

Julius-Maximilians-Universität Würzburg

Methoden der Unterschiedsprüfung

Mitschrift der Vorlesung von Dr. Rainer Scheuchenpflug
im SS 2008

Roland Pfister



Inhaltsverzeichnis

0. Vorwort	4
1. ANOVA I: Einfaktorielle ANOVA	5
1.1. Grundgedanke und Wiederholung aus QMB	5
1.1.1. Geplante Kontraste.....	6
1.1.2. Post-hoc Tests.....	6
1.2. Voraussetzungen	7
1.2.1. Unabhängigkeit der Beobachtungen	7
1.2.2. Mindestens Intervallniveau	7
1.2.3. Normalverteilung und Varianzhomogenität.....	7
1.2.4. Praktische Hinweise	9
1.3. Zusammenfassung	9
1.4. Übung I: Einfaktorielle ANOVA	10
1.4.1. Excel-Grundlagen.....	10
1.4.2. Einfaktorielle ANOVA mit SPSS	10
1.4.3. Nichtparametrische Tests	11
1.4.4. Anmerkung	11
2. ANOVA II: Mehrfaktorielle ANOVA	12
2.1. Interaktionen.....	12
2.2. ANOVA-Tabelle.....	13
2.3. Anmerkungen	13
2.3.1. Geplante Kontraste und Post-Hoc-Tests	13
2.3.2. Power (Teststärke)	13
2.3.3. Wie bekommt man signifikante Ergebnisse?	14
2.3.4. Ungleiche Stichprobenumfänge.....	14
2.4. Übung II: Mehrfaktorielle ANOVA	16
2.4.1. Excel.....	16
2.4.2. SPSS	16
2.4.3. Kontraste und Post-hoc-Tests	17
2.4.4. Anmerkung: Mehrfaktorielle ANOVA und Multiples Testen	18
3. ANOVA III: Messwiederholung	19
3.1. Einfaktorielle ANOVA mit Messwiederholung	19
3.1.1. Modell	19
3.1.2. Ablauf	20
3.1.3. ANOVA-Tabelle(n).....	20
3.1.4. Nonparametrisch: Friedman-Varianzanalyse.....	21

3.1.5.	Kontraste und Post-Hoc-Tests.....	21
3.1.6.	Voraussetzungen.....	22
3.1.7.	Missing Data.....	23
3.2. Verlaufskurven (Split-Plot)	24	
3.2.1.	Modell	24
3.2.2.	ANOVA-Tabelle	25
3.2.3.	Anmerkungen	25
3.3. Profilanalyse (MANOVA bei Split-Plot).....	26	
3.3.1.	Voraussetzungen.....	26
3.3.2.	Beispiel	26
3.4. Ungleiche Zellbesetzungen und Typ I-VI-Summen	27	
3.4.1.	Typ 1 – Summen: Sequentiell.....	27
3.4.2.	Typ 2 – Summen: Teilverquentiell	27
3.4.3.	Typ 3 – Summen: Standard.....	28
3.4.4.	Typ 4 – Summen: Nicht empfohlen	28
3.4.5.	Typ 5 – Summen: Aus der Industrie	28
3.4.6.	Typ 6 – Summen: Effective Hypothesis	28
3.5. Übung III: ANOVA mit Messwiederholung	29	
3.5.1.	Einfaktorielle ANOVA mit Messwiederholung	29
3.5.2.	Kontraste	32
3.5.3.	Split-Plot-Analyse	33
4. Multiple Regression.....	35	
4.1. Modell.....	35	
4.1.1.	Roh- und Standardwert-Regressionsgleichung	35
4.1.2.	Interpretation von Koeffizienten	35
4.1.3.	Beurteilung der Schätzung: R^2	36
4.2. Prüfung der Schätzung.....	36	
4.2.1.	Test der Koeffizienten.....	37
4.2.2.	SPSS-Outputs	37
4.3. Voraussetzungen	38	
4.3.1.	Prüfung der Voraussetzungen: Residuenplots	38
4.3.2.	Optimaler Variablenatz	39
4.4. Ausblick	40	
4.5. Übung IV: Multiple Regression	41	
4.5.1.	Excel.....	41
4.5.2.	SPSS-Einstellungen	41
4.5.3.	Nichtlineare Modelle	43
5. ANCOVA, MANOVA und MANCOVA	44	
5.1. Kovarianzanalyse (ANCOVA).....	44	
5.1.1.	Annahmen	44
5.1.2.	Weitere Anwendungen	44

5.2. Multivariate Varianzanalyse (MANOVA)	44
5.2.1. Interpretation der Ergebnisse	45
5.2.2. Bewertung	45
5.3. Multivariate Kovarianzanalyse (MANCOVA)	45
5.4. Profilanalyse: MANOVA bei Split-Plot.....	46
5.5. Übung V: ANCOVA und MANOVA	47
5.5.1. ANCOVA	47
5.5.2. MANOVA.....	47
6. Diskriminanzanalyse	48
6.1. Grundidee	48
6.2. Modell.....	49
6.3. Anwendung und SPSS	49
6.3.1. Klassifikation.....	49
6.3.2. Klassifikationsfunktionen	50
6.4. Voraussetzungen und Ausblick.....	51

0. Vorwort

Dozent: Dr. Rainer Scheuchenpflug

Termin: Mi

Klausur: - Termin: ???

Web: www.izvw.de bzw. www.psychologie.uni-wuerzburg.de/methoden

1. ANOVA I: Einfaktorielle ANOVA

1.1. Grundgedanke und Wiederholung aus QMB

Zunächst werden **univariate, einfaktorielle ANOVAs im Between-subject-Design** besprochen. Während in QMB vor allem die Berechnung einer Varianzanalyse über die Winer-Hilfsgrößen besprochen wurde, soll hier die „eigentliche“ Berechnung genauer betrachtet werden.

Diese beruht auf dem Grundgedanken, die Streuung auf Basis der Messwerte (Fehlervarianz) mit derjenigen zu vergleichen, die durch die unterschiedliche Stichprobenzugehörigkeit erzeugt wird (Schätzung der Varianz auf Basis der einzelnen Stichprobenmittelwerte; k bzw. p = Anzahl der Stichproben).

Dieses Vorgehen führt zum **F-Bruch** der Varianzanalyse:

$$F = \frac{PV}{FV} > 1$$

Die **Primärvarianz** ergibt sich aus der Streuung, welche durch die Mittelwerte der einzelnen Stichproben geschätzt wird:

$$\hat{\sigma}^2 = n \hat{\sigma}_M^2 = n \frac{\sum_{i=1}^p (\bar{x}_i - \bar{G})^2}{p-1} \quad \text{mit } \bar{G} = \text{Grand Mean},$$

während die **Fehlervarianz** durch die Varianz innerhalb der verwendeten Stichproben geschätzt wird:

$$\hat{\sigma}^2 = \frac{1}{p} \sum_{i=1}^p \hat{\sigma}_i^2 = \frac{1}{p} \sum_{i=1}^p \frac{\sum_{m=1}^n (x_{mi} - \bar{x}_i)^2}{n-1}.$$

Sollten die Stichproben aus der gleichen Grundgesamtheit stammen, so unterscheiden sich PV und FV nur unsystematisch voneinander. Sollten sie jedoch aus unterschiedlichen Grundgesamtheiten stammen (genauer: Grundgesamtheiten mit unterschiedlichen Mittelwerten aber etwa vergleichbaren Streuungen), so sollte die PV deutlich größer sein als die FV.

Die Ergebnisse der ANOVA werden häufig in einer **ANOVA-Tabelle** dargestellt. Allerdings interessieren hier eigentlich nur der F-Wert der jeweiligen Zeile, sowie dessen entsprechender p-Wert. Die univariate, einfaktorielle ANOVA liefert also nur einen F-Wert und dessen p-Wert.

Im **Fließtext** eines Artikels wird ein ANOVA-Kennwert folgendermaßen berichtet: „Der Effekt des Faktors XY war signifikant ($F(df_{\text{Nenner}}, df_{\text{Zähler}}) = 4.83, p = .003$).“ Dabei werden statistische Symbole kursiv geschrieben und inferenzstatistische Kennwerte (z.B. F-Werte) mit 2 Nachkommastellen angegeben. Zahlen die nicht größer als 1 werden können, werden ohne führende Null angegeben.

Allerdings sagt ein Omnibus-Test wie die ANOVA nichts darüber aus, welche Mittelwertsdifferenzen bedeutsam sind. Hier kommen **geplante Kontraste** oder **Post-hoc-Tests** (a posteriori-Tests) ins Spiel. Diese haben eine höhere Teststärke als α -adjustierte t-Tests und sind für die zu prüfenden Hypothesen evtl. angemessener (Gruppe 1 und 2 gegen Gruppe 3), wenn keine Originaldaten verfügbar sind.

1.1.1. Geplante Kontraste

Kontraste müssen schon vor der Durchführung der ANOVA (bzw. genauer: vor der Durchführung der gesamten Untersuchung) bekannt sein. Ansonsten würde man mit zu vielen Freiheitsgraden testen ([Schrotschussstatistik](#)).

Kontraste werden durch [Koeffizientenvektoren](#) definiert und über eine Linear-kombination dieser Vektoren geprüft. Ein Vektor besteht aus so vielen Werten wie Faktorstufen (Gruppen) vorhanden sind und wird nach festen Regeln aufgestellt (hier für gleiche Stichprobenumfänge der Subgruppen):

- Soll eine Gruppe aus dem Vergleich ausgeschlossen werden, so erhält sie den Koeffizienten 0.
- Sollen mehrere Gruppen für den Vergleich zusammengefasst werden, erhalten sie denselben Koeffizienten.
- Die Summe der Koeffizienten muss 0 ergeben.

Sollen die ersten beiden Gruppen gegen eine dritte Gruppe verglichen werden (insgesamt nur 3 Stpn) könnte ein möglicher Kontrastvektor folgendermaßen aussehen:

$$(1 | 1 | -2) \quad \text{oder allgemein:} \quad (c_1 | c_2 | c_3)$$

Diese Vektoren können schließlich über eine [t](#)-verteilte Größe (df: N-k) oder über eine [F](#)-verteilte Größe (df₁ = 1, df₂ = N-k) geprüft werden.

