text{cov}(x, y)}{s_x * s_y}$
- r entspricht der Kovarianz zweier z -transformierter Variablen, ist gegenüber linearer Transformation invariant (z.B. Verdoppelung der Variablen)
- bei minimalen Reg-Residuen & damit minimalem Standardschätzfehler, also keiner Varianz der y -Werte um Reg-gerade ist $r = \pm 1$ (negativ bei Reg-gerade mit neg. Steigung und positiv bei pos. Steigung), beide Reg-geraden (die für x und y) sind identisch
- da die \hat{y} -Werte linear aus x -Werten hervorgehen und r gg.über Lineartransf. invariant, muß gelten : Korrelation zw. x und y ist gleich der zwischen y und \hat{y}

Überprüfung von Korrelationshypotesen :

- man unterscheidet empirische Korr. r der Stichprobe & die wahre Korr. ρ (spricht man - glaube ich - „rho“) der Pop
- bei Nullhypo $\rho = 0$ & $H_1 : \rho \neq 0$ wird mittels speziellen Signifikanztest über t-Verteilung geprüft
- bei einer spezifischen Nullhypo wie $\rho = a$ (ungleich 0) und der Alternativ-Hypo $\rho \neq a$ muß davon ausgegangen werden (bei Gültigkeit der Nullhypo), daß sich die Korrelationen bei unendlich vielen Stichproben links ($\rho < 0$)- oder rechtssteil ($\rho > 0$) verteilen (= Stichprobenkennwerteverteilung), d.h. die Werte müssen erst mit der Fisher Z-Transformation „behandelt“ werden - dann verteilen sie sich annähernd normal (s.u.)
- nach dieser Fisher's Z-Transformation lassen sich mit den transformierten Werten Hypos fast aller Art untersuchen, meist durch Einsetzen der Werte in Standardnorm.vert.-z-Gleichung (also nochmal, diesmal „normal“ z-transform.), dann z-Wert (Stand.norm.-z !) in Tabelle vergleichen usw.
- manchmal auch über χ^2 -Vert. mit den Fisher-Z-transformierten Werten testen (bei k unabhängigen Stichproben)
- es lassen sich natürlich auch Konfidenzintervalle der Fisher-Z-Werte - analog zur „normalen“ Konfidenzintervallberechnung - bilden
- Fishers Z-Transformation (hat nichts mit „normaler“ z-Transformation zu tun) :
- wenn in Pop zwischen zwei Merkmalen ein Zusammenhang mit Korrelation ungleich Null, ergibt sich - wie schon erwähnt - bei (unendlich) vielen Stichproben eine rechts- oder linkssteile Verteilung der Korrelationen
- diese läßt sich wieder in Normalverteilung transformieren durch Fishers Z-Transf. - je weiter ρ von +-1 entfernt & je größer n, desto normaler
- weiterhin sind die transformierten Werte Teil einer Verhältnisskala (fester Nullpunkt), man kann nämlich bei Korr. nicht davon ausgehen, daß $r = 0,4$ halb so groß wie $r = 0,8$
- wenn transformiert zeigt sich : Zuwachsraten i. oberen Korr.-Bereich bedeutsamer
- auch Mittelwerte & Varianzen von Korr. sind nicht interpretierbar, wenn nicht transformiert
- Bei Mittelwertbildung von z.B. 3 Korrelationen also erst transformieren, dann Mittel der Fisher-Z-Werte bilden, dann dieses Mittel rücktransformieren : dabei werden höhere Korr. stärker berücksichtigt als niedrige

Korrelation für zwei dichotome Variablen mittels Φ -Koeffizient (sprich Phi) :

- Wenn beide Merkmalsausprägungen d. beiden Variablen jew. mit 0 und 1 kodiert (z.B. w/m, Brille/Nicht-B.), kann man (Produkt-Moment-)Korrelation über diese sich dann ergebenden Meßreihen durchführen & erhält den Φ -Koeffizient mit $\sqrt{\chi^2/n}$
- enger Zusammenhang mit 4-Felder- χ^2 , deshalb Signifikanzprüfung mittels diesem: $\chi^2 = n * \Phi^2$ mit $df = 1$
- Φ -Koeffizient nur im üblichen Wertebereich von -1 bis +1, wenn Aufteilung der Stichprobe in Alternative von x der in die Alternative von y entspricht, also bei identischen Randverteilungen, ansonsten geringerer Φ_{maximal} (mit Formel berech.)
- gelegentlich Aufwertung eines empirischen Φ -Wertes am Φ_{max} , um vergleichbar mit Produkt-Moment-Korrelation zu machen, aber fragwürdig, da auch P-M-Korr. nur bei identischen Randverteilungen Wertebereich von -1 bis +1 aufweist

