

Varianzanalyse

Bisher gab es Verfahren, mit denen man über zwei Stichproben ermitteln konnte, ob ein Merkmal in zwei verschiedenen Populationen unterschiedlich ausgeprägt ist. Bei z.B. Intervallskalenniveau wäre dies ein t-Test für unabhängige Stichproben.

Da aber die meisten psychologischen Untersuchungen die wechselseitigen Beeinflussungen vieler verschiedener Variablen erfassen wollen, ist der t-Test dort ziemlich nutzlos.

Und dabei springt jetzt die Varianzanalyse ein.

Die Varianzanalyse ermöglicht die simultane Kontrolle einer oder mehrerer unabhängiger Variablen, welche auf ein Merkmal (abhängige Variable) einwirken.

Verfahren, mit denen gleichzeitig mehrere abhängige Variablen geprüft werden können, findet man bei „Multivariate Verfahren“.

Der Begriff „Varianzanalyse“ ist historisch gewachsen und beinhaltet alle Verfahren in diesem Kapitel. Würde man nur der Definition des Begriffes „Varianzanalyse“ folgen – Analyse der Varianz bzw. Streuung von Daten –, würde u.a. auch der t-Test darunterfallen. Deshalb von der folgenden Definition nicht irritieren lassen, auch wenn fast alle Verfahren aus dem Teil B2 darunter fallen.

Die Varianzanalyse führt Mittelwertvergleiche unter zu Hilfenahme von Varianzen bzw. Streuungen von Werten durch.

Einfaktorielle Versuchspläne:

Die einfaktorielle Varianzanalyse ist die einfachste Version aller Varianzanalysen.

Sie überprüft die Auswirkungen einer p-fach gestuften unabhängigen Variable auf eine abhängige Variable.

Sie überprüft also, ob p Stichproben aus Grundgesamtheiten mit denselben Mittelwerten stammen – also ob es (Mittelwerts-) Unterschiede gibt oder nicht.

Als Beispiel: Die einfaktorielle Varianzanalyse untersucht die Auswirkungen einer unabhängigen Variable (Medikamente), auf eine abhängige Variable (Stimmung von Depressiven), wobei die unabhängige Variable p-fach gestuft sein kann (Medikament A, Medikament B, etc.).

Die Frage ist nun: Stammen alle Stichproben (A, B, etc.) aus Grundgesamtheiten (alle Menschen, die die Medikamente A, B etc. nehmen) mit denselben Mittelwerten (die Medikamente wirken alle gleich gut /schlecht), oder gibt es Unterschiede (Medikament z.B. A ist besser als alle anderen).

Vom englischen *Analysis of Variances* leitet sich das auch im deutschen übliche Akronym *ANOVA* her.

Zum allgemeinen Verständnis erst einmal ein paar Begriffe:

Begriffe der Varianzanalyse:

Abhängige Variable (AV): das Merkmal, dessen Varianz durch die Varianzanalyse untersucht wird. Eine AV muss intervallskaliert sein.

Unabhängige Variable (UV): auch Faktor genannt; ein oder mehrere Merkmale, die am Zustandekommen von Unterschieden bei der AV beteiligt sein können. Das Skalenniveau der UV ist beliebig, jede Versuchsperson (Vp) muss nur eindeutig einer (Faktor-) Stufe der UVs zuordbar sein.

Stufen: Anzahl der Kategorien der UV, z.B. wäre eine einfaktorielle Varianzanalyse mit der Untersuchung von Parteipräferenzen von Studenten bei den im Bundestag vertretenen Parteien (PDS, SPD, Grüne, FDP, Union) fünf-fach gestuft (Sie hätte fünf Faktorstufen).

Quadratsumme (QS): Die Quadratsumme ist die Summe der quadrierten Abweichungen aller Messwerte vom Mittelwert.

Grundprinzip der einfaktoriellen Varianzanalyse:

Datenschema: Der ANOVA liegt das folgende Datenschema zugrunde:

Faktor A					
a_1	a_2	...	a_i	...	a_p
x_{11}	x_{12}	...	x_{1i}	...	x_{1p}
x_{21}	x_{22}	...	x_{2i}	...	x_{2p}
\vdots	\vdots		\vdots		\vdots
x_{m1}	x_{m2}	...	x_{mi}	...	x_{mp}
\vdots	\vdots		\vdots		\vdots
$x_{n_1 1}$	$x_{n_2 2}$		$x_{n_i i}$		$x_{n_p p}$
\bar{A}_1	\bar{A}_2		\bar{A}_i		\bar{A}_p
Stufenmittelwerte:					
$\bar{A}_i = \frac{1}{n_i} \cdot \sum_{m=1}^{n_i} x_{mi}$					
Gesamtmittelwert:					
$\bar{G} = \frac{1}{N} \sum_{i=1}^p \sum_{m=1}^{n_i} x_{mi}, \text{ wobei } N = \sum_{i=1}^p n_i$					

Hypothesen:

$$H_0: \mu_1 = \dots = \mu_i = \dots = \mu_p$$

$$H_1: \mu_i \neq \mu_{i'}$$

Die H_1 besagt also nicht, dass sich alle Mittelwerte unterscheiden, sondern dass sich mindestens zwei Mittelwerte unterscheiden, also ob mindestens ein Effekt vorliegt

Wenn die H_0 zutrifft und die Populationsmittel gleich sind, ist auch die Summe der quadrierten Effekte gleich Null. Die H_0 ließe sich also auch wie folgt formulieren:

$$H_0: \sum_{i=1}^p \alpha_i^2 = \sum_{i=1}^p (\mu_i - \mu)^2 = 0.$$

Wenn die H_1 zutrifft, die Populationsmittel also nicht alle gleich sind, ist auch die Summe der quadrierten Effekte größer Null:

$$H_1: \sum_{i=1}^p \alpha_i^2 = \sum_{i=1}^p (\mu_i - \mu)^2 > 0.$$

Der Einfluss einer UV, der sich in unterschiedlichen Mittelwerten zeigt, kann über eine Summe von Abweichungsquadraten (Quadratsumme) der Effekte ausgedrückt werden.

In dem Modell der einfaktoriellen ANOVA werden zwei Quellen der Variation unterschieden: das **Treatment und der Fehler**.

Mehrere t-Tests statt einer Varianzanalyse?

Bei nur zwei Stichproben sind der t-Test für unabhängige Stichproben und die einfaktorielle Varianzanalyse identisch. Könnte man jetzt (rein theoretisch) nicht einfach jede Menge t-Test rechnen?

Es gibt zwei Probleme:

- 1.) Bei z.B. einem Alphaniveau von 5% und 100 gerechneten t-Tests werden aufgrund der Wahrscheinlichkeit ca. 5(%) der 100 Tests signifikant für die H_1 sein, auch wenn eigentlich die H_0 fest gilt.
- 2.) Es gibt das Problem der Alpha-Kumulierung (kommt später noch mal):
Drei t-Tests mit einem Alphaniveau von je 0,05 haben zusammen ein Alphaniveau von $(1 - \alpha)^3 = (0,95)^3 = 0,857375!$

Quadratsummenzerlegung:

Die einfaktorielle Varianzanalyse geht von folgendem aus:

Man registriert eine Gesamtvarianz aller Messwerte. (=QS_{tot})

Es wird nun gefragt, ob der Anteil der Treatmentquadratsumme – die Unterschiede, die auf den Faktor zurückzuführen sind – **hinreichend größer ist als die Fehlerquadratsumme** – die individuellen Unterschiede in der Gruppe.

Totale Quadratsumme:

Die gesamte Variation der Messwerte geht in die totale Quadratsumme ein:

$$QS_{tot} = \sum_{i=1}^p \sum_{m=1}^{n_i} (x_{mi} - \bar{G})^2.$$

QS_{tot} ist also die Summe der Abweichungen aller Messwerte vom Gesamtmittelwert.

Treatmentquadratsumme:

Die Mittelwertsunterschiede (geschätzte Effekte der UV) gehen in der ANOVA in die Treatmentquadratsumme ein:

$$QS_A = \sum_{i=1}^p n_i \cdot (\bar{A}_i - \bar{G})^2.$$

QS_A (später auch QS_B , QS_C , etc.) ist also die Summe der mit n_i multiplizierten quadrierten jeweiligen (= einzelnen) (Faktor-) Stufenmittelwerte abzüglich des Gesamtmittelwertes.

QS_A ist also der Unterschied bei den Messwerten, der nur auf die Faktorstufen (z.B. die Art des Medikaments zurückzuführen ist.

Wäre Unterschiedlichkeit der Werte ganz vom Treatment abhängig, dürften sich die Messwerte der Personen in einem Treatment nicht unterscheiden

Fehlerquadratsumme:

Die individuellen Unterschiede je Gruppe (Fehlerkomponenten) werden mit folgender Fehlerquadratsumme untersucht:

$$QS_{Fehler} = \sum_{i=1}^p \sum_{m=1}^{n_i} (x_{mi} - \bar{A}_i)^2.$$

QS_{Fehler} ist als die Summe der quadrierten Abweichungen aller einzelnen Messwerte vom jeweiligen Stufenmittelwert.

QS_{Fehler} ist so etwas ähnliches wie die Regressionsresiduen, es ist der Teil, der nicht von QS_A erfasst wird, also z.B. die individuellen Unterschiede oder Experimentalängel.

Quadratsummenzerlegung:

Für jede einfaktorielle ANOVA (ohne Messwiederholung) gelten folgende Gleichungen.

$$QS_{tot} = QS_A + QS_{Fehler}$$

$$df_{tot} = df_A + df_{Fehler}$$

Die Varianzen verhalten sich nicht additiv.

Varianzschätzungen:

Wenn man die Quadratsummen errechnet hat, kann man die einzelnen Varianzen (also Treatmentvarianz, Fehlervarianz und Gesamtvarianz) schätzen.

Die Division der Quadratsummen durch ihre Freiheitsgrade führt zu folgenden Varianzschätzungen:

$$\text{Treatmentvarianz: } \hat{\sigma}_A^2 = \frac{QS_A}{df_A}, \text{ mit } df_A = p-1.$$

$$\text{Fehlervarianz: } \hat{\sigma}_{Fehler}^2 = \frac{QS_{Fehler}}{df_{Fehler}}, \text{ mit } df_{Fehler} = N-p.$$

Bei Gültigkeit der H_0 stammen die einzelnen Werte unter den Faktorstufen (A_i) aus der gleichen Population, d.h. die Varianz der Stufenmittelwerte ist gleich dem Standardfehler des Mittelwertes, da auch bei der ANOVA eine n große Stichprobe eine Varianz von $\hat{\sigma}_{\bar{x}}$ hat.

Bei Gültigkeit der H_0 sind sowohl $\hat{\sigma}_A^2$ als auch $\hat{\sigma}_{Fehler}^2$ erwartungstreue Schätzer der gemeinsamen Populationsvarianz σ_ε^2 .

Wenn die H_0 gilt, dann:

$$E(\hat{\sigma}_A^2) = E(\hat{\sigma}_{Fehler}^2) = \sigma_\varepsilon^2$$

Je größer die Unterschiede zwischen den Mittelwerten ist, desto größer ist die Treatmentvarianz; die Fehlervarianz bleibt unverändert. Daher ist bei Gültigkeit der H_1 zu erwarten, dass $\hat{\sigma}_A^2$ größer ausfällt als $\hat{\sigma}_{Fehler}^2$.