Sollen die Kontraste teststark bleiben, so müssen sie unabhängig (orthogonal) sein, da sonst auf das gleiche Datenmaterial zurückgegriffen wird. Eine Prüfung auf [Orthogonalität](#) ist auf Basis der Koeffizientenvektoren möglich, da zwei Kontraste genau dann orthogonal sind, wenn das Skalarprodukt der entsprechenden Vektoren 0 ergibt¹:

$$\sum_{i=1}^p c_{1i} c_{2i} = 0$$

Insgesamt sind dabei immer [k-1 orthogonale Kontraste möglich](#). Dies entspricht der Anzahl der Freiheitsgrade der Treatmentquadratsumme. Sollen mehr Kontraste geprüft werden, ist eigentlich auch hier eine α -Adjustierung notwendig.

1.1.2. Post-hoc Tests

Ergeben sich bestimmte Hypothesen erst nach dem Blick auf die Ergebnisse (z.B. PPE), dann muss ein Post-hoc-Test durchgeführt werden, der in Rechnung stellt, dass implizit alle möglichen Einzelvergleiche durchgeführt wurden. Hierdurch wird der Test selbstverständlich konservativ.

Post-hoc-Tests spezifizieren dabei normalerweise eine kritische Differenz, deren Überschreitung für einen signifikanten Unterschied spricht.

- Fisher's LSD (least significant difference)
 - Duncan's Multiple Range Test
 - Neuman-Keuls
 - Tukey
 - Scheffé
- 
- liberal
konservativ

¹ Orthogonalitätsprüfung in Excel: Vektoren untereinander schreiben, =PRODUKT in den Spalten und =SUMME über die Produkte anwenden; fertig.

1.2. Voraussetzungen

Die (Between-)ANOVA macht folgende Annahmen:

- Unabhängigkeit der Beobachtungen
- Normalverteilung
- Varianzhomogenität
- Mindestens Intervallniveau

Wirklich kritisch sind vor allem die erste und die letzte Annahme, da die ANOVA sehr empfindlich auf eine Verletzung dieser beiden reagiert.

1.2.1. Unabhängigkeit der Beobachtungen

Die Unabhängigkeit der Beobachtungen kann man zwar bei entsprechender methodischer Umsicht bei der Versuchsplanung annehmen – ob diese Annahme erfüllt ist lässt sich jedoch nicht statistisch prüfen.

1.2.2. Mindestens Intervallniveau

Mittelwerte und Varianzen müssen sinnvoll interpretierbar sein. Dies ist jedoch ein Problem der entsprechenden Skalierung („The numbers don't know where they come from.“). Statistikprogramme funktionieren demnach nach dem **GI-GO-Prinzip (Garbage In, Gargabe Out)**: Wenn man Mist in die Analyse reinsteckt, kommt auch Mist raus.

1.2.3. Normalverteilung und Varianzhomogenität

Gegen eine Verletzung dieser beiden Voraussetzungen ist das Verfahren relativ robust. Zudem existieren für diese Voraussetzungen Tests und Adjustierungen, die hier genauer betrachtet werden sollen.

Allgemein möchte man bei Tests auf die Gültigkeit von Voraussetzungen die H_0 beibehalten. Um den hier drohenden β -Fehler zu begrenzen wird daher per Konvention $\alpha = 0.20$ gesetzt. Nur wenn der Test dann nicht signifikant wird, darf die ANOVA (streng genommen) verwendet werden.

Dieses Vorgehen hat allerdings zwei Probleme: Generell sind die Tests oft deutlich empfindlicher gegen Verletzungen der Voraussetzungen als die ANOVA selbst und außerdem testet man am gleichen Datenmaterial:

„To make the preliminary test on variances is rather like putting to sea in a rowing boat to find out whether conditions are sufficiently calm for an ocean liner to leave port“ (Box, 1953).

Daher werden nun Möglichkeiten vorgestellt, wie eine ANOVA auch bei schießen Verteilungen oder ungleichen Varianzen eingesetzt werden kann.

1.2.3.1. Möglichkeit 1: Freiheitsgrade adjustieren

Die Adjustierung der Freiheitsgrade führt i.A. zu nicht-ganzzahligen Freiheitsgraden. F-Werte müssen daher aus den Tabellenwerten interpoliert werden (bzw. generell mit SPSS/Excel berechnet werden).

Dieses Vorgehen entspricht der Idee des Welch-Tests.

1.2.3.2. Möglichkeit 2: Datentransformation

Bei schießen Messwertverteilungen werden bestimmte Transformationen empfohlen. Nach der Anwendung einer bestimmten Transformation rechnet man mit den transformierten Daten weiter:

- **Logarithmieren:** Bei linkssteilen Verteilungen wie RTs bzw. wenn multiplikatives Zusammenwirken vorliegen kann.
- **Quadratwurzel:** Bei linkssteilen Verteilungen; wird für Anzahlen empfohlen.
- **Arcusssinus:** Bei Anteilswerten.

1.2.3.3. Möglichkeit 3: Ignorieren wenn n groß genug

Sind die **Stichproben gleich groß und $N > 30$** , dann ist eine Verletzung insb. der Homoskedastizitätsannahme nicht besonders schwerwiegend und kann daher ignoriert werden. Dies gilt bis zu Varianzverhältnissen von bis zu 1:5 (kleinste : größte Varianz).

Eine derartige Verletzung wird insbesondere bei ungleichen Stichprobenumfängen n_i bedeutsam.

1.2.3.4. Möglichkeit 4: Rangwerte

Eine weitere Möglichkeit, mit schießen Verteilungen oder ungleichen Varianzen umzugehen, ist das Verwenden von Rangwerten. Ein parameterfreies Äquivalent zur ANOVA stellt der **Kruskal-Wallis-H-Test** dar.

Bei diesem Test wird eine approximativ X^2 -verteilte Prüfgröße berechnet, die solange gut funktioniert, solange keine **Rangbindungen** vorliegen. Ist dies der Fall bietet sich die Berechnung einer **normalen ANOVA auf Basis der Rangwerte** an.

Generell hat sich gezeigt, dass t-Test und ANOVAs bei Rangdaten zu praktisch denselben Ergebnissen kommen, wie spezielle nonparametrische Tests (obwohl hier mit Sicherheit die Normalverteilungsannahme verletzt ist).

1.2.3.5. Möglichkeit 5: Resampling/Randomisierung

Bei **Resamplingverfahren** oder **Randomisierungstests** werden die erhobenen Daten im Prinzip als Grundgesamtheit aufgefasst. Anhand dieser GG wird geprüft, wie wahrscheinlich es ist, die empirisch gefundenen Mittelwertsunterschiede bei zufälliger Verteilung der Werte auf die Stpn zu erhalten.

Hierzu werden alle möglichen (Randomisierung) oder eine Zufallsauswahl aller möglichen Sortierungen der Messwerte (approximative Randomisierung) aufgestellt und die interessierende Stichprobenstatistik berechnet². Damit erhält man ohne Umwege über Verteilungsannahmen eine empirische Verteilung der Stichprobenstatistik, die als Prüfverteilung dienen kann (**tatsächlich parameterfreie Testung!**). Hier muss eine Verletzung der Varianzhomogenität also keine Bauchschmerzen mehr bereiten.

² Resampling: ZmZ, Randomisierung: ZoZ

Diese Vorgehensweise führt jedoch zu einer Explosion der Anzahl möglicher Kombinationen und ist noch kein etabliertes Standardverfahren. In SPSS ist es daher nur für die einfaktorielle ANOVA integriert

1.2.4. Praktische Hinweise

Während der Versuchsplanung muss auf die Unabhängigkeit der Beobachtungen zwingend geachtet werden. Zudem sind ungleiche Stichprobenumfänge sehr ungünstig, da diese in einem Verlust an Power resultieren und die ANOVA dann wenig robust gegen die Verletzung der Homoskedastizitätsannahme ist. Allerdings stellt sich auch hier die Frage, was genau „ungleich groß“ bedeutet.

1.3. Zusammenfassung

Die einfaktorielle, univariate Varianzanalyse ist ein Verfahren zur Prüfung von Mittelwertsunterschieden bei mehr als 2 Stichproben und beruht auf folgendem Datenmodell³:

$$y_{mi} = \mu + \alpha_i + \varepsilon_{mi} \quad \varepsilon \sim N(0; \sigma) \quad \sum_{i=1}^k \alpha_i = 0$$

Das Verfahren setzt neben unabhängigen Daten und intervallskalierten Messwerten die Normalverteilung der Messwerte in der GG sowie Varianzhomogenität voraus. Solange die Stichproben gleich groß sind ist sie jedoch robust gegen eine Verletzung der Annahmen.

Durch den Omnibus-Test ANOVA wird die eigentlich interessierende Hypothese jedoch meist nicht beantwortet. Hier finden geplante Kontraste oder Post-hoc-Tests ihre Anwendung. Daher wird die Bedeutung der ANOVA zurzeit kritisch diskutiert, etwa dahingehend, ob sie durch die alleinige Berechnung geplanter Kontraste oder auch Randomisierungsverfahren ersetzt werden könnte.

Diese Regelung (Ersetzen der ANOVA durch reine Kontrastanalysen) wurde in die DGPs-Richtlinien zur Manuskriptgestaltung von 2008 aufgenommen.