Einfaktorielle Varianzanalyse (Kap 7) :

	i	i	i	
m				
m				
	\bar{A}_i	\bar{A}_i	\bar{A}_i	\bar{G}

m : 1 ... n

i : 1... p

- Hypothesen :
- $H_0 : \mu_1 = \mu_2 = \mu_3 = \dots \mu_p$
- $H_1 : \mu_i \neq \mu_i'$
- wie t-test oder χ^2 , aber Zusammenwirken & gegenseitige Beeinflussung mehrerer Variablen berücksichtigt
- eigtl. alle Verfahren Varianzanalysen, da alle Unterschiedlichkeit (Varianz) von VPN in Bezug auf abhängige Var. auf eine od. mehrere unabhängige Variable zurückführen
- einfaktorielle VA überprüft eine p-fach gestufte Variable / Faktor / Treatment
- Faktor A in p Stufen, Mittelwert unter einer Faktorstufe \bar{A}_i , Gesamtmittelwert \bar{G}
- warum nicht mehrere t-tests ? bei 100 t-tests mit $\alpha=0,5$ werden ca. 5 zufällig signifikant
- wenn aber nur 2 Stichproben bringen VA & t-test identische Ergebnisse
- Ansatz der VA : zu welchem Anteil kann Varianz auf Treatment zurückgeführt werden ?

Quadratsummenzerlegung :*Totale Quadratsumme / Totale Varianz :*

- Gesamtvarianz $\hat{\sigma}_{tot}^2$ aller Meßwerte ermitteln als Schätzwert für Pop-varianz, ergibt sich als Summe der quadrierten Abweichungen aller Werte vom Gesamtmittelwert (= totale Quadratsumme QS_{tot}) dividiert durch Freiheitsgrade der Varianz ($p \cdot n - 1$) - > die Varianz wäre als $QS/p \cdot n$ (oder N, halt Gesamtzahl aller Meßwerte) definiert, ist aber eine Schätzvarianz & die ist ja anders definiert (siehe S.4)
- $df_{tot} = p \cdot n - 1$

Treatmentquadratsumme :

- Anteil der QS_{tot} , der auf Treatment zurückgeht
- wenn Unterschiedlichkeit der Werte ganz von Treatment abhängig, dann dürften sich Meßwerte der Personen in einem Treatment nicht unterscheiden
- als beste Schätzung für Wirkung des Treatment nehmen wir also Durchschnittsmeßwert im Treatment (\bar{A}_i), jeder A_i - Wert wird also durch \bar{A}_i ersetzt
- jetzt Addition von jeder quadrierten ($\bar{A}_i - \bar{G}$)-Differenz (nicht jeden $(\bar{A}_i - \bar{G})^2$ -Wert nur einmal nehmen, sondern n mal !), ergibt QS_{treat}
- df : Werte innerhalb eines Treatments durch \bar{A}_i festgelegt & von den p \bar{A}_i 's können nur p-1 variieren, da sie \bar{G} ergeben müssen, also $df_{treat} = p - 1$
- die geschätzte Varianz $\hat{\sigma}_{treat}^2$ ergibt sich also mit $QS_{treat} / p - 1$

Fehlerquadratsumme :

- restlicher Varianzanteil, alle Unterschiede, die nicht auf Treatment zurückzuführen sind
- manifestiert sich in Unterschiedlichkeit der Meßwerte innerhalb des Treatments
- wir ziehen alle Werte jeweils vom Gruppenmittelwert \bar{A}_i ab (entspricht Regressionsresiduen) und quadrieren diese Differenz, dann Summation in dieser Spalte & fertig ist $QS_{\text{Fehler}(i)}$, es ergeben sich also so viele einzelne QS_{Fehler} wie Treatments
- $df_{(i)}$: Summe der (unquadrierten) Abweichung von \bar{A}_i muß Null sein, also $n-1$
- bei Berechnung der einzelnen Fehlervarianzen $1 - p$ ($QS_{\text{Fehler}(i)}$ durch $df_{\text{Fehler}(i)}$) muß sich Homogenität dieser herausstellen, d.h. eine nichtsignifikante Unterscheidung (da Störbedingungen bei allen Treatments gleich sein müssen)
- Gesamt- QS_{Fehler} erhält man durch Addition der einzelnen QS_{Fehler} , die Gesamt-Fehler-Varianz $\hat{\sigma}_{\text{Fehler}}^2$ erhält man durch Teilen von Gesamt- QS_{Fehler} durch Summe der df 's, also Gesamt- $df = p \cdot (n-1)$
- die Quadratsummen Fehler & Treatment addieren sich zu QS_{tot}
- die df 's addieren sich ebenso
- nicht additiv verhalten sich hingegen die Varianzen

Varianzaufklärung :