Wenn die H_1 gilt, dann:

$$E(\hat{\sigma}_A^2) > E(\hat{\sigma}_{Fehler}^2) [= \sigma_\varepsilon^2]$$

Die ANOVA testet, ob die Treatmentvarianz größer ist als die Fehlervarianz; wenn dies der Fall ist, wird geschlossen, dass nicht alle Mittelwerte gleich sind. Die Frage ist nun, ob der Unterschied signifikant ist.

Ist QS_{Fehler} größer als QS_A , erübrigt sich ein Test (unbedeutende Treatmenteffekte).

Entscheidung:

Die geschätzten Varianzen vom Treatment und vom Fehler verteilen sich unter der H_0 gemäß einer F-Verteilung mit $df_A = p-1$ Zähler- und $df_{\text{Fehler}} = N-p$ Nennerfreiheitsgraden:

$$F_{\text{emp}} = \frac{\hat{\sigma}_A^2}{\hat{\sigma}_{\text{Fehler}}^2}$$

Die H_0 wird verworfen, falls $F_{\text{emp}} > F_{(df_A, df_{\text{Fehler}}, 1-\alpha)}$

Aufgrund der Quadrierung in QS_A geht die Information über die Richtung einer Abweichung (positiv/negativ) verloren. Mit der ANOVA können daher keine gerichteten Hypothesen getestet werden.

D.h. wir testen ungerichtete Hypothesen mit einem F-Test rechtsseitig (bei einem Alpha-niveau von 1% schneiden wir an der F-Kurve rechts 1% ab).

Effektgrößen:

Um Effekte verschiedener Untersuchungen vergleichbar zu machen, werden auch bei der ANOVA Effekte zu Effektgrößen standardisiert:

$$\varepsilon = \frac{d}{2} \sqrt{\frac{p+1}{3 \cdot (p-1)}}$$

Wobei „d“ die Differenz zwischen dem höchsten und niedrigsten Wert – dividiert durch die Streuung – darstellt:

$$d = \frac{\mu_{\max} - \mu_{\min}}{\sigma}$$

Wenn man vor einer Untersuchung eine Effektgröße festlegt, hat dies den Vorteil (neben dem größeren Aufwand), dass der Stichprobenumfang kalkulierbar ist. (Die Formel dazu steht nicht im Bortz.) **Somit pendelt man gut zwischen zu großen und zu kleinen Stichproben.**

Die Effektgröße kann auch über die Varianzaufklärung (siehe nächster Punkt) ermittelt werden:

$$\varepsilon = \sqrt{\frac{\eta^2}{1 - \eta^2}}$$

Varianzaufklärung:

Wie viel der Gesamtvarianz über das Treatment aufgeklärt werden kann, wird über folgenden Koeffizienten ermittelt:

$$\eta^2 = \frac{QS_A}{QS_{tot}}$$

Während η^2 ein Stichprobenkennwert ist, ergibt sich eine **Schätzung** für die Varianzaufklärung in der Population über:

$$\hat{\omega}^2 = \frac{QS_A - (p - 1) \cdot \hat{\sigma}_{Fehler}^2}{QS_{tot} + \hat{\sigma}_{Fehler}^2}$$

Voraussetzungen:

In der H_0 wird behauptet, dass die Daten aus identischen Populationen stammen und in der H_1 , dass sich die normalverteilten Populationen nur bei ihren Mittelwerte unterscheiden. Daher werden als Voraussetzungen (nötig für F_{emp} , der ANOVA egal) genannt:

- Normalverteilte Fehlerkomponenten (= normalverteilte Populationen) sonst: Beeinflussung von Alpha & Beta
- Homogene Fehlervarianzen (= varianzhomogene Populationen) (bei gleichen Stichprobengrößen egal, sonst erhebliche Gefährdung des F-Tests)
- Unabhängige Fehlerkomponenten (= unabhängige Stichproben).

Ferner muss die AV intervallskaliert sein, das Skalenniveau der UV ist mindestens nominal.

Die Varianzanalyse gilt insgesamt als ziemlich robust, sofern die Stichproben gleich groß sind. Genauso (weitgehend) verhält es sich mit dem Scheffé-Test und Trendtests.

Es gibt übrigens auch ANOVAs, bei denen nur Mittelwerte, Varianzen und Stichproben bekannt sind, die einzelnen Werte aber nicht (z.B. bei Metaanalysen). Formeländerungen im Bortz S. 251

Einzelvergleiche:

Führt die einfaktorielle ANOVA zu einem signifikantem F-Wert, können wir davon ausgehen, dass sich mindestens zwei Mittelwerte unterscheiden. Die Frage ist nur, welche.

Durch Einzelvergleiche (oder auch Kontraste) finden wir heraus, zwischen welchen einzelnen Treatmentstufen signifikante Unterschiede bestehen.

A-priori (= spezifische Hypothesen im Vorhinein vorhanden) Einzelvergleiche erfordern aufwendige mathematische Verfahren (Bortz S. 253-260), laufen aber letztendlich auf die Teilung der Differenz (D) der jeweiligen Stufenmittelwerte (A_i) durch die Varianz der Streuung der Mittelwertsdifferenzen. Das Ganze kann über einen t- oder F-Test geprüft werden.

$$t = \frac{D}{\sqrt{\hat{\sigma}(D)}} \qquad F = \frac{D^2}{\hat{\sigma}(D)}$$

Alpha-Fehler-Korrektur:

Man könnte natürlich auch, um Einzelvergleiche durchzuführen, eine Anzahl(p) t-Tests rechnen.

ANOVA

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

t-Tests

$$H_0: \mu_1 = \mu_2 \quad H_0: \mu_2 = \mu_3 \quad H_0: \mu_3 = \mu_4$$

Wenn hierbei die globale Testentscheidung der ANOVA auf einem α Niveau von z.B. 5% getroffen werden soll, müssen die einzelnen Tests auf einem anderen, kleineren Niveau α' getestet werden, denn:

Wenn drei Tests auf jeweils $\alpha = 5\%$ durchgeführt würden, läge die Wahrscheinlichkeit, dass mindestens ein Test fälschlicherweise für die H_1 spricht, bei 14,26%:

$$1 - (1 - \alpha)^q = 1 - (1 - 0,05)^3 = 0,1426.$$

Diesem Problem der **α -Fehlerkumulierung** entgeht man, indem jeder Einzelvergleich auf α' durchgeführt wird:

$$\alpha' = 1 - (1 - \alpha)^{1/q} = 1 - (1 - 0,05)^{1/3} = 0,01695.$$

Aufaddiert ergibt das ein Alphaniveau von ungefähr 0,05.

Eine oftmals hinreichende Approximation zur Berechnung von α' bietet die

Bonferoni-Korrektur:

$$\alpha' = \frac{\alpha}{q}$$

Diese obigen Überlegungen gehen davon aus, dass die t-Tests voneinander unabhängig sind. **Sind Tests abhängig, wird das Alpha kleiner.** (Eine perfekte Abhängigkeit benötigt nur einen Test, da der alle anderen Ergebnisse perfekt vorhersagt.)
Deswegen sollte man bei abhängigen Tests die obigen Verfahren für unabhängige Tests benutzen, auf diese Weise ist man durch ein großes Alpha „auf der sicheren Seite“.

t-Tests und andere Einzelvergleichsverfahren können gegenüber der einfaktoriellen Varianzanalyse Vorteile bieten, wenn **vor** einer Untersuchung (**a priori**) spezifische Hypothesen über die

Mittelwertsunterschiede vorliegen. Wenn dies nicht der Fall ist, sollte die Varianzanalyse gerechnet werden, da sie ökonomischer ist und die Varianzaufklärung zu bestimmen erlaubt.

Eine bewiesene, a priori aufgestellte, Hypothese hat den höchsten denkbaren Informationsgehalt (ist nur sehr schwierig zu finden, a posteriori ist eben einfacher)

Ferner kann über Einzelvergleiche **im Nachhinein (a posteriori)** untersucht werden, welche Mittelwerte für ein signifikantes Ergebnis verantwortlich sind.

Das geht u.a. mit dem Scheffé-Test.

Ein a posteriori gefundenes Ergebnis zu einer a priori aufgestellten Hypothese zu erklären, ist der Wissenschaft nicht dienlich (100% aller Hypothesen würden bewiesen)
„Die Begründung und Überprüfung einer Hypothese mit ein und demselben Datensatz ist wissenschaftlich nicht haltbar.“ (Bortz)

Scheffé-Test:

Es gibt mehrere a-posteriori-Vergleichsverfahren, darunter den Scheffé-Test (andere Verfahren werden im Bortz nicht behandelt).

Der Scheffé-Test hat sich Verletzungen gegenüber als relativ robust erwiesen; er entscheidet tendenziell eher konservativ (= zugunsten der H_0). Aufgrund Letzterem **ist der Alphafehler des Scheffé-Tests (zwischen den Einzelvergleichen) nicht größer als der der ANOVA.**

Der vollständige Scheffé-Test erlaubt nicht nur den Vergleich von jeweils zwei Mittelwerten, sondern auch von Kombinationen mehrerer Mittelwerte folgender Art:

- (1) $(\bar{x}_1 + \bar{x}_2)/2 - (\bar{x}_3 + \bar{x}_4)/2$
- (2) $\bar{x}_2 - (\bar{x}_3 + \bar{x}_4)/2$
- (3) usw.

Da man solche Vergleiche (Kontraste) oftmals nicht sinnvoll durchführen kann, begnügt man sich in der Regel mit Paarvergleichen. Hierfür werden die Beträge aller Differenzen von je zwei Mittelwerten mit folgender kritischen Differenz verglichen:

$$Diff_{krit} = \sqrt{\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right) \cdot (p - 1) \cdot \hat{\sigma}_{Fehler}^2 \cdot F_{(p-1, N-p, 1-\alpha)}} \cdot$$

Wenn $|\bar{A}_i - \bar{A}_{i'}| > Diff_{krit}$, kann diese Differenz als maßgeblich für die Signifikanz der ANOVA angesehen werden.

Das bedeutet aber nicht, dass die Einzelvergleichshypothese bestätigt wurde, sondern nur, dass diese beiden Paare zu einem großen Teil für die signifikante H_1 der (einfaktoriellen) ANOVA verantwortlich sind.

Trendtests:

Trendtests dienen zur besseren Vorhersage der AV durch die UV(s).

Durch Trendtests wird die Treatmentquadratsumme in Anteile zerlegt, die auf verschiedene Trends (linear, quadratisch, kubisch, etc.) in den Mittelwerten der AV zurückzuführen sind.

Die Durchführung von Trendtests setzt voraus, dass neben der AV (wie immer bei ANOVAs) auch die UV intervallskaliert ist.