³ „Wenn man das Datenmodell korrekt hinschreiben will, gehören $\varepsilon \sim N(0; \sigma)$ und $\sum \alpha_i = 0$ auf jeden Fall dazu.“

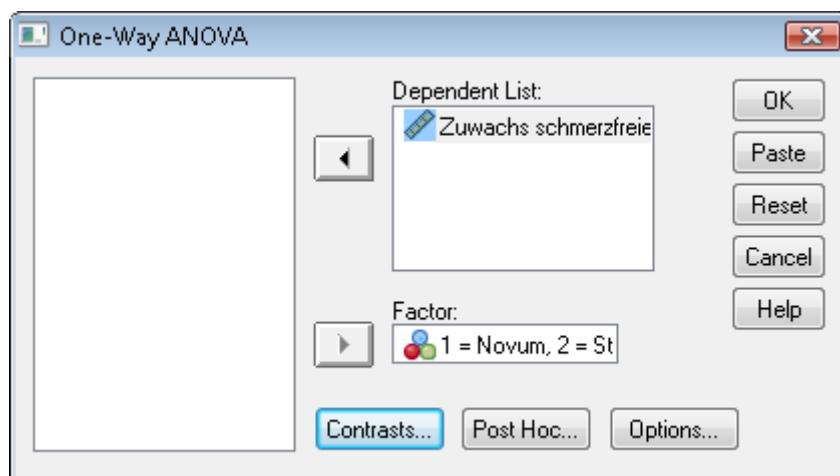
1.4. Übung I: Einfaktorielle ANOVA

1.4.1. Excel-Grundlagen

Finden sich im Skript zur Einführung ins wissenschaftliche Arbeiten (S. 26). Behandelt wurden verschiedene Methoden zur Erzeugung von Zufallszahlen (Standardfunktionen Zufallszahl und Zufallsbereich sowie die entsprechende Analysefunktion) und die Berechnung einer einfaktoriellen ANOVA in Excel.

1.4.2. Einfaktorielle ANOVA mit SPSS

Achtung: In SPSS muss jeder Fall in einer eigenen Zeile stehen. Die Excel-Liste muss also entsprechend umkodiert werden. Die einfaktorielle ANOVA findet sich unter *Analyze\Compare Means*.



Unter *Contrasts* und *Post Hoc* lassen sich die entsprechenden Einzelvergleiche auswählen. Kontraste werden hier sinnigerweise als Koeffizientenvektoren eingegeben. Unter *Options* finden sich unter anderem Mittelwertplot und Test auf Varianzhomogenität.

Im Output⁴ ist schließlich folgendes enthalten:

- Information zur Varianzhomogenität (sollte nicht signifikant sein)
- ANOVA-Tabelle
- Tests der definierten Kontraste
- Post-hoc-Tabelle mit allen möglichen Einzelvergleichen
- **Homogene Untergruppen (Homogeneous Subtests):** Hier wird untersucht, ob sich nachträglich Untergruppen identifizieren lassen, indem Stichproben, die sich nicht signifikant voneinander unterscheiden, als Gruppe aufgefasst werden. Bei einer einzigen dreistufigen UV ist diese Prozedur recht zweckfrei.
- Diagramm: Kann optisch und inhaltlich noch aufgebessert werden (z.B. Nachkommastellen verringern oder Fehlerbalken einfügen (nicht möglich)). Häufig macht es daher Sinn, den Chart Builder zu verwenden, oder ganz auf andere Programme (z.B. SigmaPlot) zurückzugreifen.

⁴ Falls in einer Zelle des Outputs viele Sternchen zu sehen sind (*****), so ist in dieser Zelle eine Zahl enthalten, die aufgrund ihrer Genauigkeit nicht mit der vorgegebenen Anzahl an Nachkommastellen angezeigt werden kann. Hierfür muss man die Zelle einfach etwas breiter machen.

1.4.3. Nichtparametrische Tests

Liegen nach einer Rangtransformation keine Rangbindungen vor, so kann eine Rangvarianzanalyse (Kruskal-Wallis-H) mit simultanen Wilcoxon-Tests berechnet werden. Bei Rangbindungen bietet sich eine normale ANOVA auf den Rangdaten an.

1.4.3.1. Rangvarianzanalyse

Die angemessene Rangvarianzanalyse findet sich unter *Nonparametric Tests \ K Independant Samples*. Die nachgeschalteten Wilcoxon-Tests (äquivalent: Mann-Whitney U-Tests) finden sich unter *Nonparametric Tests \ 2 Independant Samples*. Bei beiden Verfahren muss der jeweilige Bereich der **Gruppierungsvariable** festgelegt werden.

1.4.3.2. ANOVA über Rangdaten

Im Menü **Transformieren** kann eine Rangfolge der Daten gebildet werden (*Rank Cases*). Anschließend berechnet man die ANOVA wie gewohnt.

1.4.4. Anmerkung

In der Übungssitzung zu mehrfaktoriellen Varianzanalysen wurde nochmals eine einfaktorielle ANOVA „zu Fuß“ mit Excel berechnet, d.h. tatsächlich die entsprechenden Varianzschätzer berechnet und dann miteinander verglichen. Das wird hier allerdings nicht aufgegriffen.

2. ANOVA II: Mehrfaktorielle ANOVA

Sind zwei oder mehr Faktoren gegeben, so muss neben dem **Haupteffektmodell** zusätzlich die **Interaktion** zwischen verschiedenen Faktoren berücksichtigt werden. Man kann jedoch auch fordern, dass alle Interaktionsterme Null sein müssen (überprüfbare Hypothese). Im zweifaktoriellen Fall gilt für einen Messwert y_{mij} einer Person m, die die Faktorstufen i und j der Faktoren A und B erhalten hat:

$$y_{mij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{mij}$$

Generell gilt dabei, dass sich alle **Effekte eines Faktors bzw. einer Interaktion zu Null summieren müssen**:

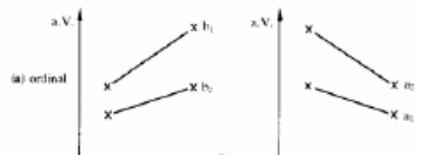
$$\sum \alpha_i = \sum \beta_j = \sum_{\text{Zeilen}} \alpha\beta_{ij} = \sum_{\text{Spalten}} \alpha\beta_{ij} = 0$$

Der Faktor ε umfasst dabei alle Effekt, die auf nicht erfasste Faktoren zurückgehen sowie wahre Fehler. Diese **Fehlerwerte** hingegen müssen **normalverteilt** sein, mit **Erwartungswert 0** und **fester Varianz σ^2** für allen Zellen. Ohne Normalverteilung wäre die Verteilung des F-Bruchs nicht bekannt und damit auch nicht nachzuschlagen (technischer Grund), zudem handelt es sich hier um eine Aneinanderreihung kleiner, unabhängiger Fehler (sachlogischer Grund: Zentraler Grenzwertsatz). Die Annahme $E(\varepsilon) = 0$ hat einen rein technischen Hintergrund, da sich sonst der Grand Mean ändern würde. Die Feste Varianz σ^2 resultiert aus der Voraussetzung, dass unter H_0 alle Stichproben (Zellen) aus derselben Grundgesamtheit gezogen werden.

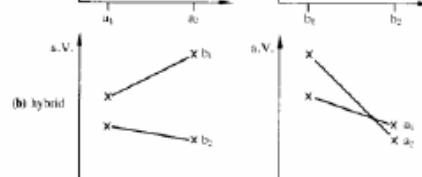
2.1. Interaktionen

Werden die Interaktionen geplottet (Faktor A auf x-Achse mit Faktor B als verschiedene Linien und umgekehrt), so können drei Arten von Interaktionen unterschieden werden. Generell gilt: Nichtparallele Linien deuten auf Interaktionen hin.

- **Ordinale Interaktion**



- **Hybride oder semi-ordinale Interaktion**



- **Disordinale Interaktion**

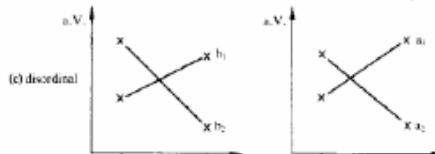


Abb. 8.2a–c. Klassifikation von Interaktionen: a ordinale Interaktion, b hybride Interaktion, c disordinale Interaktion

Anmerkung: Streng genommen werden hier keine Interaktionen dargestellt, sondern vielmehr die Stichprobenmittelwerte, die u.a. einen Interaktionsterm enthalten. Um nur die Interaktion darzustellen, müsste man von den Stichprobenmittelwerten die jeweiligen Randmittelwerte (als Schätzer der Effekte) abziehen und den Grand Mean addieren

2.2. ANOVA-Tabelle

Im mehrfaktoriellen Fall ergeben sich die Quadratsummen der einzelnen Bestandteile (Haupteffekte und Interaktionen) durch die Bereinigung der Gleichung um irrelevante Effekte. Hier werden dementsprechend nur die relevanten Mittelwerte / Randmittelwerte betrachtet.

Für jeden Faktor und jede Interaktion wird also eine Abweichungsquadratsumme berechnet. Diese Abweichungsquadratsumme entspricht einem Schätzer für die Grundgesamtheitsvarianz, die auf den vorhanden Faktor (bzw. Interaktion) zurückgeht. Diese Varianz muss standardisiert werden ($MS = SS/df$), sodass sie mit der standardisierten Fehlervarianz (MS_{err}) verglichen werden kann.

Q.d.V	SS	df	MS	F	p
Faktor 1	-	p-1	SS_1/df_1	$F_1 = MS_1/MS_{\text{err}}$	
Faktor 2	-	q-1	SS_2/df_2	$F_2 = MS_2/MS_{\text{err}}$	
Interaktion	-	(p-1)(q-1)	SS_{IA}/df_{IA}	$F_{IA} = MS_{IA}/MS_{\text{err}}$	
Fehler/error	-	pq(n-1)	$SS_{\text{err}}/df_{\text{err}}$		
Total	-	pqn-1			

Nebenbei: Der Erwartungswert einer F-verteilten Zufallsvariable (Bruch PV/FV) ist statistisch gesehen $E(X) = df_2/(df_2-2)$. Bei einigermaßen großen Stichprobenumfängen ist dies für praktische Zwecke gleich 1.