- ergibt sich durch $(QS_{\text{treat}} / QS_{\text{tot}}) * 100\%$
- besagt, daß XX % der Varianz durch Treatment aufgeklärt (erklärt) werden
- Quotient $QS_{\text{treat}} / QS_{\text{tot}}$ heißt auch Korrelationskoeffizient η^2 (sprich eta), enthält alle auf die verschiedenen Trends (s.u.) zurückgehenden Zusammenhänge auf Stichprobenebene
- auf Populationsebene ist $\hat{\omega}^2$ (sprich omega) das Äquivalent (allerdings nur geschätzt aus Stichprobendaten)

Signifikanztest :

- kommt die Varianzaufklärung zufällig zustande oder nicht ?
- bei Gültigkeit von H_0 (die Pops stimmen überein bzw. deren μ 's sind gleich) gilt, daß die Treatmentvarianz $\hat{\sigma}_{\text{treat}}^2$ eine erwartungstreue Schätzung der Fehlervarianz $\hat{\sigma}_{\text{Fehler}}^2$ ist :
 - H_0 lautet nämlich $\mu_1 = \mu_2 = \mu_3 = \dots \mu_p$
 - als Schätzwert für die Pop- μ 's nehmen wir die Faktorstufenmittelwerte \bar{A}_i
 - diese streuen aber - auch bei Gültigkeit von H_0 um \bar{G}
 - bei H_0 -Gültigkeit stammen nun die einzelnen Werte unter den Faktorstufen aus der gleichen Pop, d.h. die Varianz der \bar{A}_i -Werte ist analog zum Standardfehler des Mittelwertes
 - also gilt : \bar{A}_i -Varianz $\hat{\sigma}_{\bar{A}}^2 = \hat{\sigma}_{\bar{x}}^2$.
 - $\hat{\sigma}_{\bar{x}}^2$ ist definiert als $\hat{\sigma}^2 / n$, also gilt $\hat{\sigma}_{\bar{A}}^2 = \hat{\sigma}^2 / n$
 - da $\hat{\sigma}_{\text{treat}}^2 = \hat{\sigma}_{\bar{A}}^2 * n$ ergibt sich $\hat{\sigma}_{\text{treat}}^2 = \hat{\sigma}^2$
 - wenn die einzelnen Fehlervarianzen der Stichproben homogen sind, stellt $\hat{\sigma}_{\text{Fehler}}^2$ eine erwartungstreue Schätzung von $\hat{\sigma}^2$ dar.
 - Bei H_0 muß also gelten : $\hat{\sigma}_{\text{treat}}^2 = \hat{\sigma}_{\text{Fehler}}^2$
- somit kann man über den F-Test prüfen : $F = \hat{\sigma}_{\text{treat}}^2 / \hat{\sigma}_{\text{Fehler}}^2$

- wenn dieser empirische F-Wert größer als kritischer bei Zählerdf = df_{treat} und Nennerdf = df_{Fehler} , dann mit xx-Niveau signifikant unterschiedliche Varianzen, d.h. mindestens zwei Treatments (ausgedrückt durch ihre μ 's unterscheiden sich signifikant)
- wenn $\text{Varianz}_{\text{Fehler}}$ größer $\text{Varianz}_{\text{Treat}}$ erübrigt sich Test, da Treatmenteffekte unbedeutend, F-Test überprüft also einseitig (gerichtet), Treatvarianz soll ja größer als Fehlervarianz sein, entspricht jedoch der ungerichteten H_0 !

Einzelvergleiche (Scheffé-Test) :

- Welche μ 's sind signifikant unterschiedlich ?
- a-priori-Methoden : zu kompliziert
- a-posteriori-Methoden : z.B. Scheffé-Test :
- Absolutbeträge aller Mittelwertsdifferenzen $\text{diff}_{\text{emp}(i)}$ (z.B. $\bar{A}_1 - \bar{A}_2$) bilden, dann kritische Differenz über Formel berechnen, wenn $\text{diff}_{\text{emp}(i)} > \text{diff}_{\text{crit}}$, dann signifikant für das μ -Paar, auf das dies zutrifft. Absolutbetrag meint : immer positiven Wert.