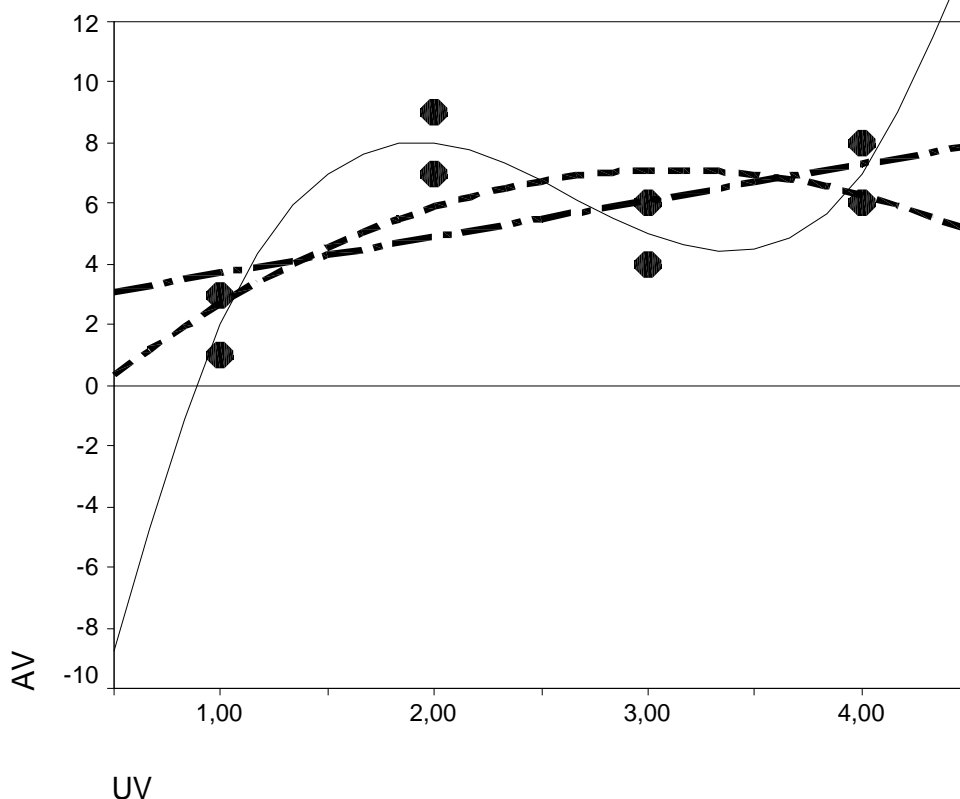
Wenn die UV intervallskaliert ist, kann über die Daten einer ANOVA auch eine Regression gerechnet werden.

Die allgemeine Regressionsgleichung lautet:

$$\hat{y} = a + b_1 \cdot x + b_2 \cdot x^2 + \dots + b_q \cdot x^q.$$

Bei p Faktorstufen lassen sich maximal $p - 1$ Trendkomponenten (hier: $p - 1 = q$) untersuchen.

Grafik (samt Werten) einer linearen, einer quadratischen und einer kubischen Regression.



Neben einem linearen Zusammenhang können in Datensätzen auch quadratische oder kubische Zusammenhänge bestehen. Je mehr Trendkomponenten berücksichtigt werden, desto besser gelingt die Vorhersage der AV, d.h. desto kleiner wird die Residualvarianz.

Es gibt auch die Möglichkeit, monotone Trends zu untersuchen. Dafür ist kein Intervall-, sondern nur Ordinal-, teils sogar nur Nominalskalenniveau nötig. Man vergleicht einfach nur die Mittelwerte, wobei man bei Nominalskalenniveau versucht, eine hypothetische Systematik (hier eine Rangfolge) vorherzusagen.

Durch eine monotone Trendhypothese wird eine Rangfolge der Treatment-Mittelwerte vorgegeben.

Zusammenhänge:

r^2 vs. η^2

Bei zwei Gruppen stimmt der Determinationskoeffizient r_{lin}^2 aus der linearen Regression mit der Varianzaufklärung aus der ANOVA überein:

$$r_{lin}^2 = \eta^2, \text{ wenn } p = 2.$$

Bei mehreren Gruppen ist dies in der Regel nicht der Fall:

$$r_{lin}^2 \neq \eta^2, \text{ wenn } p > 2.$$

F-Test in der Regression:

In der Regression gibt es folgende Varianzkomponentenzerlegung:

$$\begin{aligned} s_y^2 &= s_{\hat{y}}^2 + s_{y^*}^2 \\ \Leftrightarrow \frac{\sum (y_i - \bar{y})^2}{n} &= \frac{\sum (\hat{y} - \bar{y})^2}{n} + \frac{\sum (y^* - 0)^2}{n} \\ \Leftrightarrow \frac{QS_{tot}}{n} &= \frac{QS_{reg}}{n} + \frac{QS_{res}}{n}. \end{aligned}$$

Analog zum F-Test in der ANOVA lassen sich in der bivariaten Regression Varianzen schätzen und testen:

$$F = \frac{QS_{reg}/df_{reg}}{QS_{res}/df_{res}} = \frac{QS_{reg}/1}{QS_{res}/n-2}.$$

Wenn $F > F_{(1, n-2, 1-\alpha)}$, lässt sich schließen, dass die UV in der AV auf Populationsebene Varianz aufklärt.

Rangvarianzanalyse:

Die Rangvarianzanalyse (**Kruskal-Wallis-Test, H-Test**) ist ein nichtparametrisches Pendant zur einfaktoriellen ANOVA.

Nicht über diesen Punkt wundern, er steht nicht im Bortz, wurde aber in den B3-Folien behandelt (ganz am Ende).

Hypothese:

H_0 : Die p Stichproben stammen aus Populationen mit denselben zentralen Tendenzen.

Prüfgröße

Die Herleitung der Prüfgröße H orientiert sich an der parametrischen ANOVA. Wenn die Populationsvarianz bekannt ist, kann die $H_0: \mu_1 = \mu_2 = \dots = \mu_p$ mit folgender χ^2 -verteilten Prüfgröße getestet werden:

$$\chi^2 = \sum_{i=1}^p \frac{(\bar{x}_i - \bar{x})^2}{\sigma^2/n_i}$$

Wie beim U -Test werden beim H -Test alle Messwerte in eine gemeinsame Rangreihe gebracht. Diese Ränge werden von 1 bis N verteilt, so dass der kleinste Wert die 1, der nächst größere die 2 usw. erhält.

Die Variablen der obigen Gleichung werden beim H -Test durch die entsprechenden Variablen ersetzt, die sich bei einer Berechnung über Ränge ergeben:

$$H = \frac{N-1}{N} \cdot \sum_{i=1}^p \frac{(\bar{R}_i - \bar{R})^2}{\sigma_R^2/n_i}$$

hierbei sind

- \bar{R}_i der Mittelwert der Rangplätze in Gruppe i ,
- $\bar{R} = (N+1)/2$ das Gesamtmittel aller Rangplätze,
- $\sigma_R^2 = (N^2 - 1)/12$ die Varianz aller Rangplätze
- $\frac{N-1}{N}$ eine Korrektur bezüglich der endlichen Grundgesamtheit.

Da \bar{R} und σ_R^2 nur von N abhängen, sind für H folgende Schreibweisen möglich:

$$H = \frac{N-1}{N} \cdot \sum_{i=1}^p \frac{(\bar{R}_i - \frac{N+1}{2})^2}{\frac{N^2-1}{12}/n_i} = \frac{(N-1) \cdot \sum_{i=1}^p n_i (\bar{R}_i - \frac{N+1}{2})^2}{\frac{N^3-N}{12}}$$

Entscheidung:

Unter der H_0 ist die Statistik H asymptotisch χ^2 -verteilt mit $df = p - 1$:

$$\text{Falls } H > \chi_{p-1, 1-\alpha}^2, \text{ Annahme der } H_1.$$

Voraussetzungen:

Der H -Test setzt voraus, dass die Messwerte aus stetigen Verteilungen stammen, die dieselbe Form besitzen und sich nur hinsichtlich ihrer Lage bzw. zentralen Tendenz unterscheiden. Im Vergleich zur ANOVA liegt dem H -Test also ein weniger restriktives Modell zugrunde, denn bei der ANOVA wird zusätzlich vorausgesetzt, dass die Verteilungen normal sind.

Der hier beschriebene asymptotische Test sollte nur eingesetzt werden, wenn die Stichproben nicht zu klein sind ($n_i > 5$ oder $p \geq 4$). Bei kleineren Stichproben sollte die exakte Verteilung von H nach der Randomisationsmethode bestimmt werden.

Mehrfaktorielle Versuchspläne:

Wird eine einfaktorielle ANOVA nicht signifikant (die Fehlervarianz ist zu groß), kann dies (u.a.) auf das Einwirken anderer Faktoren zurückzuführen. Also machen wir doch einfach eine mehrfaktorielle ANOVA.-

Ein Vorteil der mehrfaktoriellen ANOVA besteht darin, auch das Zusammenwirken (die Interaktion) mehrerer Variablen zu erfassen.

Zu beachten ist jedoch, dass z.B. bei 4 dreifach gestuften Faktoren $3 \times 3 \times 3 \times 3 = 3^4 = 81$ Gruppen untersucht werden müssen. Da n auch eine gewisse Größe haben sollte, müsste man bei z.B. $n = 10$ insgesamt 810 Leute untersuchen.

Deswegen ist es wichtig, in der Planung genau zu überlegen, welche und ob Faktoren wichtig sein können, was unnötig ist und was nicht (das ist jetzt eine Sache der Methodenlehre).

Die mehrfaktorielle Varianzanalyse entspricht der einfaktoriellen Varianzanalyse, sie ist eine Art Erweiterung. Sie baut auf dem selben System wie die einfaktorielle ANOVA auf, alle Definitionen und Bezeichnungen sind identisch, einzig die Formeln sind leicht verändert, und es kommt die Quadratsumme der Interaktion sowie weitere Treatmentquadratsummen hinzu. Das ist alles.

Zweifaktorielle Varianzanalyse:

Mit der zweifaktoriellen Varianzanalyse wird überprüft, wie eine AV von zwei UVn (Faktoren) beeinflusst wird. Hierbei können nicht nur die Effekte der beiden Faktoren analog zur einfaktoriellen ANOVA getrennt beurteilt werden, sondern auch die Wechselwirkung (die Interaktion) der Faktoren kann getestet werden.

Im folgenden wird nur der Fall gleichgroßer Zellenbesetzungen $n_{ij} = n$ behandelt. Für ungleichgroße Zellen gilt das beschriebene Verfahren – im Gegensatz zur einfaktoriellen ANOVA – nicht!

Hypothesen:

Faktor A	$H_{0A}: \mu_1 = \mu_2 = \dots = \mu_p$ $H_{1A}: \mu_i \neq \mu_i'$
Faktor B	$H_{0B}: \mu_1 = \mu_2 = \dots = \mu_q$ $H_{1B}: \mu_j \neq \mu_j'$
Interaktion $A \times B$	$H_{0A \times B}: \mu_{ij} = \mu_i + \mu_j - \mu$ $H_{1A \times B}: \mu_{ij} \neq \mu_i + \mu_j - \mu$ für mindestens ein Paar i, j

Datenschema:

		Faktor A					
		a_1	a_2	...	a_i	...	a_p
Faktor B	b_1	x_{111}	x_{211}		x_{i11}		x_{p11}
		x_{112}	x_{212}		x_{i12}		x_{p12}
		\vdots	\vdots		\vdots		\vdots
		x_{11m}	x_{21m}	...	x_{i1m}	...	x_{p1m}
		\vdots	\vdots		\vdots		\vdots
		x_{11n}	x_{21n}		x_{i1n}		x_{p1n}
		\vdots	\vdots		\vdots		\vdots
	b_j	x_{1j1}	x_{2j1}		x_{ij1}		x_{pj1}
		x_{1j2}	x_{2j2}		x_{ij2}		x_{pj2}
		\vdots	\vdots		\vdots		\vdots
		x_{1jm}	x_{2jm}	...	x_{ijm}	...	x_{pjm}
		\vdots	\vdots		\vdots		\vdots
x_{1jn}		x_{2jn}		x_{ijn}		x_{pjn}	
b_q	x_{1q1}	x_{2q1}		x_{iq1}		x_{pq1}	
	x_{1q2}	x_{2q2}		x_{iq2}		x_{pq2}	
	\vdots	\vdots		\vdots		\vdots	
	x_{1qm}	x_{2qm}	...	x_{iqm}	...	x_{pqm}	
	\vdots	\vdots		\vdots		\vdots	
	x_{1qn}	x_{2qn}		x_{iqn}		x_{pqn}	

- Der Faktor A wird nummeriert von 1 bis p.
- Der Faktor B wird nummeriert von 1 bis q.
- Das Zeichen i steht für ein beliebiges A.
- Das Zeichen j steht für ein beliebiges B.
- Die Messwerte x (VPN) in den einzelnen Bedingungen werden nummeriert von 1 bis n.
- Das Zeichen m steht für ein beliebiges x.