2.3. Anmerkungen

2.3.1. Geplante Kontraste und Post-Hoc-Tests

Wird der Haupteffekt eines Faktors mit mehr als 2 Stufen signifikant, müssen wie im einfaktoriellen Fall nachgeschaltete Einzelvergleiche über geplante Kontraste oder Post-Hoc-Tests durchgeführt werden. Auch Interaktionen lassen sich theoretisch über Kontraste prüfen, allerdings werden hier normalerweise einfach die entsprechenden Mittelwertdiagramme interpretiert.

2.3.2. Power (Teststärke)

Auftraggeber wie der DFG verlangen entsprechende Schätzungen teilweise bereits bei Projektanträgen. Zur Berechnung der Power gibt es spezielle Programme wie GPower (Buchner, Faul & Erdfelders):

<http://www.psych.uni-duesseldorf.de/aap/projects/gpower>

Alternativ lässt sich die Power auch aus Tabellen entnehmen. Beispielsweise geben Aron, Aron & Coups (2006) eine Tabelle mit notwendigen Stichprobenumfängen an, um mit 80% Power einen Effekt einer bestimmten Größe zu entdecken. Neben der Effektstärke ist der notwendige Stichprobenumfang natürlich auch vom Design (Menge der Faktorstufen) abhängig.

Die Schätzung der Effektstärke kann auf verschiedenen Wegen erfolgen. Im einfachsten Fall sind Werte in der Literatur dokumentiert oder auf Basis von Voruntersuchungen abschätzbar. Speziell bei der Evaluation neuer Treatments kann auch erst ein Effekt einer bestimmten Minimalgröße interessieren.

2.3.3. Wie bekommt man signifikante Ergebnisse?

Der F-Bruch wird größer, wenn:

- die wahren Mittelwertsunterschiede größer werden
- die Fehlervarianz abnimmt (weil genauer gemessen wird)
- bei konstanter Fehlervarianz mehr Personen untersucht werden (weil die df der MS_{error} proportional zu n sind)

2.3.4. Ungleiche Stichprobenumfänge

Ungleiche Stichprobenumfänge sind ärgerlich, weil das n_i in der Herleitung der Quadratsummen nicht mehr ausgeklammert werden kann und die Effekte nicht mehr unabhängig geschätzt werden können. Bei ungleichen Stichprobenumfängen ist die QS_{gesamt} also nicht mehr die Summe der restlichen Quadratsummen. Man hat also bei ungleichen Stichprobengrößen zwei Probleme:

- Das Verfahren ist [bei Verletzung der Varianzhomogenität nicht mehr robust](#) (das betrifft auch die einfaktorielle ANOVA)
- Das Design ist nicht mehr orthogonal – [die Effekte können nicht mehr unabhängig geschätzt werden](#) (das betrifft nur die mehrfaktorielle ANOVA).

Hierfür gibt es 6 verschiedene Lösungsansätze ([Typ I-VI-Summen](#)), die jeweils für spezifische Fragestellungen und Datensituationen angemessen sind (siehe 3.4).

Sind trotz aller Vorkehrungen ungleiche Stichprobenumfänge vorhanden und es ist kein Auffüllen möglich, so bieten sich zwei Vorgehensweisen an:

- Pbn aus allen Stichproben per Zufall entfernen, bis die Stichproben wieder gleich groß sind (casewise deletion). Dies ist insbesondere bei Messwiederholungsdesigns häufig die Voreinstellung.
- Ersetzen der fehlenden Daten durch den Zellenmittelwert oder eine geeignete Schätzung.

2.3.5. Feste und zufällige Effekte

Bisher wurden als unabhängige Variablen nur solche Faktoren betrachtet, deren Stufen vom VL explizit hergestellt wurden: Geschlecht (m/w), Alter (3 Gruppen) oder Treatment (Novum, Standard, Wartegruppe). Bei einer weiteren Untersuchung könnte man die Ausprägungen des Faktors dabei exakt wieder so herstellen.

Wird ein Haupteffekt signifikant, kann man sagen, dass ein Unterschied irgendwo zwischen genau diesen Stufen des Faktors vorliegt. [In diesem Fall spricht man von festen Effekten \(fixed factors\)](#).

Es kann jedoch auch sein, dass die Stufen der UV eine Zufallsauswahl aus vielen möglichen Stufen darstellen, z.B. Therapeutenpersönlichkeiten, Testleiter, Probanden. In diesem Fall kann man unter Umständen bei signifikantem Haupteffekt über die untersuchten Stufen hinaus auf die „Population“ dieser UV generalisieren. Man kann jedoch keine Aussagen über die realisierten Ausprägungen der UV treffen, da diese mehr oder weniger zufällig auftritt ([zufällige Effekte, random effects](#)).

Technisch berechnen sich die SS für feste und zufällige Effekte anders.

Beispiel aus Bortz: Leistungstest unter 3 Testinstruktionen (fixed factor), der von 8 unterschiedlichen Testleitern durchgeführt wird (random factor). Wird der Faktor „Testleiter“ signifikant, kann man sagen, dass die Testergebnisse von Testleitern generell abhängen – nicht nur von den 8 eingesetzten Testleitern. Die Aussage bezieht sich also auf die Population der möglichen Testleiter, aus denen die 8 verwendeten per Zufall gezogen wurden.

Ist der Faktor „Instruktion“ signifikant (ohne dass die Interaktion $TL * Inst$ signifikant wird), kann man sagen, dass die Testergebnisse von der Instruktion generell abhängen, also auch bei anderen TL der Population auftreten werden.

Ist die Interaktion signifikant, dann hängen die Instruktionseffekte davon ab, welcher der 8 Testleiter den Test durchführt. Der Haupteffekt der Instruktion lässt sich dann nicht mehr einfach so generalisieren.

Anmerkung: Wenn ein eigentlich zufälliger Faktor in einer ANOVA hypothetisch als fest angenommen wird, kann es vorkommen, dass ein anderer, zuvor signifikanter Faktor plötzlich nicht mehr signifikant wird. Der Effekt würde also für die Population der Effekte des zufälligen Faktors gelten (alle Testleiter), nicht jedoch für die Stichprobe (die 8 verwendeten Testleiter). Da dies ziemlich widersinnig ist, sollte ein zufälliger Faktor daher nur interpretiert werden, wenn er auch bei einer Analyse signifikant wird, bei der er als fester Faktor angenommen wird.

2.4. Übung II: Mehrfaktorielle ANOVA

2.4.1. Excel

Achtung: Die Analyse-Funktion „Zweifaktorielle Varianzanalyse“ in Excel kann nicht mit Interaktionseffekten umgehen weil sie nur mit einer Beobachtung pro Zelle rechnet. → Schwachsinn.

Natürlich geht auch die zweifaktorielle ANOVA zu Fuß in Excel, siehe Tabelleblatt *02_c_2fakt_Übung.xls*. Die Daten entsprechen dem Folienbeispiel aus der Themensitzung (alte, mittelalte und junge Fahrer, Männer und Frauen) und wurden mit einem Seed = 12 in Excel erstellt. Interessant ist vor allem die Schätzung der Interaktionseffekte (ab Zeile 152).

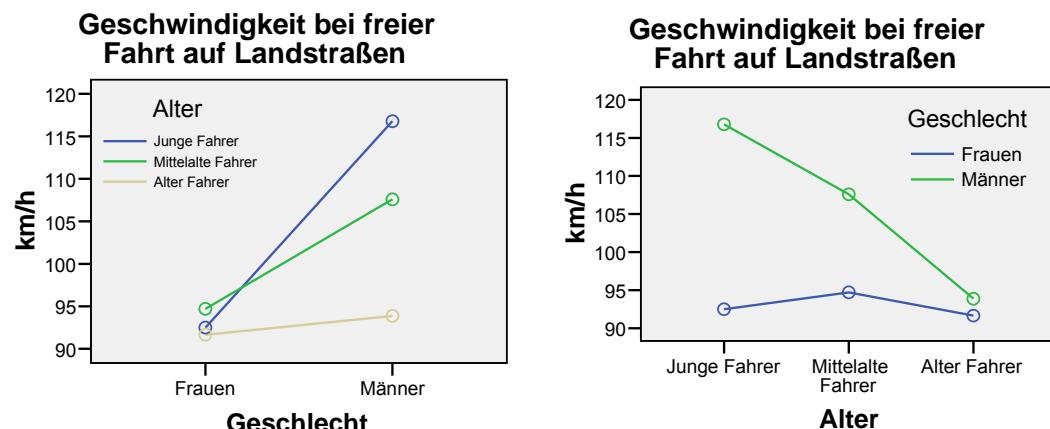
Die Berechnung per Hand beruht dabei auf der direkten Berechnung der Schätzer für die Grundgesamtheitsvarianz. MS_{error} berechnet sich dabei wie gehabt als Mittel der Varianzschätzer der einzelnen Stichproben, die Schätzung der drei Treatmentvarianzen erfolgt über $\hat{\sigma}_X^2 = \hat{\sigma}^2 \cdot n$ (Achtung: Eigentlich werden hier drei Tests berechnet!)

2.4.2. SPSS

Zur Kodierung: Die Faktoren lassen sich als **feste vs. Zufallsfaktoren** eingeben. Letztere Option stellt in Rechnung, dass zwei an sich gleiche Realisationen einer UV in der Praxis unterschiedlich sein können. Klassisches Beispiel hierfür ist die Therapievergleichsstudie. Werden hier mehrere Patienten mit derselben Therapieform aber unterschiedlichen Therapeuten behandelt, so erzeugt dieses Vorgehen unausweichlich Varianz trotz gleicher Ausprägung der UV.

Es wird daher auch nicht mehr von einer „Stufe“ der UV gesprochen. In der Interpretation muss man daher auch entsprechend vorsichtiger argumentieren: „Varianz in Faktor A beeinflusst die Ergebnisse systematisch“. Bei diesem Vorgehen

SPSS spuckt sofort Interaktionsdiagramme aus, allerdings ohne Fehlerbalken. Wie immer deuten nichtparallele Linien auf eine Interaktion hin.



Hier liegt eine disordinale Interaktion vor, da sich die Linien im linken Diagramm kreuzen und im rechten auch keine gleichgerichteten Steigungen zu beobachten sind.