Trendtests :

- spezielle Form der Einzelvergleiche
- zerlegen die Treatmentsumme in Anteile, die auf versch. Trends in den Mittelwerten der AV zurückzuführen sind.
- Bei p Faktorstufen lassen sich maximal p-1 Trendkomponenten untersuchen :

1. Ordnung	linear
2. Ordnung	quadratisch
3. Ordnung	kubisch
4. Ordnung etc	quartisch etc

- alle QS_{trend} addieren sich zur QS_{treat} , alle df_{trend} 's addieren sich df_{treat} (p-1)
- Wenn nur ein linearer und ein non-linearer Trend untersucht wird (der alle non-lin. umfasst), ergibt sich für den linearen Trend ein df von 1 und für den df_{nonlin} die restlichen p-2 df's.
- für jeden Trend unterschiedliche C-Koeffizienten, summieren sich zu Null
- für jeden Trend wird QS & dann Varianz berechnet (oft gleich der QS, da $df = 1$) & durch F-Bruch an Fehlervarianz auf Signifikanz getestet
- zuerst QS_{lin} , dann ergibt sich durch $QS_{\text{treat}} - QS_{\text{lin}}$ die QS_{nonlin} , wenn diese signifikant, lohnt es sich, die höheren Trends auch zu untersuchen
- die Trends stehen orthogonal aufeinander (statistisch unabhängig)
- jetzt läßt sich auch die Korrelation unabh. Var. - abh. Var. einfacher berechnen mit

$$r_{\text{lin}} = \sqrt{\frac{QS_{\text{lin}}}{QS_{\text{tot}}}} \quad r_{\text{lin}} \text{ berücksichtigt allerdings nur die auf den linearen Trend}$$

zurückgehende Korrelation, r_{quad} berücksichtigt lin + quad, und r_{total} berechnet sich

$$\text{mit } \sqrt{\frac{QS_{\text{lin}} + QS_{\text{quad}} + \dots + QS_{\text{trend}(p-1)}}{QS_{\text{tot}}}} = \sqrt{\frac{QS_{\text{treat}}}{QS_{\text{tot}}}} = \eta^2$$

- wenn QS_{lin} nicht sign., dann ist r_{lin} auch recht niedrig
- UV und AV müssen intervallskaliert sei, also die UV wenigstens subjektiv äquidistant skaliert.
- Es können mehrere Trendkomponenten signifikant werden, die Stichproben der Faktorstufen müssen *unabhängig* sein, also nicht bei VA mit Meßwiederholung verwendbar.

α -Fehler-Kumulierung :

- Es können nicht einfach mehrere Tests, z.B. t-tests gerechnet werden anstatt einer VA, wegen der α -Fehler-Kumulierung : es ergibt sich größeres α , da unter Annahme der stochastischen Unabhängigkeit der t-tests sich das neue α mit $1-(1-\alpha)^{\text{Anzahl der t-tests (p-1)}}$ ergibt (siehe Wahrscheinlichkeitsrechnung), also z.B. mit $1-0,95^2 \approx 0,1$, also 10% bei zwei t-tests
- man muß neues α' berechnen, mit diesem α' muß man rechnen, um mehrere t-tests durchzuführen & doch das gewünschte α zu erhalten
- vereinfachte Formel : $\alpha' = \alpha / (p-1)$ (Bonferoni-Korrektur)
- wenn Treatments nicht voneinander unabhängig, dann genügt (bei perfekter Abhängigkeit) ein Test, da alle anderen Tests zum gleichen Ergebnis führen würden, α -Korrektur also nicht notwendig, wäre der Fall, wenn u.g. Erziehungsvariablen eine Korrelation von 1 hätten
- bei wachsender Abhängigkeit also konservativere α -Korrektur
- α -Korrektur auch nötig, wenn z.B. globale Hypo : Erziehungsstil beeinflusst Sozialverhalten der Kinder & Erziehungsstil wie auch Soz.verh. wären durch je 5 Variablen operationalisiert. Es ergeben sich hier eigtl. 25 Einzelhypos (wenn Faktoren jeweils unabhängig) - so etwas wird nebenbei gesagt mit der sog. kanonischen Korrelationsrechnung zusammenklamüsert.

Voraussetzungen :

- Intervallskalenniveau
- QS-Zerlegung ohne weitere Voraussetzungen, nur F-Signifikanz-Test hat Voraussetzungen :
 1. Fehlerkomponenten in Pops normalverteilt
 2. Fehler-Varianzen homogen in Pops (Kontrolle durch Bartlett-Test, aber sehr sensibel, wenn Stichproben gleich groß lieber über F_{\max} -Test)
 3. Fehlerkomponenten voneinander unabhängig, d.h. Treatment- & Fehlereffekte additiv
- gilt fast alles auch für Scheffé-Test & Trendanalysen
- bei Stichproben größer 10 wenig Bedenken

zu 2. : F_{\max} -Test :

- ist gegeben durch Quotient der größten der i Fehlervarianzen durch die kleinste
- eigene Tabelle, abhängig von Anzahl der Treatments & der Freiheitsgrade der einzelnen Varianz (n-1)

zu 3. :

- Fehlerkomponenten normalerweise dann unabhängig, wenn Untersuchungseinheiten den Treatmentstufen zufällig zugeordnet & unter den jeweiligen Treatmentstufen unterschiedliche Stichproben untersucht
- letzteres ist nicht der Fall bei VA mit Meßwiederholung