Quadratsummen:

In dem Modell der zweifaktoriellen ANOVA werden vier Quellen der Variation unterschieden:

Quadratsumme des Haupteffektes A :

$$QS_A = n \cdot q \cdot \sum_{i=1}^p (\bar{A}_i - \bar{G})^2.$$

Quadratsumme des Haupteffektes B:

$$QS_B = n \cdot p \cdot \sum_{j=1}^q (\bar{B}_j - \bar{G})^2.$$

Quadratsumme der Interaktion A×B:

$$QS_{A \times B} = n \cdot \sum_{i=1}^p \sum_{j=1}^q (\overline{AB}_{ij} - \overline{AB}_{ij}')^2.$$

\overline{AB}_{ij} ist der beobachtete Zellenmittelwert (einfach in der Tabelle nachsehen) und \overline{AB}_{ij}' ist der zu erwartende Zellenmittelwert für den Fall, dass es keine Interaktion gibt.

Wenn nur die Faktoren A und B wirken würden, wären folgende Zellenmittelwerte zu erwarten:

$$\overline{AB}_{ij}' = \bar{A}_i + \bar{B}_j - \bar{G}.$$

Die Interaktion misst Auswirkungen, die durch die Kombination einzelner Faktorstufen entsteht. (z.B. Medikament B1 wirkt bei Frauen besser, Medikament B2 bei Männern)

Fehlerquadratsumme:

$$QS_{Fehler} = \sum_{i=1}^p \sum_{j=1}^q \sum_{m=1}^n (x_{ijm} - \overline{AB}_{ij}')^2.$$

Die **totale Quadratsumme** berechnet sich über:

$$QS_{total} = \sum_{i=1}^p \sum_{j=1}^q \sum_{m=1}^n (x_{ijm} - \bar{G})^2.$$

Quadratsummenzerlegung:

Sofern alle Zellen gleich stark besetzt sind, besteht folgende additive Beziehung:

$$QS_{tot} = QS_A + QS_B + QS_{A \times B} + QS_{Fehler} .$$

Die Fehlerquadratsumme ist bei der zweifaktoriellen ANOVA maximal gleich groß, meist aber kleiner als die Fehlerquadratsumme bei der einfaktoriellen ANOVA.

Sie hat sich um die Quadratsumme (QS_{Zellen}) der erwarteten Zellenmittelwerte \overline{AB}_{ij} vermindert.

$$QS_{tot} = QS_{Zellen} + QS_{Fehler}$$

Freiheitsgrade und Varianzschätzungen:

Die Freiheitsgrade verhalten sich genauso additiv wie die Quadratsummen.

Quelle	Varianzen	Freiheitsgrade
<i>A</i>	$\hat{\sigma}_A^2 = \frac{QS_A}{df_A}$	$df_A = p-1$
<i>B</i>	$\hat{\sigma}_B^2 = \frac{QS_B}{df_B}$	$df_B = q-1$
<i>A</i> × <i>B</i>	$\hat{\sigma}_{A \times B}^2 = \frac{QS_{A \times B}}{df_{A \times B}}$	$df_{A \times B} = (p-1) \cdot (q-1)$
<i>Fehler</i>	$\hat{\sigma}_{Fehler}^2 = \frac{QS_{Fehler}}{df_{Fehler}}$	$df_{Fehler} = p \cdot q \cdot (n-1)$

Hypothesenprüfung:

Sofern mit der ANOVA nur Aussagen über die realisierten Treatmentstufen angestrebt werden, sind bei Gültigkeit der H_0 sämtliche Varianzen erwartungstreue Schätzer der (allen Zellen) gemeinsamen Populationsvarianzen. Die Entscheidungsregeln lauten daher:

Annahme der ...	
H_{1A} falls ...	$F = \frac{\hat{\sigma}_A^2}{\hat{\sigma}_{Fehler}^2} > F_{(df_A, df_{Fehler}, 1-\alpha)}$
H_{1B} falls ...	$F = \frac{\hat{\sigma}_B^2}{\hat{\sigma}_{Fehler}^2} > F_{(df_B, df_{Fehler}, 1-\alpha)}$
$H_{1A \times B}$ falls ...	$F = \frac{\hat{\sigma}_{A \times B}^2}{\hat{\sigma}_{Fehler}^2} > F_{(df_{A \times B}, df_{Fehler}, 1-\alpha)}$

Varianzaufklärung:

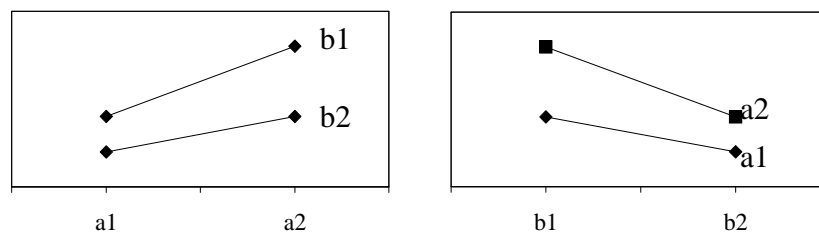
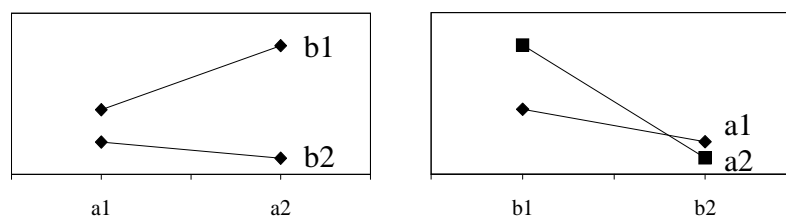
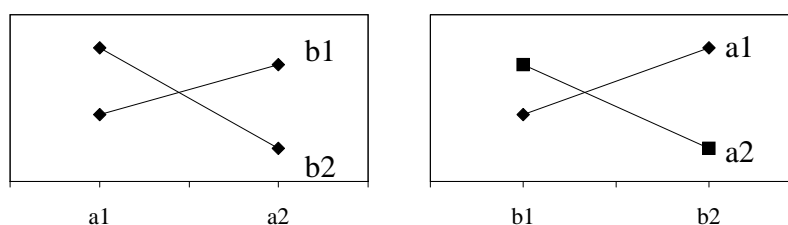
Wie viel der Gesamtvarianz über die einzelnen Treatments aufgeklärt wird, erfasst man mit derselben Formel wie bei der einfaktoriellen ANOVA:

$$\eta_A^2 = \frac{QS_A}{QS_{tot}} \quad \text{bzw.} \quad \eta_B^2 = \frac{QS_B}{QS_{tot}} \quad \text{bzw.} \quad \eta_{A \times B}^2 = \frac{QS_{A \times B}}{QS_{tot}}$$

$$\eta^2 = \eta_A^2 + \eta_B^2 + \eta_{A \times B}^2 = \frac{QS_A}{QS_{total}} + \frac{QS_B}{QS_{total}} + \frac{QS_{A \times B}}{QS_{total}}$$

Interaktionsdiagramme:

Um signifikante Ergebnisse besser interpretieren zu können, stellt man sie grafisch dar. Verlaufen die Geraden nicht parallel, besteht zwischen den Faktoren eine Interaktion

Ordinal**Hybrid bzw. semidisordinal****Disordinal**

Da die Interpretation von Interaktionseffekten häufig durch die gleichzeitige Berücksichtigung von Haupteffekten verfälscht wird, nimmt man gelegentlich die Interpretation an den Diagrammen der residualen Mittelwerte (= empirischer minus erwarteter Mittelwert) vor.

Wenn eine Interaktion signifikant geworden ist, erstellt man ein Interaktionsdiagramm. Dabei gibt es drei mögliche Interaktionen:

1.) Ordinal:

Bei einer ordinalen Interaktion weisen beide Linienzüge den gleichen Trend auf, so dass beide in einem Diagramm steigen und beide im anderen Diagramm fallen.

Beide Diagramme sind interpretierbar.

2.) Hybrid bzw. semidisordinal:

Bei einer hybriden Interaktion zeigen die Linienzüge in einem Diagramm gegenläufige Trends auf, so dass sie sich im anderen Diagramm schneiden.

Der eine Haupteffekt (hier: B) ist eindeutig interpretierbar (hier links: b_2 immer kleiner als b_1 ; siehe letzte Seite), der andere Haupteffekt (hier A) sollte nicht interpretiert werden.

3.) Disordinal:

Bei einer disordinalen Interaktion schneiden sich in beiden Diagrammen die Geraden.

Die Interpretation der Haupteffekte ist hier unmöglich, man kann höchstens die Unterschiede von (z.B.) b_1 und b_2 im Vergleich mit den Stufen des Faktors A sinnvoll interpretieren.

Feste und zufällige Faktoren:

Bisher war es so, dass man bei der ANOVA nur (inferenzstatistisch durch F-Test abgesicherte) Aussagen über die in einer Untersuchung aufgestellten Treatmenteffekte machen wollte.

Nun möchte man gelegentlich Aussagen generalisieren können, auch wenn man nicht alle Möglichkeiten untersucht.

Dazu ist es wichtig, feste und zufällige Faktoren zu klassifizieren:

- Von einem **festen Faktor** (fixed factor) spricht man, wenn es bei der inferenzstatistischen Absicherung nur um den Vergleich der tatsächlich realisierten Stufen dieses Faktors geht. D.h. man stellt vorher **festen Faktoren** auf, z.B. die Faktoren Verhaltens-, Gestalt- und analytischer Therapie. Diese werden dann untersucht, ohne dass man über andere Therapieformen Aussagen treffen will.
- Bei einem **zufälligen Faktor** (random factor) geht es nicht um den Vergleich von bestimmten Treatmentstufen, sondern es soll auf alle möglichen, auch nicht realisierte Stufen des Faktors geschlossen werden. D.h. wir untersuchen den Einfluss von Therapeuten auf den Therapieerfolg bei z.B. drei Therapeuten, aber wollen nun eine generelle Aussage über den Einfluss von jedem (und nicht nur diesen drei) Therapeuten aufstellen.

Die Interpretation eines zufälligen Faktors ist somit viel weitergehend als die eines festen Faktors. Deswegen gibt es auch eine weitere Voraussetzung:

Die mit allen möglichen Faktorstufen verbundenen Treatmenteffekte müssen normalverteilt sein.