Diese Beobachtung wird in der Interpretation rituell mit dem Satz „[...] Diese Haupteffekte werden durch eine Interaktion modifiziert...“ konstatiert. Eine mögliche Interpretation findet sich auf Folie 66:

„Es ergaben sich signifikante Haupteffekte für Geschlecht ($F(1, 294) = 119.7$), Alter ($F(2, 294) = 34.5$) und eine signifikante Interaktion ($F(2, 294) = 28.2; p$ in allen Fällen < 0.0005). Das Mittelwertsdiagramm zeigt, dass Frauen **tendenziell** geringere Geschwindigkeiten wählen als Männer, und dass ältere Fahrer **tendenziell** langsamer fahren als jüngere. Diese Effekte werden durch die Interaktion modifiziert: Während in der Gruppe der männlichen Fahrer ein deutlicher Effekt des Alters auf die Geschwindigkeit erkennbar ist, wirkt sich das Alter bei weiblichen Fahrern kaum auf die Geschwindigkeitswahl aus.“

2.4.3. Kontraste und Post-hoc-Tests

„Tendenziell“ ist hier ein Platzhalter für die nicht durchgeführten Kontrast- bzw. Post-hoc-Analysen. Diese werden notwendig, **sobald ein oder mehrere Faktoren mehr als 2 Stufen haben**. Deswegen begnügt man sich auch gerne mit 2x2-Designs.

Allerdings **machen Kontraste oder Post-Hoc-Tests nur dann Sinn, wenn keine Interaktion vorliegt**. Wenn eine Interaktion signifikant wird, gibt es zwei Herangehensweisen: **Simple effects analysis** oder **interaction comparison**.

2.4.3.1. Simple Effects Analysis

Die **Simple Effects Analysis** testet auf jeder Stufe des zweiten Faktors, ob sich die Stufen des ersten Faktors unterscheiden. Bei mehrstufigen Faktoren können Kontraste nachgeschaltet werden.

Beispiel für den Alterseffekt: Unterscheiden sich Frauen unterschiedlichen Alters? Unterscheiden sich Männer unterschiedlichen Alters?

Der Ansatz ähnelt also der Durchführung univariater ANOVAs innerhalb der Stufen eines Faktors.

2.4.3.2. Interaction Comparison

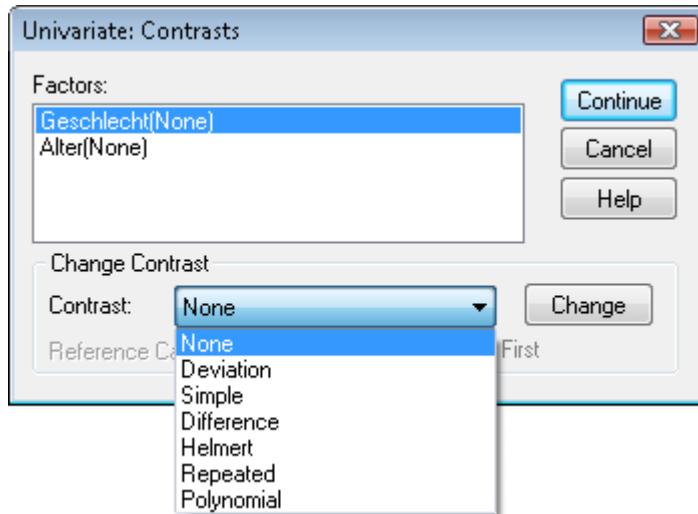
Die **Interaction Comparison** ist ein spezieller Kontrast, der aus sachlogischen Überlegungen abgeleitet sein muss.

Ärgerlicherweise lassen sich beide Analysen nur über die SPSS-Syntax berechnen und nicht über die normale Benutzeroberfläche einstellen. Zudem können verschiedene Kommandos zum Ziel führen.

2.4.3.3. SPSS-Kontraste

Kontraste bzgl. der Haupteffekte zu berechnen ist streng genommen nur möglich, wenn die Interaktion nicht signifikant wurde. Wenn dies doch der Fall ist, so muss auf die Simple Effect Analysis (separate einfaktorielle ANOVAs auf den einzelnen Faktorstufen) oder auf die Interaction Comparison zurückgegriffen werden. Letztere muss auf a-priori-Hypothesen beruhen und kann in SPSS nur über die Syntax kodiert werden.

SPSS bietet einige andere Kontraste an, die in den meisten Fällen genügen. Aber eben nur meistens. Kontrastdefinitionen in der Syntax: s. 3.5.2.



- **Abweichung (Deviation):** Vergleicht jede Stufe mit dem Grand Mean.
- **Einfach (Simple):** Vergleicht jede Stufe mit einer Referenzkategorie (letzte oder erste Stufe). Gut für Kontrollgruppendesign.
- **Differenz (Difference):** Vergleicht jede Stufe mit dem Mittelwert aller kleineren Faktorstufen.
- **Helmert:** Vergleicht jede Faktorstufe mit dem Mittel aller höheren Stufen.
- **Wiederholt (Repeated):** Vergleicht aufeinanderfolgende Faktorstufen (1 mit 2, 2 mit 3...) Teilweise bei MWH-Designs sinnvoll.
- **Polynomial:** Für Trendtests; linear, quadratisch, kubisch (Exponent des Polynoms = Anzahl der Kategorien – 1).

2.4.3.4. Post-Hoc-Tests

Eigentlich wird nur der Scheffé-Test regelmäßig verwendet.

2.4.4. Anmerkung: Mehrfaktorielle ANOVA und Multiples Testen

Streng genommen werden bei einer mehrfaktoriellen ANOVA immer mehrere Tests am selben Datenmaterial durchgeführt (wenigstens HE₁, HE₂ und WW), sodass auch die F-Tests mit einem Bonferroni-adjustierten α durchgeführt werden müssten.

Dies wird besonders bei ANOVAs mit vielen Faktoren kritisch, wird aber auch dort selten gemacht.

3. ANOVA III: Messwiederholung

Ein Messwiederholungsdesign liegt dann von, wenn von einem Probanden mehrere Messwerte erhoben wurden, sei es weil eine Variable zu verschiedenen Zeitpunkten gemessen wurde oder weil mehrere Variablen beobachtet wurden.

Synonyme (je nach Autor und Komplexität des Modells): [Within-subjects](#), repeated measures, Messwiederholungsdesign, one score per cell, [Random-Block](#), randomized blocks, matched-randomized changeover, crossover, [split-plot](#) ([Verlaufskurven](#))...

3.1. Einfaktorielle ANOVA mit Messwiederholung

Zunächst soll der einfachste Fall, also eine univariate, einfaktorielle ANOVA im Messwiederholungsdesign betrachtet werden. Es liegt also nur eine Stichprobe vor, wobei von den n VPn jeweils k Messungen erhoben wurden. Zusätzlich wird davon ausgegangen, dass es sich um einen Zeitfaktor handelt (vorher, nachher, später). [sog. Klassischer Ansatz].

3.1.1. Modell

Da mehrere Messungen einer Versuchsperson vorliegen, lassen sich die Eigenschaften der Versuchsperson getrennt vom Fehlerterm darstellen und bestimmen:

$$y_{mi} = \mu + \alpha_i + \pi_m + \varepsilon_{mi}$$

Der Fehlerterm ε_{mi} enthält dabei sowohl „echte“ Fehler als auch die Interaktion der Versuchsperson m mit der Stufe i des Messwiederholungsfaktors. Diese Effekte sind nicht zu trennen und werden somit als Fehler aufgefasst. [Der ursprüngliche Fehlerterm der between-Variante wird also in die beiden Komponenten \$\pi\$ und \$\varepsilon\$ zerlegt](#).

Insgesamt sind in dem Modell also feste Effekte (α) sowie zufällige Effekte (π und ε) enthalten. Da die VP allgemein als „Block“ bezeichnet wird und einen zufälligen Effekt (π) in das Design einbringt, heißt die Messwiederholungs-ANOVA auch häufig [Random-Block-Design](#).

3.1.1.1. Zusatzannahmen

Wie im bekannten Fall gilt:

- 1.) $\sum \alpha_i = 0$. Technische Voraussetzung, damit \bar{G} auch \bar{G} bleibt.

Zusätzliche Annahmen:

- 2.) Die π_m sind in der Population normalverteilt mit $\mu = 0$ und Varianz σ^2_p .
- 3.) Die ε_{mi} sind innerhalb jeder Stufe der MWH normalverteilt mit $\mu = 0$ und fester Varianz σ^2_e .
- 4.) Die Fehler ε_{mi} sind unabhängig von den VP-Faktoren π_m .

Alternative Formulierung: 1.) Das Merkmal (y) muss auf allen Stufen des MWH-Faktors normalverteilt sein ([Normalverteilung](#)). 2.) Die Populationsvarianzen müssen auf allen Stufen gleich sein ([Varianzhomogenität](#)) und 3.) (neu): Bei mehr als 2 Stufen müssen die Varianzen aller möglichen Differenzen gleich sein ([Zirkularitätsbedingung](#) bzw. [Sphäritätsannahme](#)).

3.1.1.2. Zirkularitätsbedingung

Sobald ein Messwiederholungsfaktor mehr als 2 Stufen hat braucht man also eine zusätzliche Voraussetzung zu Normalverteilung und Homoskedastizität.

Eine **hinreichende Bedingung (Homogenität der Kovarianzen)** besagt, dass die Kovarianzen (Korrelationen) der Messwerte für alle Paare von Stufen der Messwiederholung gleich sein müssen. Man kann synonym von Kovarianzen und Korrelationen sprechen, da ja die Homoskedastizitätsannahme gemacht wird. ($r_{xy} = \text{cov}_{xy}/(s_x \cdot s_y)$).