Mehrfaktorielle VA (zwei, drei & mehr) (Kap 8) :

	i	i	i	
j	xx	xx	xx	$\overline{B_j}$
	xx	xx	xx	
j	xx	xx	xx	$\overline{B_j}$
	xx	xx	xx	
	$\overline{A_i}$	$\overline{A_i}$	$\overline{A_i}$	\overline{G}

Hier $\overline{A_1 B_1}$

Index m gibt die VP in der Zelle an, läuft von 1 bis m, d.h. n bezeichnet VPN pro Zelle

$X_{ijm} : i = 1 \dots p \text{ (A)}$
 $j = 1 \dots q \text{ (B)}$
 $m = 1 \dots n \text{ (VP pro Zelle)}$

- Hypothesen :
 - $H_{0(A)} : \mu_1 = \mu_2 = \mu_3 = \dots \mu_p$
 - $H_{1(A)} : \mu_i \neq \mu_i'$
 - $H_{0(B)} : \mu_1 = \mu_2 = \mu_3 = \dots \mu_q$
 - $H_{1(B)} : \mu_j \neq \mu_j'$
 - $H_{0(\text{Interaktion})} : \mu_{ij} = \mu_i + \mu_j - \mu$ (zwischen Faktoren keine Interaktion)
- wenn einfakt. VA kein signifikantes Ergebnis, liegt dies an zu kleiner Treatmentvarianz und/oder zu großer Fehlervarianz
- die „wahre“ Treatmentbedeutsamkeit ist nicht zu beeinflussen, wohl aber die Fehlervarianz, wodurch natürl. die relative Bedeutsamkeit d. Treatments steigt
- Reduktion der Fehlervarianz durch ...
 - *Variablen konstant halten* : wenn möglichst viele Variablen, die potentiellen Einfluß auf abhängige Variable haben, konstant gehalten werden, können sie Fehlervarianz nicht beeinflussen (z.B. nur Männer untersuchen, Geschlecht hat dann keinen Einfluß). Nachteil : Ergebnisse nicht verallgemeinerbar, nur auf Männer anwendbar, also keine externe Validität
 - *Variablen kontrollieren* : andere Störvariablen (die die abhängige potentiell beeinflussen) mit erheben & im nachhinein rausrechnen (Kovarianzanaly.)
 - *Störvariablen systematisch variieren*, also z.B. VPN auch nach Geschlecht gruppieren : mehrfaktorielle VA
- Die zweifakt. VA untersucht also Beziehung zw. A und B, bzw. Wirkung von A auf B und umgek. Ist A od. B solo interpretierbar, od. nur Zusammenw. der beiden?

Quadratsummenzerlegung :*Totale QS :*

- wie bei einfaktorieller VA

Zellen-QS :

- wie müßten Werte beschaffen sein, wenn nur von den beiden Faktoren abhängig?
Alle müßten gleich sein in Zelle
- alle Zellenwerte durch ihr jeweiliges Mittel $\overline{A_i B_j}$ ersetzen
- alle quadrierten Abweichungen von \overline{G} berechnen (wenn 5 Werte in Faktorstufe, dann auch fünf mal, also n-mal) und addieren (über alle Faktorstufen)
- setzt sich zusammen aus der QS der Haupteffekte & der Interaktion

Fehler-QS :

- ergibt sich aus den aufsummierten quadrierten Abweichungen aller Zelleneinzelwerte von ihrem jeweiligen Zellenmittelwert
- Fehler-QS ist gg.über der einfaktoriellen meist kleiner (jedoch nie größer), um den gleichen Betrag, wie Zellen-QS gg.über Treatment-QS größer geworden ist,
 $QS_{\text{Zellen}} + QS_{\text{Fehler}} = QS_{\text{tot}}$

Haupteffekte-QS :

- QS_A entspricht QS_{treat}
- QS_B analog zu QS_A : Ersetzen der Meßwerte unter den B-Stufen durch den jeweiligen Mittelwert der B-Stufe, quadrierte Abweichungen $\overline{B_j} - \overline{G}$ bilden (jeweils n-mal), alle addieren
- QS_{Zellen} ist nicht gleich $QS_A + QS_B$, sondern größer, enthält noch die Inter.- $QS_{A \times B}$

Interaktions-QS :

- wann enthält QS_{Zellen} nur Unterschiede der Faktoren A und B ? Wenn Faktorenauspräg. A alle Stufen d. Faktors B in *konstanter* Weise beeinflussen & umgekehrt
- es müßten dann andere Zellenmittelwerte $\overline{A_i B_j}'$ existieren mit $\overline{A_i B_j}' = \overline{A_i} + \overline{B_j} - \overline{G}$
- die quadrierten Abweichungen der realen empir. Zellenmittelwerte von den theoretischen errechneten (bei keiner Interaktion) werden mal der Anzahl n der Werte in der Faktorstufe genommen & aufsummiert, das ergibt die $QS_{A \times B}$
- Interaktions-QS ist nur durch einen eigenständigen Effekt zu erklären, der über A und B alleine hinausgeht
- alle Effekte sind wechselseitig unabhängig