Bei der zweifaktoriellen ANOVA lassen sich drei Modelle unterscheiden, wobei sich die Prüfvarianzen wie folgt unterscheiden:

zu prüfende Varianz	Prüfvarianz		
	Modell I A fest B fest	Modell II A fest B zufällig	Modell III A zufällig B zufällig
$\hat{\sigma}_A^2$	$\hat{\sigma}_{Fehler}^2$	$\hat{\sigma}_{A \times B}^2$	$\hat{\sigma}_{A \times B}^2$
$\hat{\sigma}_B^2$	$\hat{\sigma}_{Fehler}^2$	$\hat{\sigma}_{Fehler}^2$	$\hat{\sigma}_{A \times B}^2$
$\hat{\sigma}_{A \times B}^2$	$\hat{\sigma}_{Fehler}^2$	$\hat{\sigma}_{Fehler}^2$	$\hat{\sigma}_{Fehler}^2$

Beispiel am **Modell II**: Ein signifikanter fester Faktor A besagt in diesem Fall, dass die Unterschiede zwischen den einzelnen Faktorstufen von A (also A_1, A_2, A_3 , etc.) nicht nur bei den zufällig untersuchten Faktorstufen von B gelten, sondern bei allen möglichen B.

Das Ganze gilt aber nur, wenn die Interaktion (zw. A & B) nicht signifikant ist.

Wichtig ist, dass der feste Faktor gegen die Fehlervarianz signifikant ist (d.h. es gibt tatsächlich Unterschiede bei A), da es sonst zu der Aussage kommen kann, dass bei festem Faktor B nichts signifikant ist, bei zufälligem Faktor schon: „Bei bestimmten B hat der Faktor A keinen signifikanten Einfluss, bei allen möglichen B hat der Faktor A einen signifikanten Einfluss.“

Bei der einfaktoriellen ANOVA wird auch die Varianz eines zufälligen Faktors gegen die Fehlervarianz getestet, dementsprechend ist es egal, ob man da mit festen oder zufälligen Faktoren rechnet.

Effektgrößen und Einzelvergleiche:

Effektgrößen spielen auch bei der mehrfaktoriellen ANOVA eine Rolle, wer mehr wissen will lese nach im Bortz auf S.293 / 294.

Durch **Einzelvergleiche** bei der mehrfaktoriellen ANOVA finden wir heraus, zwischen welchen einzelnen Treatmentstufen und bei welchen Interaktionen signifikante Unterschiede bestehen.

A-priori Einzelvergleiche erfordern aufwendige mathematische Verfahren (Bortz S. 295-301), man arbeitet hierbei wieder mit den Differenzen (D) zwischen den einzelnen Stufen.

A posteriori Einzelvergleiche sind auch bei der mehrfaktoriellen ANOVA mit Variationen des Scheffé-Tests möglich.

Drei- und mehrfaktorielle Varianzanalysen:

Drei- und mehrfaktorielle Varianzanalysen laufen vom Prinzip genau wie eine zweifaktorielle ANOVA ab. Zu beachten ist neben den weiteren Faktoren (QS_C , etc.) noch die Interaktion. Es gibt bei der drei- und mehrfaktoriellen ANOVA Interaktionen 2. und höherer Ordnung; d.h. Interaktionen nicht nur zwischen zwei, sondern auch drei und mehr Faktoren.

Alle Interaktionen müssen abgedeckt sein, um QS_{tot} zu erreichen:

$$QS_{tot} = QS_A + QS_B + QS_C + QS_{A \times B} + QS_{A \times C} + QS_{B \times C} + QS_{A \times B \times C} + QS_{Fehler} .$$

Die drei- und mehrfaktorielle Varianzanalyse gleicht der zweifaktoriellen Varianzanalyse, nur muss man bei ihr viel mehr rechnen.

Dementsprechend gibt es z.B. auch mehr Nullhypothesen, hier oben wären es z.B. 7 H_0 .

Effektgrößen spielen auch bei der drei- und mehrfaktoriellen ANOVA eine Rolle, wer mehr wissen will lese nach im Bortz auf S.305.

A-priori Einzelvergleiche laufen wie bei der zweifaktoriellen ANOVA ab, sie erfordern aufwendige mathematische Verfahren (Bortz S. 295-301; 304), man arbeitet hierbei wieder mit den Differenzen (D) zwischen den einzelnen Stufen.

A posteriori Einzelvergleiche sind auch bei der drei- und mehrfaktoriellen ANOVA mit weiteren Variationen des Scheffé-Tests möglich.

Ungleiche Stichprobengrößen:

Die ganzen bisherigen Verfahren der mehrfaktoriellen Varianzanalyse versagen bei ungleich großen Stichproben. Sind die Stichprobengrößen ungleich, ist eine wichtige Eigenschaft der ANOVA, nämlich die Unabhängigkeit (Orthogonalität) der Haupteffekte und der Interaktionseffekte, nicht mehr gegeben. Deswegen werden solche ANOVAs auch als nichtorthogonale Varianzanalysen bezeichnet. Hierbei gibt es (theoretisch) 4 Verfahren:

I.) Missing-data Techniken:

Missing-data Techniken sind eigentlich generiert worden für ursprünglich gleichgroße Stichproben, bei denen ein zu vernachlässigender Prozentsatz der Daten unbrauchbar ist.

Man ersetzt die fehlenden Daten durch den jeweiligen Mittelwert. Dabei ist wichtig zu beachten, dass somit der F-Test eher progressiv (zugunsten der H_1) entscheidet.

II.) Proportional geschichtete Stichproben:

Stehen die ungleich großen Stichproben (regelmäßig) in einem bestimmten Verhältnis zueinander (also z.B. 1 : 4) kann man über ein paar einfache Formelmodifikationen (Bortz S. 311) eine ANOVA berechnen. Das tolle hierbei: **Diese ANOVA gehört noch zu den orthogonale Varianzanalysen!**

III.) Ausgleich durch das harmonische Mittel:

Sind die Stichproben ungleich und nicht proportional geschichtet, kann man die einzelnen Stichprobenumfänge durch das harmonische Mittel (Bortz S.40) aller Stichprobenumfänge ersetzen. Ist das Verhältnis vom größten zum kleinsten Stichprobe kleiner als 5, kommt man zu halbwegs genauen Ergebnissen, sofern man den Formeln im Bortz S. 311/ 312 folgt.

IV.) Das allgemeine lineare Modell (ALM):

Dies ist eine Technik der „Multivariaten Verfahren“, es wird in einem eigenen Kapitel im B4-Teil behandelt.

Voraussetzungen:

Die zugrundeliegenden Modelle der mehrfaktoriellen Varianzanalysen sind letztlich gleich mit dem der einfaktoriellen ANOVA, d.h. man nimmt mit den Nullhypothesen an, dass die Daten je Zelle aus identischen Normalverteilungen stammen. Also lauten die Voraussetzungen wiederum

- Normalverteilung,
- Varianzhomogenität,
- Unabhängigkeit.

Es gibt übrigens sogar ANOAVAs mit $n = 1$ Stichprobengrößen. Dies ist möglich durch spezielle Formeln, die man auf S. 314 – 316 im Bortz findet.

Versuchspläne mit Messwiederholungen:

In der u.a. psychologischen Forschung werden oftmals Tests mit mehreren Messzeitpunkten durchgeführt, z.B. in der Therapieforschung (Wirksamkeit der Therapie), in der Gedächtnisforschung (Lernerfolge), in der Wahrnehmungspsychologie (Wiedererkennung), usw.

Um diese Tests berechnen zu können, benötigt man die Varianzanalyse mit Messwiederholungen.

Einfaktorielle Varianzanalyse mit Messwiederholung:

Die einfaktorielle Varianzanalyse mit Messwiederholung prüft, ob zwischen den Mittelwerten von p abhängigen Stichproben Unterschiede bestehen.

Bei $p = 2$ entspricht die einfaktorielle Varianzanalyse mit Messwiederholung dem t-Test für abhängige Stichproben.

Die einfaktorielle Varianzanalyse mit Messwiederholung entspricht der einfaktorielle Varianzanalyse ohne Messwiederholung, nur wird ein neuer Faktor eingeführt, der sogenannte **Blockfaktor** P . Da wir bei der einfaktoriellen ANOVA mit Messwiederholung mit Veränderungen an (z.B.) Personen rechnen, müssen die x -Werte nicht nur anhand der Faktorstufen A_i , sondern auch hinsichtlich der jeweiligen Person P_m gekennzeichnet und geordnet werden.

Bei (Versuchs-) Personen wäre P_1 die Versuchsperson Nr. 1, welche eine Anzahl p von Messungen A (also A_1 bis A_p) über sich ergehen lassen muss.

Die Abweichungen der jeweiligen Personen vom Gesamtmittelwert ergeben den Blockfaktor.

Datenschema:

Vpn	Faktor A						Σ
	a_1	a_2	...	a_i	...	a_p	
1	x_{11}	x_{12}	...	x_{1i}	...	x_{1p}	P_1
2	x_{21}	x_{22}	...	x_{2i}	...	x_{2p}	P_2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
m	x_{m1}	x_{m2}	...	x_{mi}	...	x_{mp}	P_m
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	x_{n1}	x_{n2}	...	x_{ni}	...	x_{np}	P_n
Σ	A_1	A_2		A_i		A_p	

$$\bar{P}_m = \frac{1}{p} \cdot \sum_{i=1}^p x_{im}$$

$$\bar{A}_i = \frac{1}{n} \cdot \sum_{m=1}^n x_{im}$$

$$\bar{G} = \frac{1}{n \cdot p} \cdot \sum_{i=1}^p \sum_{m=1}^n x_{im}$$

Während die Stufen des Faktors A in der Regel aus bestimmten Gründen vom Untersucher festgesetzt werden (z.B. Zeitpunkt 1, 2, etc.), stellen die Vpn eine Zufallsauswahl aus einer Population dar. Somit ist A ein fester Faktor, die Variable Versuchspersonen stellt einen zufälligen Faktor dar (= Blockfaktor P).

Hypothesen:

Die einfaktorielle Varianzanalyse mit Messwiederholung prüft (wie die einfaktorielle Varianzanalyse ohne Messwiederholung), ob sich mindestens zwei Mittelwerte unterscheiden:

$$H_0 : \mu_1 = \dots = \mu_i = \dots = \mu_p$$

$$H_1 : \mu_i \neq \mu_{i'}$$

Bei Untersuchungen der Depressivität über einen Therapiezeitraum hinweg würde die Nullhypothese demnach behaupten, dass die Therapie keinen Einfluss auf die Depressivität hat.

Quadratsummen:

QS_{Res} der Varianzanalyse mit Messwiederholung entspricht dem QS_{Fehler} der Varianzanalyse ohne Messwiederholung – es erfasst die individuellen Unterschiede. Zusätzlich erfasst es sogar noch die Interaktionseffekte zwischen Treatment und Vpn.

Treatment:	$QS_A = n \cdot \sum_{i=1}^p (\bar{A}_i - \bar{G})^2$
Blockfaktor:	$QS_P = p \cdot \sum_{m=1}^n (\bar{P}_m - \bar{G})^2$
Residual:	$QS_{Res} = \sum_{i=1}^p \sum_{m=1}^n (x_{mi} - \bar{A}_i - \bar{P}_m + \bar{G})^2$
Gesamt:	$QS_{tot} = \sum_{i=1}^p \sum_{m=1}^n (x_{mi} - \bar{G})^2$

\bar{P}_m ist der Mittelwert aller Messwerte der Vp m.

Quadratsummenzerlegung:

$$QS_{tot} = QS_A + QS_P + QS_{Res}$$

Die Freiheitsgrade verhalten sich genauso.