Die „Reihenfolge“ der VPn muss also auf Stufe 1 und 2 ähnlich sein wie zwischen den Stufen 2 und 3. Diese Annahme ist insbesondere dann gern verletzt, wenn Stufe 1 und 2 zeitlich nahe beieinander liegen und Stufe 3 deutlich später erfolgt bzw. wenn Deckeneffekte auftreten. Ersteres ist darauf zurückzuführen, dass zeitliche weit auseinanderliegende Messwerte im Allgemeinen schwächer korrelieren als zeitlich nahe Messwerte.

Diese Zusatzannahme kann in Form der **Sphäritätsannahme (engl.: Sphericity oder Compound symmetry)** etwas strenger und mathematisch sauberer gefasst werden. Die Sphäritätsannahme ist eine **notwendige und hinreichende Bedingung** für die Geltung des Modells aber nicht mehr natürlichsprachlich auszudrücken. Diese Annahme wird u.a. in SPSS mit **Mauchly's Test auf Sphärität** geprüft.

Für den Moment sollen die Voraussetzungen allerdings erst einmal als erfüllt angesehen werden.

3.1.2. Ablauf

Zur Erinnerung: Es werden n Versuchspersonen untersucht (Laufindex: m), die zu k verschiedenen Zeitpunkten gemessen werden (Laufindex des MWH-Faktors: i).

Wie bisher werden die Effekte im Datenmodell wieder über Randmittelwerte geschätzt. Hierzu wird die Abweichungsquadratsumme der Werte vom Grand Mean wieder in einzelne Komponenten zerlegt und Varianzschätzer auf Basis der einzelnen Faktoren berechnet.

Die einzige Änderung ergibt sich für die Berechnung der Fehler-QS, die nun so ähnlich berechnet wird, wie die Interaktionsquadratsummen in der zweifaktoriellen ANOVA (mit denselben df). Sie enthält ja auch die IA VP * Zeitpunkt.

3.1.3. ANOVA-Tabelle(n)

Der Übersicht halber wird auch hier nur ein Teil der von SPSS generierten Tabellen dargestellt. Eigentlich interessiert dabei immer nur die orangene Zeile.

Q.d.V	SS	df	MS	F	p
Faktor 1 (ε)	-	k-1	SS_1/df_1	$F_1 = MS_1/MS_{\text{Res}}$	
Faktor 2 (π)	-	n-1			
Residuum (ε)	-	(n-1)(k-1)			
Total	-	nk-1			

Anmerkung: **Der Fehler heißt jetzt Residuum** (weil die IA enthalten ist).

Im Bortz findet sich eine etwas umfangreichere Tabelle, die der ursprünglichen Idee der Additivität von Quadratsummen und Freiheitsgraden folgt. Grundlegend ist hier die Unterscheidung von Variation zwischen den VPn (entspricht π ; im MWH-Design uninteressant) und Variation innerhalb der VPn. Letztere wird unterteilt in Variation aufgrund des Treatments (α) und Residuum (ε).

Q.d.V	SS	df	MS	F	p
Zwischen VPn	X	n-1			
Innerhalb VPn	Y	n*(k-1)			
Treatment	T	k-1	SS ₁ /df ₁	F ₁ = MS ₁ /MS _{Res}	
Residuum (ε)	U	(n-1)(k-1)			
Total	Z	nk-1			

Dabei gilt sowohl für Quadratsummen als auch dazugehörige Freiheitsgrade: $X + Y = Z$ mit $Y = T + U$. Wirklich interessant bleibt trotzdem nur die orange eingefärbte Zeile. Was die veränderte Darstellung bringt ist fraglich ☺.

Im SPSS-Output ist zunächst auch nur ein kleiner Teil relevant:

Tests of Within-Subjects Effects

Measure: Ärgerwerte (vor, direkt nach und nach 60 min Pause; Treatment: Zwickeln, Kitzeln, Beissen).

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Treatment	70.933	2	35.467	40.151	.000
	70.933	1.997	35.517		
	70.933	2.000	35.467		
	70.933	1.000	70.933		
Error (Treatment)	7.067	8	.883		.003
	7.067	7.989	.885		
	7.067	8.000	.883		
	7.067	4.000	1.767		

Die beiden mittleren Verfahren (Greenhouse-Geisser und Huynh-Feldt) nehmen Adjustierungen der Freiheitsgrade vor, um einer Verletzung der Sphäritätsannahme Rechnung zu tragen (vgl. 3.1.6.3). Die Untergrenze (Lower-bound) ist der konservativste Test.

3.1.4. Nonparametrisch: Friedman-Varianzanalyse

Das nichtparametrische Äquivalent zur einfaktoriellen ANOVA mit Messwiederholung ist die **Friedman-(Rang-)Varianzanalyse** bzw. der **Kendall-W-Koeffizient**, die beide kurz in der Übung angesprochen werden.⁵

3.1.5. Kontraste und Post-Hoc-Tests

Die Kodierung von Kontrasten in MWH-Designs ist in SPSS nicht ganz selbsterklärend. Daher findet sich eine detaillierte Beschreibung in der Übung (siehe 3.5.2).

⁵ Anmerkung: Zu mehrfaktoriellen ANOVAs mit Messwiederholung (vgl. 3.2) existieren keine nichtparametrischen Gegenstücke.

3.1.6. Voraussetzung: Sphäritätsannahme

Within-ANOVAs haben diverse Vorteile. So laufen Experimente etwa effizienter ab, systematische Unterschiede können leichter entdeckt werden (Kontrolle der Personenvarianz) und manche Designs, wie etwa Vorher-Nachher-Designs, zwingen zur mehrfachen Messung.

Probleme können dabei immer dann auftreten, wenn eine Interaktion zwischen VP-Mittelwert und Zeitpunkt auftreten (z.B. wenn eine gute VP wenig dazulernen: Deckeneffekt) oder wenn sich die VPn nur wenig systematisch voneinander unterscheiden (dann kann ein Between-Plan empfindlicher sein).

Zusätzlich muss man aufgrund der Sphäritätsannahme davon ausgehen, dass keine Lerneffekte auftreten, die nicht im Treatment enthalten sind – die mehrfache Testung also solche also keinen Effekt auf die Messwerte hat. Zusätzlich müssten sämtliche Kovarianzen konstant sein.

Wie bereits erwähnt ist die [Sphäritätsannahme](#) gerne verletzt, sodass man sich überlegen sollte, wie in einem solchen Fall vorgegangen werden kann. Einige Möglichkeiten sind:

- 1.) Ignorieren
- 2.) Statistisch prüfen
- 3.) Adjustieren
- 4.) Verfahren anwenden, die diese Annahme nicht benötigen

3.1.6.1. Ignorieren

Um überhaupt etwas rechnen zu können ist dies der einfachste Weg („Ich gehe davon aus, dass sie gilt“). Dieser Weg muss bei vielen Faktorstufen gewählt werden, weil man die anderen Verfahren sonst nicht anwenden kann.

3.1.6.2. Statistisch prüfen

Eine Prüfung ist beispielsweise über [Mauchly's Test auf Sphärität](#) möglich. Hier gelten allerdings ähnliche Probleme wie bei den anderen betrachteten Voraussetzungstests: Streng genommen führt man wieder mehrere Tests am selben Datenmaterial durch und die Voraussetzungstests sind empfindlicher gegen Verletzungen der Annahme als das Verfahren selbst.

3.1.6.3. Adjustieren

Bei Verdacht auf Verletzung der Annahmen (etwa aufgrund statistischer Prüfung) kann wie bei einfaktoriellen ANOVAs eine Adjustierung der Freiheitsgrade vorgenommen werden.

SPSS berechnet automatisch zwei Varianten: Greenhouse-Geisser und Huynh-Feldt, wobei die Verfahren für praktische Zwecke selten zu unterschiedlichen Ergebnissen kommen. Auch die sehr konservative Untergrenze (Lower-bound) kann hierfür herangezogen werden.

3.1.6.4. Verfahren ohne Sphäritätsannahme

Die relevanten Verfahren basieren auf dem allgemeinen linearen Modell und können auf dieselben Fragestellungen angewendet werden, ohne von der Sphäritätsannahme abhängig zu sein.

Sobald man Zugang zu einem Computer hat ist das beschriebene Verfahren also weitestgehend obsolet (Ausnahme: sehr kleines n), da dann das ALM verwendet werden sollte. Die Idee hinter der multivariaten Auswertung wird unter 5.4 besprochen. Im Prinzip beruht sie aber auf einer trickreichen Umformulierung der Datensituation.

3.1.7. Missing Data

Auch hier gibt es 4 Möglichkeiten:

- 1.) Casewise deletion
- 2.) Ersetzen fehlender Werte
- 3.) Analyse mit und ohne Missing Data berechnen
- 4.) Missing Data als Datum

3.1.7.1. Casewise Deletion

VPn mit unvollständigen Daten können komplett ausgeschlossen werden. Dieses Vorgehen ist zwar die Standardeinstellung vieler Statistikpakete, allerdings nur dann akzeptabel, wenn nur wenige Fälle ausfallen und diese Fälle per Zufall über die Stichprobe verteilt sind. Und das ist unwahrscheinlich.

3.1.7.2. Ersetzen fehlender Werte

Hier können verschiedene Methoden verwendet werden, um Schätzwerte für die fehlenden Zellen zu bestimmen:

- A priori-Schätzung (z.B. „last value carried forward“; das macht aber Bauchweh, weil der letzte bekannte Wert einfach weiter eingesetzt wird. Schlauer: Multiple Regression.)
- Zellenmittelwert
- Multiple Regression, die auf vollständigen Datensätzen berechnet wird
- Komplexere Schätzverfahren wie Erwartungswertmaximierung

Bei einem beliebten Verfahren wird der fehlende Wert durch einen Schätzwert aus den Faktorstufen geschätzt. Dabei wird für jede relevante Between- und Within-Faktorstufe sowie für die VP der Mittelwert der vorhandenen Daten für die Schätzung verwendet. Eine Formel findet sich auf den Folien 48ff.

3.1.7.3. Analyse mit und ohne Missing Data berechnen

Die erhaltenen Ergebnisse können dann miteinander verglichen werden, um so einen Eindruck der Güte der Schätzung zu erhalten.