Varianzschätzungen :

- QS_{Fehler} einer einfakt. VA niemals kleiner als die QS_{Fehler} der entspr. zweifakt. VA
- das gilt aber nicht für die entsprechenden Varianzen, wg. der veränderten df's (verändert, da neuer Faktor eingeführt)

Signifikanztests :

- Prüfung der drei Nullhypos über drei F-Brüche

Varianzaufklärung :

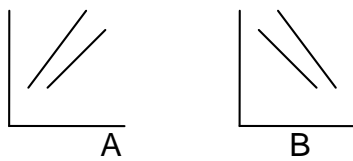
- für jeden Faktor (A und B) sowie die Interaktion einzeln
- über z.B. η^2 mit anstatt der Treatmentvarianz im Zähler die A-, B- oder AxB-Varianz

Interaktionsdiagramme :

- Jetzt werden die beiden Interaktionsdiagramme für die A*B-Interaktion gezeichnet (nur wenn signifikantes F_{A*B}) :
- Wenn Geraden gleiche Steigungsrichtung, also gleichen Trend haben, darf Haupteffekt als solcher interpretiert werden. In Bild A wird immer der HE A interpretiert, in Bild B der HE B.

3 Formen der Interaktion :

Interpret. werden muß:



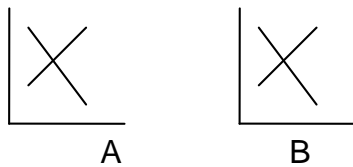
ordinale Interaktion

A,B



hybride Interaktion

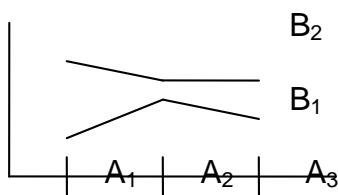
B (da Trends/Steigungsrichtung in Bild B gleich)



disordinale Interaktion

--

Abgetragen bei diesen Diagrammen werden die $\overline{A_i B_j}$ -Werte :



$\overline{A_3 B_2}$

dieses Diagramm ist analog zu einem der obigen A-Bilder

$\overline{A_1 B_2}$

- Drei- oder mehrfaktorielle VA einfach analog zur zweifaktoriellen

VA mit Meßwiederholungen (Kap 9) :

	5 h davor	5 sek davor	5 h danach	\bar{P}_m
VP ₁				\bar{P}_m
VP ₂				
	\bar{A}_i	\bar{A}_i	\bar{A}_i	\bar{G}

VPN : m von 1 bis n Faktorstufen i von 1 bis p

- für abhängige Stichproben, ist genau wie die einfakt. VA eine Erweiterung des t-tests, hier des abhängigen
- bei Veränderungen über d. Zeit wie auch bei parallelisierten Stichpr. (z.B. Paare)
- Hypothesen wie in einfaktorieller VA

Quadratsummenzerlegung :

$$QS_{\text{tot}} = QS_{\text{Zwischen den VPN}} + QS_{\text{Innerhalb der VPN}}$$

$$QS_{\text{tot}} = QS_{\text{ZW VPN}} + QS_{\text{TREAT}} + QS_{\text{RES}}$$

$$df_{\text{tot}} = df_{\text{Zwischen den VPN}} + df_{\text{Innerhalb der VPN}}$$

$$df_{\text{tot}} = df_{\text{ZW VPN}} + df_{\text{TREAT}} + df_{\text{RES}}$$

- QS_{tot} wie bei einfaktorieller VA
- QS_{treat} wie bei einfaktorieller VA
- QS_{RES} erfasst das Rauschen in den Zellen wie QS_{Fehler} und außerdem noch Interaktionseffekte $VPN \times \text{Treatment}$ (äußert sich z.B. in untersch. Reaktion der einzelnen VPN auf Dauerbelastungssituation)
- $QS_{\text{Zwischen den VPN}}$ ohne Bedeutung, lediglich a-priori-Unterschiede, ergibt sich als Summe der quadr. Abw. des <Mittelwertes aller Meßwerte einer VPN> vom <Gesamtmittelwert>
- $QS_{\text{Innerhalb der VPN}}$ ergibt sich als Summe der quadr. Abw. der <Meßwerte> vom <Mittelwert aller Meßwerte einer VPN>, durchlaufend über alle VPN
- F-Signifikanz-Test mit $\hat{\sigma}_{\text{treat}}^2 / \hat{\sigma}_{\text{res}}^2$