Varianzschätzungen:

Quelle		Freiheitsgrade
Treatment	$\hat{\sigma}_A^2 = \frac{QS_A}{df_A}$	$df_A = p - 1$
Residual	$\hat{\sigma}_{Res}^2 = \frac{QS_{Res}}{df_{Res}}$	$df_{Res} = (n - 1) \cdot (p - 1)$

Unter der H_0 gilt: $E(\hat{\sigma}_A^2) = E(\hat{\sigma}_{Res}^2) = \sigma_{Res}^2$.

Prüfung der Nullhypothese:

Annahme der H_1 , falls

$$F = \frac{\hat{\sigma}_A^2}{\hat{\sigma}_{Res}^2} > F_{(df_A, df_{Res}, 1-\alpha)}$$

Dieser F-Bruch kann auch zum Test der Nullhypothese herangezogen werden, wenn A ein zufälliger Faktor ist.

Trendtests und Einzelvergleiche:

Ebenso wie bei der einfaktoriellen ANOVA ohne Messwiederholung können auch bei dieser ANOVA Trendtests, a priori Einzelvergleiche und a posteriori Einzelvergleiche (= Scheffé-Test) durchgeführt werden.

Als Prüfvarianz muss jedoch $\hat{\sigma}_{Res}^2$ anstatt $\hat{\sigma}_{Fehler}^2$ eingesetzt werden.

Ipsative Daten:

Eine einfaktoriellen ANOVA mit Messwiederholung lässt sich auch als einfaktoriellen ANOVA ohne Messwiederholung darstellen, wenn man die ursprünglichen Messungen der Vpn „ipsativiert“. Dabei wird von jedem individuellen Messwert p der Durchschnitt \bar{P}_m abgezogen.

Mehrfaktorielle Varianzanalyse mit Messwiederholung:

Die mehrfaktorielle ANOVA mit Messwiederholung ist einfach eine Erweiterung der einfaktoriellen ANOVA mit Messwiederholung; es treten wieder Interaktionseffekte auf, die beachtet werden müssen.

A bezeichnet weiterhin den Messwiederholungsfaktor, B ist der sogenannte Gruppierungsfaktor.

Hypothesen:

Die Hypothesen sind mit der zweifaktoriellen ANOVA ohne Messwiederholung identisch.

Datenschema:

Das Datenschema ist im Vergleich zur einfaktoriellen ANOVA mit Messwiederholung um den Faktor B erweitert worden.

	Faktor A					
Faktor B	a_1	...	a_i	...	a_p	Σ
b_1	x_{111}	...	x_{i11}	...	x_{p11}	P_{11}
	\vdots		\vdots		\vdots	\vdots
	x_{11m}	...	x_{i1m}	...	x_{p1m}	P_{1m}
	\vdots		\vdots		\vdots	\vdots
x_{11n}	...	x_{i1n}	...	x_{p1n}	P_{1n}	
\vdots	\vdots		\vdots		\vdots	\vdots
b_j	x_{1j1}	...	x_{ij1}	...	x_{pj1}	P_{j1}
	\vdots		\vdots		\vdots	\vdots
	x_{1jm}	...	x_{ijm}	...	x_{pjm}	P_{jm}
	\vdots		\vdots		\vdots	\vdots
x_{1jn}	...	x_{ijn}	...	x_{pjn}	P_{jn}	
\vdots	\vdots		\vdots		\vdots	\vdots
b_q	x_{1q1}	...	x_{iq1}	...	x_{pq1}	P_{q1}
	\vdots		\vdots		\vdots	\vdots
	x_{1qm}	...	x_{iqm}	...	x_{pqm}	P_{qm}
	\vdots		\vdots		\vdots	\vdots
x_{1qn}	...	x_{iqn}	...	x_{pqn}	P_{qn}	

Quadratsummen:

Die Quadratsummen werden erweitert um QS_{inS} (= Unterschiede zwischen den VP innerhalb der einzelnen Stichproben, also z.B. in der Stichprobe A_1, B_2) und

$QS_{A \times V_{pn}}$ (= spezifische Reaktionsweise der Vpn auf die Faktorstufen A).

A	$QS_A = q \cdot n \cdot \sum_i^p (\bar{A}_i - \bar{G})^2$
B	$QS_B = p \cdot n \cdot \sum_j^q (\bar{B}_j - \bar{G})^2$
$A \times B$	$QS_{A \times B} = n \cdot \sum_i^p \sum_j^q (\bar{AB}_{ij} - \bar{A}_i - \bar{B}_j + \bar{G})^2$
$A \times Vpn$	$QS_{A \times Vpn} = \sum_i^p \sum_j^q \sum_m^n (x_{ijm} - \bar{AB}_{ij} - \bar{P}_{jm} + \bar{B}_j)^2$
in S	$QS_{inS} = p \cdot \sum_j^q \sum_m^n (\bar{P}_{jm} - \bar{B}_j)^2$
Gesamt	$QS_{tot} = \sum_i^p \sum_j^q \sum_m^n (x_{ijm} - \bar{G})^2$

Quadratsummenzerlegung:

$$QS_{tot} = QS_A + QS_B + QS_{A \times B} + QS_{A \times Vpn} + QS_{inS}$$

Varianzschätzung:

Läuft genauso ab wie immer: jeweilige QS geteilt durch ihre df.

Prüfung der Nullhypothese:

Läuft genauso ab wie immer; man beachte nur, dass anstelle von QS_{Res} nun $QS_{A \times Vpn}$ die Prüfgröße darstellt.

$$F = \frac{\hat{\sigma}_A^2}{\hat{\sigma}_{A \times Vpn}^2} > F_{(df_A, df_{A \times Vpn}, 1-\alpha)} \Rightarrow \text{Annahme der } H_1$$

Trendtests und Einzelvergleiche:

Sind ebenso durch einfache Formelmodifikationen möglich (siehe Bortz S.327 ff.)

Mathematik total:**Dreifaktorielle Varianzanalysen mit Messwiederholung:**

Bei diesem Fall wird es ganz garstig, es gibt bei dieser Erweiterung z.B. mehrere Möglichkeiten, unterschiedlich viele Faktoren als Messwiederholungsfaktoren zu benennen.

Wer Extremsport liebt: Bortz S. 330-339

ungleiche Stichprobenumfänge:

Auch bei der mehrfaktoriellen ANOVA mit Messwiederholung gibt es Verfahren, um ungleich große Stichproben berechnen zu können, sofern einige wechselnde Bedingungen erfüllt sind (Bortz S. 329-330; S. 331-332; S. 337-338).

Bei einer einfaktoriellen Varianzanalyse mit Messwiederholung müssen die Stichproben gleich groß sein, denn sonst würden einige V_p nicht alle Messzeitpunkte absolviert haben.

Voraussetzungen:

- **Normalverteilung der Fehler- bzw. Residualkomponenten.**
- **Homogene Varianzen unter den Faktorstufen.**
- **Homogene Korrelationen zwischen den Faktorstufen.** (da Unabhängigkeit zwischen den Faktorstufen bei Messwiederholung ja schwierig ist)

Wären die Korrelationen Null (also gäbe es Unabhängigkeit), hätten wir identische Ergebnisse wie bei der ANOVA ohne Messwiederholung.

Bei Verletzung der oben genannten Bedingungen entscheiden die F-Tests progressiver, d.h. eher zugunsten der H_1 .

Sind neben homogenen Korrelationen auch homogene Varianzen nicht vorhanden, kann man eine ϵ -Korrektur anwenden.

$\frac{1}{p-1} \leq \epsilon \leq 1$, dieser Wert ϵ wird geschätzt und dann mit den Freiheitsgraden multipliziert, diese werden kleiner, die Entscheidung konservativer, d.h. später signifikant, und auf diese Weise gelangt man auch bei solchen Verletzungen zu (halbwegs) funktionalen Tests.
(siehe Bortz S. 342-346)

Kovarianzanalyse:

Um die Fehlervarianz zu reduzieren – und somit bessere Reliabilität (Zuverlässigkeit der Daten) und Validität (Gültigkeit inhaltlicher Schlussfolgerungen) zu erhalten –, gibt es mehrere Möglichkeiten:

- Man könnte mehrere Faktoren einführen, jedoch benötigt man aber dafür viele Vpn.
- Man könnte Messwiederholungen durchführen, jedoch kann es da zu Ermüdungserscheinungen seitens der Vpn (und somit zu Ergebnisverzerrung) führen.

Doch wie immer gibt es bei solchen Einleitungen auch eine Lösung, die die eben genannten Nachteile nicht hat:

DIE KOVARIANZANALYSE (ANCOVA)

Der Vorteil hierbei ist, dass die Zahl der Vpn nicht erhöht werden muss, und abgesehen von der Erhebung der KV keine weiteren Erhebungen oder Nachuntersuchungen folgen müssen.

Die Kovarianzanalyse (ANCOVA) bietet die Möglichkeit, neben der UV den potentiellen Einfluss einer weiteren Variablen (Kontrollvariable bzw. Kovariate) auf die AV zu berücksichtigen. Die Kontrollvariable (KV) muss intervallskaliert sein und wird zusätzlich zur AV an jedem Probanden miterhoben.

Geprüft wird also, wie stark der Einfluss der KV auf die Untersuchung ist. Dabei wird der Einfluss auf die AV neutralisiert.

Die ANCOVA ist eine Varianzanalyse über Regressionsresiduen, die aus der Regression der AV auf die KV resultieren. Die Frage ist also, ob die UV Mittelwertsunterschiede in den Daten erklärt, wenn der (lineare) Einfluss der KV herausgenommen wurde.

(Wir kombinieren also varianzanalytische mit regressionsanalytischen Techniken!)

Eine ANCOVA rechnet man gerne, wenn man (spätestens bei a-priori Tests herausgefundene) vorgegebene Unterschiede nicht berücksichtigen will (z.B. Intelligenz bei Lernmethoden).

ACHTUNG:

Da die Formeln bei der Kovarianzanalyse gelegentlich sehr kompliziert werden, lasse ich meist beim Summenzeichen die Ober- und Untergrenzen weg. **Die gibt es aber trotzdem noch!!!**

Einfaktorielle Kovarianzanalyse:

Dies ist die einfachste Version aller ANCOVAs, sie wird ausführlich erklärt, dafür werden die anderen Modelle später nur kurz angeschnitten.

Für eine ANCOVA muss von jeder Vp ein Messwert der abhängigen Variablen Y sowie der Kontrollvariablen X vorliegen.

(ACHTUNG: die KV ist X, die AV Y!!!)

Die Summer der x-Werte unter einer Faktorstufe werden mit $A_{x(i)}$ und die Summer der y-Werte unter einer Faktorstufe mit $A_{y(i)}$ gekennzeichnet.

Datenschema:

Faktor A							
a_1		...	a_i		...	a_p	
x_{11}	y_{11}	...	x_{1i}	y_{1i}	...	x_{1p}	y_{1p}
x_{21}	y_{21}	...	x_{2i}	y_{2i}	...	x_{2p}	y_{2p}
\vdots	\vdots		\vdots	\vdots		\vdots	\vdots
x_{m1}	y_{m1}	...	x_{mi}	y_{mi}	...	x_{mp}	y_{mp}
\vdots	\vdots		\vdots	\vdots		\vdots	\vdots
x_{n_11}	y_{n_11}		$x_{n_i i}$	$y_{n_i i}$		$x_{n_p p}$	$y_{n_p p}$

$$\bar{A}_{x(i)} = \frac{1}{n_i} \sum_{m=1}^{n_i} x_{mi}, \quad \bar{A}_{y(i)} = \frac{1}{n_i} \sum_{m=1}^{n_i} y_{mi}$$

$$\bar{G}_x = \frac{1}{N} \sum_{m=1}^{n_i} \sum_{i=1}^p x_{mi}, \quad \bar{G}_y = \frac{1}{N} \sum_{m=1}^{n_i} \sum_{i=1}^p y_{mi}$$

Als Vortest wird gerne eine einfaktorielle Varianzanalyse (mit den y-Werten, also den Werten der AV) gerechnet. Die dann gerechnete ANCOVA kann bei einer signifikanten ANOVA gerade nicht signifikant sein und umgekehrt.