3.1.7.4. Missing Data als Datum

Missing Data kann auch per se ein Datum sein. Man kann sich beispielsweise fragen, warum in bestimmten Gruppen besonders viele VPn ausfallen. Dazu wird eine neue Variable generiert, die den Ausfall kodiert (also: Gruppe ist vollständig vs. nicht vollständig). Diese Variable kann dann als Prädiktor in die Analyse integriert werden.

3.2. Verlaufskurven (Split-Plot)

Im Folgenden soll die ANOVA auf mehrere Faktoren erweitert werden, die sowohl Between- als auch Within-Faktoren sein können. Die Kombination der beiden unterschiedlichen Faktorenarten wird auch allgemein als **Split-Plot** bezeichnet.

Unter einer **Verlaufskurve** versteht man die Messung mehrerer abhängiger Variablen auf derselben Skala an mehreren Gruppen. Diese Gruppenzugehörigkeit bildet den Between-Faktor. Dabei wird der Messwertvektor einer VP häufig auch als Profil bezeichnet (**Profilanalyse**). Manche Autoren bezeichnen Abschnitte der Profile zusätzlich als **Segmente**.

Speziell bei diesen Fragestellungen (etwa: Therapie vs. Kontrollgruppe mit Messung vorher, nachher und nach 6 Wochen), interessiert weder der Haupteffekt der Gruppenzugehörigkeit, noch der Haupteffekt des Within-Faktors (es wird über die Zeit besser), sondern die Interaktion.

Konkret sind also folgende Fragestellungen enthalten:

- Sind die Profile parallel? → Wenn nein: Interaktion.
- Gibt es einen Lageunterschied zwischen den Gruppen? → Entspricht Between-Haupteffekt.
- Wenn die Profile parallel sind: Sind die Profile flach? → Frage nach einem Within-Haupteffekt.

3.2.1. Modell

Im Datenmodell der einfachen ANOVA mit MWH wurde neben Grand Mean, Haupteffekt des Treatments und Fehler ein Haupteffekt der Person eingeführt. Die Interaktion zwischen Person und Ausprägung der eigentlichen UV wurde im Fehlerterm belassen.

Hier wird das einfache Modell um einen zweiten Haupteffekt erweitert (α ist jetzt der between-Faktor, β der within-Faktor). Jede Versuchsperson $m(i)$ ist jetzt in genau einer Gruppe (Stichprobe), die durch die Ausprägung i des Faktors α markiert wird.

$$y_{mi} = \mu + \alpha_i + \beta_j + \pi_{m(i)} + \alpha\beta_{ij} + \pi\beta_{m(i)j} + \varepsilon_{mij}$$

Die Interaktion $\pi\beta_{m(i)j}$ wird also explizit hingeschrieben, obwohl sie nicht separat geschätzt werden kann. Diese Interaktion und der Fehler ε_{mij} bilden also die Residualvarianz.

Für die **Quadratsummen** (und entsprechend für die jeweiligen **df**) ergibt sich hier folgende Zerlegung:

$$\begin{aligned} SS_{\text{total}} &= SS_{\text{zwischen_VPn}} + SS_{\text{innerhalb_VPn}} = \\ &= SS_A + SS_{\text{in_Stp}} + SS_B + SS_{\text{AxB}} + SS_{\text{BxVP}} \end{aligned}$$

Mit $df_{\text{total}} = k \cdot l \cdot n - 1$ und k = Anzahl der Stufen des Between-Faktors, l = Anzahl der Stufen des Within-Faktors und n = Anzahl der VPn pro Stufe des Between-Faktors. Die Überlegungen zu Freiheitsgraden der sonstigen QSn finden sich auf den Folien 17-19.

Neben Quadratsummen und Freiheitsgraden, hängen die F-Brüche noch davon ab, ob die **Faktoren** als **fest oder zufällig** angenommen werden. Für den einfachsten Fall (beide fest) ergeben sich folgende Regeln:

- $HW_A: F = \frac{MS_A}{MS_{in_Stichprobe}}$
- $HW_B: F = \frac{MS_B}{MS_{BxVPn}}$
- $HW_{AxB}: F = \frac{MS_{AxB}}{MS_{BxVPn}}$

Der Between-Effekt wird bei **festen Effekten** also an der Restvarianz innerhalb der Stichproben geprüft (wie bei der „normalen“ ANOVA). Der Within-Effekt hingegen wird hier an der Interaktion BxVPn geprüft.

Bei **zufälligen Effekten** werden andere Prüfvarianzen im Nenner (und evtl. auch im Zähler) verwendet. Details finden sich in der Literatur. Generell werden die beiden Schätzer jedoch so konstruiert, dass sie sich nur um den Varianzbeitrag des zu prüfenden Effekts unterscheiden.

3.2.2. ANOVA-Tabelle

Q.d.V	SS	df	MS	F	p
Faktor A	SS_A	$k-1$	SS_A/df_A	MS_A/MS_{in_Stpn}	
Innerhalb Gruppen	SS_{in_Stpn}	$k(n-1)$	SS_{i_S}/df_{i_S}		
Zwischen VPn	SS_{zw_VPn}	$kn-1$			
Faktor B	SS_B	$l-1$	SS_B/df_B	MS_B/MS_{BxVPn}	
A x B	SS_{AxB}	$(k-1)(l-1)$	SS_{AxB}/df_{AxB}	MS_{AxB}/MS_{BxVPn}	
Rest innerhalb	SS_{BxVP}	$k(l-1)(n-1)$	SS_{BxVP}/df_{BxVP}		
Innerhalb VPn	SS_{in_VPn}	$k^*n^*(l-1)$			
Total	SS_{Total}	$kln-1$			

Dabei addieren sich jeweils die gleich eingefärbten Zellen einer Spalte zur nächstdunkleren Zelle, z.B. $SS_A + SS_{in_Stpn} = SS_{zw_VPn}$. Die angegebene Berechnung der F-Brüche gilt natürlich nur für feste Effekte.

3.2.3. Anmerkungen

Wie bei der einfaktoriellen ANOVA mit Messwiederholung besprochen, setzen Designs mit mehr als 2 Stufen Sphärität voraus. Diese Annahme ist jedoch häufig verletzt, speziell bei einer Messwiederholung über die Zeit.

Vor der Entwicklung des ALM wurde in diesen Fällen eine Adjustierung der Freiheitsgrade verwendet (Greenhouse-Geisser, Huynh-Feldt). Inzwischen ist es jedoch üblich, solche Designs mit multivariaten Verfahren auszuwerten, die schwächere Annahmen machen.

Bisher wurde unter Split-Plot-Designs (Verlaufskurven) die spezielle Datensituation verstanden, die sowohl klassisch als auch nach dem ALM auszuwerten ist. Manche Autoren (wie Tabachnik und Fidel) reservieren den Begriff **Profilanalyse** für die Auswertung von Messwiederholungsdesigns mit multivariaten Verfahren. Generell sollte bei genügend Freiheitsgraden immer mit dem multivariaten Ansatz ausgewertet werden, den SPSS auch automatisch berechnet.

3.3. Profilanalyse (MANOVA bei Split-Plot)

Auch hier lassen sich die bekannten drei Hypothesen aufstellen:

- Verlaufen die Profile nicht parallel? (Interaktion)
- Gibt es einen Lageunterschied zwischen den Gruppen? (Haupteffekt des Between-Faktors)
- Wenn die Profile parallel sind: Sind die Profile flach? (Within-Haupteffekt). Diese Frage stellt sich natürlich nur bei fehlender Interaktion.

3.3.1. Voraussetzungen

- Der Umfang der kleinsten Stichprobe muss größer sein als die Anzahl der Stufen des Messwiederholungsfaktors. Ist diese Voraussetzung verletzt bleibt nur die univariate ANOVA.
- Ungleiche Zellgrößen sind weniger problematisch, da die einzelnen Forschungsfragen wie in einer einfaktoriellen ANOVA getestet werden. Bei Ausfall von VPn kann ggf. eine Datenschätzung eingesetzt werden.
- Die Normalverteilungsannahme ist unkritisch, so lange Bedingung 1 erfüllt ist und die Stichprobenumfänge gleich sind.
- Ausreißer beeinflussen die MANOVA sehr stark.
- Bei gleichen Stichprobengrößen ist die Annahme der Varianzhomogenität (hier der Varianz-Kovarianz-Matrix) unkritisch.

3.3.2. Beispiel

Untersuchung in Reha-Klinik zur Schmerztherapie: Vorher/Nachher/Nach 6 Wochen werden 3 Versuchsgruppen gemessen (medizinische Behandlung, zusätzliche Standardpsychotherapie, zusätzliche neue Therapie).

Multivariate Tests^a

Effekt		Wert	F	Hypothese df	Fehler df	Signifikanz
zeit	Pillai-Spur	.266	6.175 ^a	2.000	34.000	.005
	Wilks-Lambda	.734	6.175 ^a	2.000	34.000	.005
	Hotelling-Spur	.363	6.175 ^a	2.000	34.000	.005
	Größte charakteristische Wurzel nach Roy	.363	6.175 ^a	2.000	34.000	.005
zeit * Treatment	Pillai-Spur	.046	.410	4.000	70.000	.801
	Wilks-Lambda	.954	.403 ^a	4.000	68.000	.806
	Hotelling-Spur	.048	.396	4.000	66.000	.811
	Größte charakteristische Wurzel nach Roy	.048	.837 ^b	2.000	35.000	.442

SPSS bietet 4 verschiedene Testverfahren an, deren Ergebnisse sich meist jedoch nicht voneinander unterscheiden. Die Pillai-Spur gilt als robustester Test, insbesondere bei ungleichen Zellgrößen. Wilks-Lambda und die Hotelling-Spur sind dagegen etwas stärker. Meist wird Wilks-Lambda verwendet, weil es etwas bekannter ist.

Kontraste und Post-Hoc-Tests lassen sich in SPSS nicht per Mausklick berechnen und werden in der Übung genauer unter die Lupe genommen. Mehr Infos zur multivariaten Auswertung finden sich im Kapitel MANOVA (vgl. 5.4).