Trendtest & Einzelvergleiche :

- Trendtest & Scheffé-Test auch möglich, als Prüfvarianz $\hat{\sigma}_{\text{res}}^2$ anstatt $\hat{\sigma}_{\text{Fehler}}^2$

Voraussetzungen :

- Korrelationen zw. Faktorstufen homogen
- Varianzen in Faktorstufen homogen

Homogene-Varianzen-Test für abh. Stichproben nicht vorhanden, aber ε -Korrektur :

- $\frac{1}{p-1} \leq \varepsilon \leq 1$, dieser Wert ε wird mit den Freiheitsgraden multipliziert, diese werden kleiner, die Entscheidung konservativer, d.h. später signifikant.

Kovarianzanalyse (Kap 10) :

	i		i		i	
m	yy	xx	yy	xx	yy	xx
m	yy	xx	yy	xx	yy	xx
	\bar{A}_i	\bar{B}_i	\bar{A}_i	\bar{B}_i	\bar{A}_i	\bar{B}_i

\bar{G}_y, \bar{G}_x

Variable (y oder av)

Kovariable (x oder kov)

- durch mehrfaktorielle VA kann man Fehlervarianz reduzieren, führt jedoch schnell zu hohen VPN-Zahlen
- bei Meßwiederholungs-VA weniger VPN, aber evtl. zu hohe Beanspruchung der VPN
- Kovarianzanalyse reduziert Fehlervarianz ohne höhere VPN-Anzahl & -Belastung, indem Kontrollvariablen miteingehoben werden, die intervallskaliert sein müssen
- Bedeutsamkeit dieser Kovariablen wird in KoVA untersucht, wenn z.B. a-priori-Unterschiede zwischen VPN nicht beeinflussen sollen (wie bei VA mit Meßwiederholung), oder wenn Faktor wie aktuelles Tageslicht aus Messung der idealen Arbeitsplatzbeleuchtung herauspartialisiert werden soll
- Kombination VA-Techniken mit Regressionstechniken :
- Bestimmung einer Regressionsgleichung zw. abh. Variable & Kontrollvariable (wie wenn abh. Variable nur von Kontrollvariable beeinflusst wäre)
- dann Gesamt-Regressionsresiduen y^* berechnen (über \bar{G}) : Abweichung der theoretischen Werte von den tatsächlichen empirischen
- jetzt VA über die Regressionsresiduen

Quadratsummenzerlegung :*Die (bereinigte) totale QS* :*

- ist die Summe der quadrierten Regressionsresiduen y^* , da diese schon die Abweichung beinhalten
- es ergibt sich eine meist etwas kleinere QS^*_{tot} (nie größer), als bei normaler VA, da Kontrollvariable ja rausgerechnet

Die (bereinigte) Fehler-QS :*

- für jede Gruppe wird einzelne Fehler-QS* berechnet - über drei Regressionsgleichungen mit Vorhersage KoVariable auf abh. Var. (also auch mit drei verschiedenen b-Gewichten)
- aus den drei Steigungskoeffizienten b mitteln wir das zusammengefaßte b, indem alle Zähler & alle Nenner der b 's jeweils addiert werden
- dieses wird wieder in die drei Gleichungen eingesetzt & damit Vorhersagewerte errechnet, so daß die Steigung zwar gleich ist, aber die Höhenlage untersch.
- Begründung : Voraussetzung ist, daß Fehlervarianzen homogen. Werden dann versch. Konstanten addiert, ändert dies nichts an Homogenität, die unterschiedliche Höhenlage spiegelt lediglich die untersch. Mittelwerte wieder
- nun Abzug der theoretisch errechneten Vorhersagewerte von den tatsächlichen, empirischen Werten, ergibt die Gruppen-Regressionsresiduen y^* (über \bar{A}_i , nicht Gesamt-Reg-residuen über \bar{G} wie oben !!!)
- die einzelnen errechneten y^* -Werte werden nun quadriert und addiert, fertig ist die (um die Kontrollvariable bereinigte) Fehlervarianz, die meist kleiner als die der normalen VA ist, jedenfalls nie größer

Die (bereinigte) Treatment-QS* :

- nur durch $QS^*_{\text{tot}} - QS^*_{\text{Fehler}}$ berechenbar, meist größer als unkontrollierte QS_{treat}

Freiheitsgrade :

- df^*_{tot} um einen verringert, da auch noch mit ursprünglichen y-Werten verbunden
- df^*_{Fehler} auch um einen verringert
- df^*_{treat} bleibt gleich

F-Test :

- ganz normal, halt mit den korrigierten Varianzen (korrigierte QS^* und df^*)

Voraussetzungen :