Hypothesen:

Bei der ANCOVA werden adjustierte Mittelwerte der AV getestet, die zu erwarten wären, wenn die KV keinen Einfluss hätte.

$$H_0: \mu_1^{adj} = \dots = \mu_i^{adj} = \dots = \mu_p^{adj}$$

$$H_1: \mu_i^{adj} \neq \mu_{i'}^{adj}$$

Quadratsummen:

Wie bei der einfaktoriellen ANOVA wird von folgender Quadratsummenzerlegung ausgegangen:

$$QS_{total}^* = QS_A^* + QS_{Fehler}^*$$

Da diese Quadratsummen über Residualwerte bestimmt werden, werden sie mit einem * versehen (indiziert).

Totale Quadratsumme:

Die QS_{tot}^* ist die Unterschiedlichkeit aller y-Werte, die nicht durch die Unterschiedlichkeit der x-Werte erklärt werden kann. Dies ist die Summe der quadrierten Residuen der Regression von y auf x. [Also jetzt nicht wundern, dass eine Regressionsgleichung folgt. Bei der ANCOVA wollen wir *Zusammenhänge* (d.h. z.B. Korrelationen, Regressionen etc.) herausfinden.]

Bestimmung von b^{tot} und a^{tot} in der Totalen Regressionsgleichung:

$$b^{tot} = \frac{\sum_m \sum_i (x_{mi} - \bar{G}_x) \cdot (y_{mi} - \bar{G}_y)}{\sum_m \sum_i (x_{mi} - \bar{G}_x)^2} \quad a^{tot} = \bar{G}_y - b^{tot} \cdot \bar{G}_x$$

Vorhersage der AV durch die KV:

$$\hat{y}_{mi}^{tot} = b_{yx}^{tot} \cdot x_{mi} + a_{yx}^{tot}$$

Berechnung der Residuen:

Für jede Vp berechnet man das Residuum.

$$y_{mi}^{*tot} = y_{mi} - \hat{y}_{mi}^{tot}$$

Diese resultierenden Residuen stellen die Werte der AV dar, die nicht von der KV beeinträchtigt werden.

Bestimmung der QS_{tot}^* :

$$QS_{tot}^* = \sum_m \sum_i y_{mi}^{*tot}{}^2$$

Nun kann man QS_{tot}^* mit QS_{tot} unserer einfaktoriellen ANOVA in Verbindung bringen und ausrechnen, wie stark die Kovariate das Ergebnis beeinflusst (verzerrt) hat.

Es ergibt sich eine meist etwas kleinere QS_{tot}^* (nie größer), als bei normaler ANOVA, da die KV ja rausgerechnet wurde [QS sind immer positiv. Die positive KV wurde abgezogen. => Klar?]

Ein Beispiel:

$$QS_{tot} = 50,4 \quad QS_{tot}^* = 46,45$$

$$50,4 - 46,45 = 3,95$$

$$(3,95 : 50,4) \times 100\% = 7,8\%$$

Die KV ist hier für 7,8% der Gesamtunterschiedlichkeit verantwortlich.

Wären QS_{tot}^* und QS_{tot} identisch (also kein Einfluss der KV), könnte man auch einfach eine ANOVA rechnen.

Fehlerquadratsumme:

Bei der QS_{Fehler}^* wird jeweils die Unterschiedlichkeit **innerhalb** der Gruppen betrachtet, die nicht mit der KV erklärt werden kann. D.h. im Gegensatz zur ANOVA werden die Unterschiede aller Messwertpaare pro Gruppe (=Faktorstufe A_i) berechnet und nicht die Unterschiede aller Messwerte vom Gesamtmittel.

Da wir bei der ANCOVA (anfangs) mit Regressionen arbeiten [Letztendlich erfolgt eine Varianzanalyse mit den Residuen, also ab den Quadratsummen.], ermitteln wir erst einmal b^{in} und a^{in} , was ein Teil der Fehlerregressionsgleichung darstellt.

Bei der ANCOVA wird angenommen, dass dieser Einfluss linear und in allen Gruppen gleich ist. Es werden b^{in} und jeweilige a^{in} der Faktorstufe bestimmt (das jeweilige a^{in} gleicht Mittelwertsunterschiede in den einzelnen Faktorstufen aus):

$$b^{in} = \frac{\sum_m \sum_i (x_{mi} - \bar{A}_{x(i)}) \cdot (y_{mi} - \bar{A}_{y(i)})}{\sum_m \sum_i (x_{mi} - \bar{A}_{x(i)})^2} \quad a_i^{in} = \bar{A}_{y(i)} - b^{in} \cdot \bar{A}_{x(i)}$$

Bestimmung der vorhergesagten Werte je Gruppe (wie bei QS_{tot}):

$$\hat{y}_{mi}^{in} = b^{in} \cdot x_{mi} + a_i^{in}$$

Berechnung der Residuen (wie bei QS_{tot}):

$$y_{mi}^{*in} = y_{mi} - \hat{y}_{mi}^{in}$$

Bestimmung der QS_{Fehler}^* (wie bei QS_{tot}):

$$QS_{Fehler}^* = \sum_m \sum_i y_{mi}^{*in^2}$$

Es ergibt sich eine meist etwas kleinere QS_{Fehler}^* (nie größer; obgleich – anders als bei QS_{tot}^* – rein theoretisch möglich), als bei normaler ANOVA, da die KV ja rausgerechnet wurde.

Treatmentquadratsumme:

Die QS_A^* kann nur indirekt bestimmt werden:

$$QS_A^* = QS_{tot}^* - QS_{Fehler}^*$$

Diese Quadratsumme muss nicht mit der Quadratsumme zwischen den geschätzten adjustierten Mitteln übereinstimmen: $\bar{A}_{y(i)}^{adj} = \bar{A}_{y(i)} - b^{in} \cdot (\bar{A}_{x(i)} - \bar{G}_x)$ (= nicht so wichtig)

Freiheitsgrade und Varianzschätzungen:

Quelle		
Treatment	$\hat{\sigma}_A^{*2} = \frac{QS_A^*}{df_A^*}$	$df_A^* = p - 1$
Fehler	$\hat{\sigma}_{Fehler}^{*2} = \frac{QS_{Fehler}^*}{df_{Fehler}^*}$	$df_{Fehler}^* = N - p - 1$

Unter der H_0 gilt: $E(\hat{\sigma}_A^{*2}) = E(\hat{\sigma}_{Fehler}^{*2}) = \sigma_\varepsilon^{*2}$.

Hypothesenprüfung:

Annahme der H_1 , falls

$$F = \frac{\sigma_A^{*2}}{\sigma_{Fehler}^{*2}} > F_{(df_A^*, df_{Fehler}^*, 1-\alpha)}$$

Unterschiedliche Stichprobengrößen, A-priori Einzelvergleiche:

Unterschiedliche Stichprobengrößen sind möglich, wenn man einzelne Formeln modifiziert (siehe Bortz S. 356).

A-priori Einzelvergleiche basieren bei der ANCOVA auf bereinigten AVs, also AVs mit herauspartialisierendem Einfluss der KV (siehe Bortz S. 356).

Effektgrößen:

Das Tolle bei der ANCOVA waren ja die möglichst kleinen Stichprobengrößen. Die kann man mit Effektgrößen ermitteln.

Die Formeln zur Effektgröße bei der ANCOVA sind mit denen bei der ANOVA identisch! (siehe also „Einfaktorielle Versuchspläne: Effektgrößen“)

Man beachte nur, dass die Formeln

$$d = \frac{\mu_{\max} - \mu_{\min}}{\sigma} \quad \text{und} \quad \varepsilon = \sqrt{\frac{\eta^2}{1 - \eta^2}}$$

sich auf die korrigierten Mittelwerte beziehen. Also $\bar{A}_{y(i)}^* = \bar{A}_{y(i)} - b^{in} \cdot (\bar{A}_{x(i)} - \bar{G}_x)$

Zusammenfassung:

Die ANCOVA ist also eine Methode, die varianzanalytische mit regressionsanalytischen Methoden kombiniert.

Man berechnet zuerst von der KV aus eine Regression auf die AV und kann so den Teil der AV, der mit der KV zusammenhängt, herausfinden. Mit dem Rest (den Residuen) der AV rechnet man dann eine einfache Varianzanalyse.

Die ANCOVA kann sowohl ein vorher nicht signifikantes Ergebnis signifikant machen, wie auch umgekehrt.

Letzteres kann passieren, wenn die KV mit der UV korreliert, wenn z.B. Wirkung von Schichtunterschieden (UV) auf Ausgaben für Kindererziehung (AV) gemessen wird & das Einkommen der Eltern (KV) rausgenommen wird (wäre da zwar blödsinnig, aber theoretisch machbar).

Voraussetzungen:

- Normalverteilte Fehlerkomponenten.
 - Homogene Fehlervarianzen.
 - Unabhängige Fehlerkomponenten.
 - Homogene Steigungen der Regressionen innerhalb der Gruppen.
- } genauso wie bei der ANOVA

Die ANCOVA erweist sich als mindestens ebenso robust wie die ANOVA, im letzten Punkt ist sie sogar sehr robust [Es müssen schon die Regressionen (stark) heterogen sein, die Stichproben ungleich groß und die Residuen nicht normalverteilt sein!].

Mehrfaktorielle Kovarianzanalyse:

Die mehrfaktorielle Kovarianzanalyse ist wiederum eine simple Erweiterung der einfaktoriellen Kovarianzanalyse, nur ist sie komplizierter zu berechnen.

Denn die KV muss aus allen Treatmentquadratsummen, ergo QS_A , QS_B & $QS_{A \times B}$ herausgerechnet werden. Das führt zu jeder Menge Formeln (siehe Bortz auf S. 360-363), da nur QS_{tot} und QS_{Fehler} direkt zu bestimmen sind.. Ansonsten ist alles wie bei der einfaktoriellen ANCOVA.

Kovarianzanalyse mit Messwiederholung:

Eine einfaktorielle ANCOVA mit Messwiederholung ist wieder nur eine einfache Erweiterung der einfaktoriellen ANCOVA, es wird halt nur komplizierter. Man arbeitet halt mit QS_{Res} statt QS_{Fehler} , und muss einige Formeln anwenden (Bortz S. 363-372).

Eine mehrfaktorielle Kovarianzanalyse mit Messwiederholung bedeutet jede Menge Formeln, man verliert die Übersicht, und ist wohl auch nicht notwendig zu behandeln.

Unvollständige, mehrfaktorielle Versuchspläne:

Gelegentlich kann es in der (u.a. psychologischen) Forschung dazu kommen, dass man nicht alle Möglichkeiten (Faktorstufenkombinationen) einer Varianzanalyse erfassen möchte – da es z.B. zuviel Aufwand wäre. Deswegen benötigt man Verfahren, die solches ermöglichen.

Hierbei gehen nur die Haupteffekte und Interaktionen in die Analyse ein, die wir benötigen.

Versuchspläne, bei denen nicht alle möglichen Faktorstufenkombinationen untersucht werden, bezeichnet man als unvollständige Versuchspläne.

Es kann gelegentlich auch methodisch unmöglich sein, einen vollständigen Versuchsplan aufzustellen, z.B. der Vergleich von Therapieerfolg mit den Faktorstufen Behandlungsmethode A und Therapeut B. Da es sich im Allgemeinen um spezialisierte Therapeuten handelt, können wir nicht erwarten, dass ein auf Verhaltenstherapie spezialisierter Therapeut plötzlich auch Psychoanalyse macht.

Die nun folgenden Verfahren werden nur äußerst kurz behandelt, da sie meines Wissens in keiner Veranstaltung des Bereiches Methodenlehre vorkamen und auch auf keinen Folien oder Skripten zu finden sind.

Wer es trotzdem wissen will, lese Kapitel 11 im Bortz (S. 375-396).

Hierarchische und teilhierarchische Versuchspläne:

Tritt jede Faktorstufe eines Faktors (z.B. A) nur mit einer bestimmten Faktorstufe eines anderen Faktors (z.B. B) auf, haben wir eine Verschachtelung. Die Stufen des einen Faktors (hier B) werden in die Stufen des Anderen (hier A) hineingeschachtelt (siehe Tabelle).

Versuchspläne, bei denen durch die Schachtelung des einen Faktors unter den anderen Faktor eine Hierarchie der Faktoren entsteht, bezeichnet man als (zweifaktorielle) hierarchische Versuchspläne.

Hierarchischer Versuchsplan:

(S steht jetzt für die Summe der n , ich hatte keinen Bock, die auch noch da rein zu schreiben)

a ₁		a ₂		a ₃	
b ₁	b ₂	b ₃	b ₄	b ₅	b ₆
S ₁₁₍₁₎	S ₁₂₍₁₎	S ₂₁₍₂₎	S ₂₂₍₂₎	S ₃₁₍₃₎	S ₃₂₍₃₎

Beispiel eines **vollständigen Versuchsplanes**, wobei die gefüllten Kästen **dieselben** wie die in der letzten Tabelle darstellen (ja, der Index von S ist anders):

	a ₁	a ₂	a ₃
b ₁	S ₁₁		
b ₂	S ₁₂		
b ₃		S ₂₃	
b ₄		S ₂₄	
b ₅			S ₃₅
b ₆			S ₃₆

Teilhierarchische Versuchspläne gibt es erst ab drei oder mehr Faktorstufen, der Unterschied besteht darin, dass nicht alle Faktorstufen unter den Stufen eines Faktors geschachtelt sind, z.B. wäre C nicht unter A geschachtelt. Das sähe beispielsweise so aus:

	a ₁		a ₂		a ₃	
c ₁	b ₁	b ₂	B ₃	b ₄	b ₅	b ₆
c ₂	S ₁₁₍₁₎	S ₁₂₍₁₎	S ₂₁₍₂₎	S ₂₂₍₂₎	S ₃₁₍₃₎	S ₃₂₍₃₎

Das Ganze bei den hierarchischen und teilhierarchischen Verfahren läuft wieder über Quadratsummen und Prüfvarianzen ab – natürlich mit modifizierten Formeln, ansonsten aber wie immer.

Man kann bei hierarchischen oder teilhierarchischen Verfahren keine Interaktionen zwischen den geschachtelten Faktoren testen!

Lateinische Quadrate:

Lateinische Quadrate führt man durch, um Haupteffekte zu überprüfen. Man kann sie aber nur durchführen, sofern keine Interaktionen vorliegen.

Mit lateinischen Quadraten können 3 Haupteffekte geprüft werden.

Beispiel eines lateinische Quadrates für p = 3 (p = Anzahl der Faktorstufen):

	a ₁	a ₂	a ₃
b ₁	c ₁	c ₂	c ₃
b ₂	c ₂	c ₃	c ₁
b ₃	c ₃	c ₁	c ₂

Man beachte hierbei, dass das c₃ von b₂ und a₂ eigentlich neben sich noch ein c₁ und ein c₂ hätte. Die fallen halt nur raus. (Nicht kapiert? Tabelle zur Erläuterung => Bortz S.385)

Der Rest – QS, Prüfvarianz – läuft wie immer.

Griechisch-lateinische Quadrate:

Das ist dasselbe wie ein lateinisches Quadrat, nur können vier Haupteffekte geprüft werden, sofern keine Interaktion vorliegt (Bortz S. 388).

Quadratische Anordnungen mit Messwiederholungen:

Natürlich kann man auch Messwiederholungen bei den Quadraten durchführen. Dabei ist zu beachten, dass die einzelnen Faktorstufenkombinationen voneinander unabhängig sind.

Wichtig ist auch noch, wie man die Stufen anordnet. (Bortz S. 392-396)

Theoretische Grundlagen der Varianzanalyse:

In den bisherigen Kapiteln im varianzanalytischen Teil dieses Skriptes waren die Verfahren möglichst logisch im Zusammenhang ihrer Anwendungsmöglichkeiten dargestellt worden. Die theoretische Herleitung der Verfahren und ihrer einzelnen Untertests, Vorgaben, etc. soll nun behandelt werden.

Die nun folgenden Verfahren werden – ebenso wie das letzte Kapitel – nur äußerst kurz behandelt. Alles, was ich anschneide, ist auch in den B3-Folien behandelt worden. Ich glaube nicht, dass dieses Kapitel für die Prüfung relevant ist (aber Sicherheit geht vor). Wer trotzdem noch mehr wissen will, lese Kapitel 12 im Bortz (S. 397-422).

Einfaktorielle Varianzanalyse:

In der ANOVA wird angenommen, dass sich der Messwert des Probanden m aus der Gruppe i gemäß folgendem **linearen Modell** zusammensetzt:

$$x_{im} = \mu + \alpha_i + \varepsilon_{im}.$$

Hierbei bezeichnen die drei Größen:

- μ Gesamtmittel aller untersuchten Populationen.
Dieser Parameter ist konstant für alle Messwerte.
- α_i = $\mu_i - \mu$
Effekt der Bedingung i , der jeweils für alle Objekte einer Population i konstant ist.
- ε_{im} = $x_{im} - \mu_i$
Individuelle Fehlerkomponente bzw. Unterschiede, die nicht durch das Treatment erklärbar sind.

Die Abweichung des Messwertes x_{mi} vom Gesamtmittel μ geht zum einen auf den Effekt der UV (α_i) und zum anderen auf den Einfluss anderer unbekannter Faktoren (ε_{im}) zurück.

Varianzschätzungen:

Bei Gültigkeit der H_0 sind sowohl $\hat{\sigma}_A^2$ als auch $\hat{\sigma}_{Fehler}^2$ erwartungstreue Schätzer der gemeinsamen Populationsvarianz σ_ε^2 .

Wenn H_0 gilt, dann:

$$E(\hat{\sigma}_A^2) = E(\hat{\sigma}_{Fehler}^2) = \sigma_\varepsilon^2$$

Die Fehlervarianz ist für Faktoren mit zufälligen und festen Effekten die beste Prüfvarianz für die Treatmentvarianz.

Mehrfaktorielle Varianzanalyse:

Bei der zweifaktoriellen Varianzanalyse wird angenommen, dass sich der Messwert des Probanden m aus der Bedingung ij gemäß folgendem *linearen Modell* zusammensetzt:

$$x_{ijm} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijm}.$$

μ	Gesamtmittel aller untersuchten Populationen. Konstante für alle Messwerte.
α_i	$= \mu_i - \mu$ Effekt der Bedingung i des Faktors A . Konstante für alle Objekte einer Population i .
β_j	$= \mu_j - \mu$ Effekt der Bedingung j des Faktors B . Konstante für alle Objekte einer Population j .
$\alpha\beta_{ij}$	$= (\mu_{ij} - \mu) - (\mu_i - \mu) - (\mu_j - \mu)$ $= \mu_{ij} - \mu_i - \mu_j + \mu$ Interaktionseffekt, der durch das Zusammenwirken von Stufe i und Stufe j entsteht.
ε_{ijm}	$= x_{ijm} - \mu_{ij}$ Individuelle Fehlerkomponente bzw. Unterschiede, die nicht durch das Treatment erklärbar sind.

Die Punkte bei den Mys kennzeichnen einen fehlenden (hier redundanten) Index.

Varianzanalyse mit Messwiederholung:

Bei der **einfaktoriellen Varianzanalyse mit Messwiederholung** wird angenommen, dass sich der Messwert des Probanden m aus der Bedingung i gemäß folgendem *linearen Modell* zusammensetzt:

$$x_{im} = \mu + \alpha_i + \pi_m + \alpha\pi_{im} + \varepsilon_{im}.$$

μ	Gesamtmittel aller untersuchten Populationen. Konstante für alle Messwerte.
α_i	$= \mu_i - \mu$ Effekt der Stufe i des Faktors A .
π_m	$= \mu_{.m} - \mu$ Spezifische Reaktionsweise der Person m . Konstant über alle Stufen des Faktors A .
$\alpha\pi_{im}$	Interaktionseffekt: Spezifische Reaktionsweise der Person m auf die Faktorstufe i .
ε_{im}	Individuelle Fehlerkomponente.

$\alpha\pi_{im}$ und ε_{im} können nicht getrennt voneinander geschätzt werden und werden zu einer Residualkomponente Res_{im} zusammengefasst:

$$x_{mi} = \mu + \alpha_i + \pi_m + \text{Res}_{im} .$$

Wann resultieren eigentlich abhängige Stichproben?

Abhängige Stichproben resultieren, wenn ...

- ... eine Stichprobe von n Personen sukzessive allen p Untersuchungsbedingungen ausgesetzt wird,
- ... an n Personen zu p Zeitpunkten Messwerte erhoben werden,
- ... oder wenn p Gruppen von je n Personen hinsichtlich einer Matchingvariable parallelisiert werden.

Kovarianzanalyse:

Im Vergleich zur einfaktoriellen ANOVA erweitert die **einfaktorielle Kovarianzanalyse** ihr Datenmodell um $\beta_{yx} \cdot (x_{mi} - \mu_x)$.

$$y_{mi} = \mu_y + \alpha_i + \beta_{yx} \cdot (x_{mi} - \mu_x) + \varepsilon_{mi}$$

μ_y	Gesamtmittel der AV.
α_i	Vom Einfluss der Kontrollvariablen bereinigter Effekt der Stufe i des Faktors A.
$\beta_{yx} \cdot (x_{mi} - \mu_x)$	Abhängigkeit der AV von der KV. Zwischen Y und X wird nur ein linearer Zusammenhang angenommen, der über alle Stufen der UV gleich ist.
ε_{mi}	Individuelle Fehlerkomponente.

Weitere Verfahren:

Weitere Verfahren wären im Kapitel 11 des Bortz – “Unvollständige, mehrfaktorielle Versuchspläne” – zu finden. Da ich die aber sowieso nur kurz behandelt habe, lasse ich sie hier weg. (Nachzulesen im Bortz S.414 - 416)

Allgemeine Regeln:

Im Bortz auf S. 416 bis 422 stehen allgemeine Regeln zur Erstellung von Varianzanalysen, mit allem, was das Methodikerherz begehrt: Datenmodelle, Erwartungswerte, Rechenregeln, etc. Das Tolle daran ist, dass man so jede mögliche Verfahrenskombination selbst erstellen kann. **Noch besser hingegen ist, dass uns das für die Prüfung völlig wurscht sein kann.**