- neben denen für die normale VA :
- homogene Steigungen der Regressionen innerhalb der Stichproben
- allerdings sehr robust

Wichtig :

- Die Kovarianzanalyse kann sowohl ein vorher nicht signifikantes Ergebnis signifikant machen, wie auch umgekehrt.
- Letzteres, wenn Ko-Var. mit unabh. Var. korreliert, wenn z.B. Wirkung von Schichtunterschieden (unabh.) auf Ausgaben für Kindererziehung (abh.) gemessen wird & Einkommen der Eltern (Kont.-Var.) rausgenommen wird

Zweifaktorielle KoVa (nicht so wichtig) :

- bei zweifaktorieller KoVa müssen QS_A , QS_B , $QS_{A \times B}$ und QS_{Fehler} korrigiert werden, so daß die korrigierten QS^* für die Haupteffekte & die Interaktion nicht mehr einzeln subtraktiv aus der korrigierten QS^*_{tot} & QS^*_{Fehler} bestimmt werden können - wie das bei einfakt. KoVa mit QS^*_{treat} möglich war
- trotzdem bleibt das Grundprinzip erhalten : Zur Berechnung der korrigierten Haupteffekte bzw. Interaktionen subtrahieren wir die korrigierte Fehler- QS^* von einer QS^* , die nur Fehleranteile & Anteile des jeweils interessierenden Haupteffektes (Interaktion) enthält
- die Fehler- QS^* erhalten wir analog zur einfakt. KoVa : es wird hierbei für jede Zelle ein b-Gewicht bestimmt
- die korrigierten QS der Haupteffekte (Interaktion) erhalten wir auf indirektem Wege :
- Zunächst Zusammenfassung unkorrigierte $QS_{\text{Haupteffekt (Interaktion)}}$ mit unkorrigierter QS_{Fehler}
- dann Korrektur dieser zusammengefaßten QS
- dann Subtraktion der korrigierten QS^*_{Fehler} von der zusammengefaßten, korr. QS^*
- es kommt raus die korrigierte $QS^*_{\text{Haupteffekt (Interaktion)}}$
- df 's der Haupteffekte & Interaktion gegenüber der zweifakt. VA nicht verändert

Begriffe, Methodenlehre (Ausschnitte) :

Der größte Teil der Methodenlehre ist hier ausgelassen - nachzulesen in z.B. dem Methodenskript von Claudia Evers.

- interne Validität : inwieweit sind Ergebnisse logisch eindeutig interpretierbar, wird das gemessen, was gemessen werden soll, int. Val. sinkt mit wachsender Anzahl von mögl. Störfaktoren & Alternativerklärungen für das Ergebnis (meist gut bei Experimenten)
- externe Validität : sind Ergebnisse generalisierbar (meist gut bei Feldstudien), wird von interner Val. natürlich auch beeinflusst ...
- Reliabilität : mißt der Test die Variablen in richtiger Quantität, nachgeprüft z.B. durch re-test-reliabilität (Testwiederholung) - hier spielt Fehleranfälligkeit der abh. Var. eine große Rolle
- Faktorenanalyse :
 - Ziel : Herausheben von Faktoren in Datenstruktur
 - verschiedene Verfahren unter diesem Sammelbegriff, z.B. Hauptkomponentenanalyse
 - wenn zwei Variablen gemessen werden lassen sich über Korrelation zw. diesen Gemeinsamkeiten erkennen
 - was sie aber genau messen, läßt sich besser erkennen, wenn Korrelation mit weiteren Variablen vorhanden, die evtl. eindeutiger zu interpretieren sind
 - bei 20 Variablen müssen aber schon 190 Korrelationen analysiert werden, rein intuitiv geht das nicht mehr, sondern nur mit math. Verfahren Faktorenanalyse lassen sich anhand der Korr. voneinander (fast) unabh. Gruppen / Faktoren herausbilden, die dann allerdings hypothetischen Charakter haben
 - FA wirkt somit datenreduzierend
 - FA ist heuristisch, hypothetisch : man könnte auch andere Faktoren bilden, Faktoren müssen nachher noch interpretiert werden & das ist nie eindeutig möglich
 - bei der Hauptkomponentenanalyse (PCA = Principal Components Analysis) finden sich folgende Kennwerte :
 - Faktorwerte : Werte der VPN auf den errechneten z-standardisierten Faktoren
 - Faktorladung : ist die Korrelation eines Faktors mit einer Variablen (Variable lädt mit xx% auf Faktor)
 - Kommunalität einer Variablen : gibt an, in welchem Ausmaß die Variable durch Faktoren erfaßt / aufgeklärt wird
 - Eigenwert eines Faktors : gibt an, wieviel von Gesamtvarianz aller Variablen durch Faktor erfaßt wird, umso größer, je stärker Variablen korrelieren