3.4. Ungleiche Zellbesetzungen und Typ I-VI-Summen

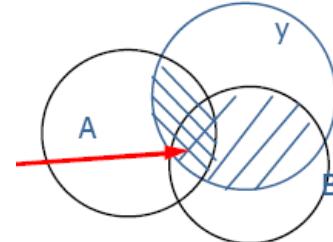
Die klassische Varianzanalyse geht davon aus, dass alle Stichproben gleich groß sind. Dadurch werden die Gleichungen einfach und es ist möglich, die Effekte unabhängig voneinander zu schätzen.

Bei einfaktoriellen ANOVAs werden ungleiche Stichprobenumfänge speziell bei Varianzinhomogenität problematisch (df adjustieren).

Man kann zwar immer versuchen, die Stichproben gleich groß zu halten (z.B. Pbn per Zufall aus „zu großen“ Stichproben hinauswerfen oder fehlende Daten schätzen), allerdings ist dies nicht immer angemessen. Spätestens wenn der Verdacht nahe liegt, dass VPn systematisch auf bestimmten Stufen fehlen muss man anders arbeiten.

Speziell für mehrfaktorielle ANOVAs gilt zusätzlich, dass die Gesamt-QS nicht der Summe der QS der Effekte entspricht. Die Tests der Haupteffekte und Interaktionen sind also nicht mehr unabhängig voneinander. Dies liegt daran, dass auf Basis der Kenntnis der Ausprägung eines Faktors über die zugrunde liegende Stichprobengröße Rückschlüsse auf die Ausprägung des zweiten Faktors möglich sind (= Korrelation der UVn!).

Zur Lösung dieser Probleme gibt es 6 verschiedene Lösungsansätze (sog. **Typ I-IV-Summen**), die jeweils für spezifische Fragestellungen und Datensituationen angemessen sind. Keine von ihnen kann jedoch in jedem Fall angewendet werden.



Standard ist dabei die Typ-3-Summe; siehe 3.4.3, bei der die gemeinsame Varianz gar nicht betrachtet wird (konservativ). Das Vorgehen der einzelnen Methoden soll nun kurz betrachtet werden.

3.4.1. Typ 1 – Summen: Sequentiell

Wie bei sequentiellen/hierarchischen Modelltests in der Regressionsanalyse baut man auch hier der Reihe nach die Parameter in das Modell ein und prüft, ob das Zufügen eines Parameters die Varianzaufklärung (R^2) erhöht (vgl. 4.3.2.1).

Typ 1 – Summen werden deshalb auch als **hierarchische oder sequentielle Summen** bezeichnet.

Die gemeinsame Varianz zweier Faktoren wird also dabei dem ersten Faktor zugerechnet, der in das Modell aufgenommen wird.

Vorteil: Die Typ 1 – Quadratsummen behalten die mathematisch elegante Eigenschaft der Additivität. Allerdings hängen ihre Ergebnisse von der Reihenfolge der Eingabe der Faktoren ab.

3.4.2. Typ 2 – Summen: Teilsequentiell

Bei den **teilsequentiellen Typ 2 – Summen** wird die QS für einen Effekt berechnet, indem man gegen alle Effekte desselben oder geringeren Grades kontrolliert. Daher ist die Reihenfolge der Modellparameter egal, sodass diese Summen für Regressionsfragestellungen, Haupteffekt-ANOVAAs, faktorielle

ANOVAs mit gleichen Zellgrößen oder Hierarchische Nested Designs tauglich sind.

Bei ungleichen Zellgrößen sind die Tests / Hypothesen allerdings nicht interpretierbar und damit unbrauchbar.

3.4.3. Typ 3 – Summen: Standard

Die häufig, bspw. in SPSS und Statistica, als Standardeinstellung verwendeten Typ 3 – Quadratsummen erlauben am ehesten eine „normale“ Interpretation der Effekte. Dabei wird die gemeinsame Varianz der UVn herausgerechnet und nicht berücksichtigt.

Nur der Anteil der Varianz, den ein Faktor inkrementell beiträgt, geht auch in die Berechnung der QS ein. Dies führt zu einer konservativeren Testung.

Die Typ 3 – Summen sind die einzige Methode, mit der faktorielle ANOVAs mit ungleichen Zellgrößen getestet werden können. Dabei werden KQ-Mittelwerte verglichen (Schätzung für Randmittel).

Der Vorteil der Anwendbarkeit bei ungleichen Zellbesetzungen wird jedoch für einen Nachteil erkauft: Typ 3 – Summen sind bei Missing Cells nicht interpretierbar. Hier muss auf eine der anderen Varianten zurückgegriffen werden.

3.4.4. Typ 4 – Summen: Nicht empfohlen

Dieser Typ der Varianzzerlegung wurde speziell für Datensätze mit Missing Cells entwickelt. Allerdings hängen die Ergebnisse von der Ordnung der Stufen der kategorialen Prädiktoren ab und die getesteten Hypothesen sind abhängig vom Muster der höherwertigen Interaktionen und damit nicht interpretierbar.

Finger weg!

3.4.5. Typ 5 – Summen: Aus der Industrie

Typ 5 – Summen stellen eine Spezialmethode für Designs mit vielen Faktoren und vielen Stufen dar, die nicht vollfaktoriell geprüft werden können. Eine Interpretation ist möglich, aber schwierig.

3.4.6. Typ 6 – Summen: Effective Hypothesis

Die Typ 6 – Quadratsumme eines Effekts ergibt sich, indem man von der Gesamt-QS die Quadratsummen aller anderen Effekte abzieht. Bei Missing Cells kann dieses Vorgehen dazu führen, dass die resultierende QS keine Freiheitsgrade mehr aufweist und somit nicht testbar ist. [Besser nicht als verzerrt testen.]

Typ 6 – Quadratsummen können nicht bei Nested Designs, Separate-Slope-ANCOVAs oder Mixed Models eingesetzt werden.

3.5. Übung III: ANOVA mit Messwiederholung

3.5.1. Einfaktorielle ANOVA mit Messwiederholung

Alle SPSS-Outputs sind aus: 03_f_Eigenes SPSS-Beispiel_Output.spo.

Im Output werden zunächst **multivariate Tests** berichtet (bzw. vorher eine Tippfehlerkontrolle), von denen meist Pillai oder Wilks verwendet wird. Wie immer werden von SPSS alle Maße berichtet, die irgendwie berechnet werden können (hier: 5 Tests bei der gegebenen Datenlage möglich). Die eigentliche MWH-ANOVA findet sich also weiter unten.

Multivariate Tests(b)

Effect		Value	F	Hypothesis df	Error df	Sig.
Treatment	Pillai's Trace	.954	30.866(a)	2.000	3.000	.010
	Wilks' Lambda	.046	30.866(a)		3.000	.010
	Hotelling's Trace	20.578	30.866(a)		3.000	.010
	Roy's Largest Root	20.578	30.866(a)		3.000	.010

a Exact statistic

b Design: Intercept

Within Subjects Design: Treatment

Während hier eigentlich nur die Spalte der Signifikanz interessiert (wenn $< .05$ dann gibt es einen MWH-Effekt), ist eine niedrige Wahrscheinlichkeit in **Mauchlys Test auf Sphärität** ungünstig, da diese Annahme dann nicht erfüllt ist. Da es sich hier im einen Voraussetzungstest handelt (H_0 soll beibehalten werden), müsste fairerweise ein höheres Alpha (.20/.25) verwendet werden.

Ist die Sphäritätsannahme verletzt (hier nicht der Fall), dann müssen die df der **Tests auf Innersubjekteffekte** adjustiert werden (Greenhouse-Geisser oder Huynh-Feldt), sodass sich auch die MS ändern. Da dies hier bei Treatment- und Fehlervarianzschätzer gleichermaßen passiert, ändern sich die F-Brüche nicht. Das muss aber nicht so sein. Alternativ lässt sich natürlich gleich der multivariate Test verwenden.

Tests of Within-Subjects Effects

Measure: Ärgerwerte

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Treatment	Sphericity Assumed	70.933	2	35.467	40.151	.000
	Greenhouse-Geisser	70.933	1.997	35.517	40.151	.000
	Huynh-Feldt	70.933	2.000	35.467	40.151	.000
	Lower-bound	70.933	1.000	70.933	40.151	.003
(Treatment)	Sphericity Assumed	7.067	8	.883		
	Greenhouse-Geisser	7.067	7.989	.885		
	Huynh-Feldt	7.067	8.000	.883		
	Lower-bound	7.067	4.000	1.767		

a Exact statistic

b Design: Intercept

Within Subjects Design: Treatment

Die Residuen (Error) können theoretisch auch auf Signifikanz geprüft werden. Das geht über die Syntax – allerdings ist das in den seltensten Fällen wirklich interessant.

Schließlich berechnet SPSS automatisch einen **Test der Innersubjektkontraste**. Dieser macht allerdings nur Sinn, wenn die Zeitpunkte gleich weit voneinander entfernt sind.

nander entfernt sind. Ist dies nicht der Fall, muss ein Trendtest mit der multiplen Regression durchführen, da dort die Lage der Mittelwerte auf der X-Achse (Zeitabstand) explizit kodiert wird. In der ANOVA hingegen werden fortlaufend Werte kodiert, und stillschweigend Gleichabständigkeit impliziert.

Zusätzlich gibt SPSS einen [Test der Zwischensubjekteffekte](#) aus. Da in diesem Fall aber gar keine Between-Faktoren enthalten waren ist die Tabelle überflüssig. Eine Signifikanz des konstanten Terms sagt lediglich, dass der Grand Mean ungleich 0 war (super...).

Zum [Plot](#): SPSS zeigt einen gestutzten Wertebereich, der ggf. einen falschen Eindruck der Veränderung entstehen lässt. Hier empfiehlt sich eine Anpassung der Y-Achse auf den gesamten Wertebereich.

3.5.1.1. Einzelvergleiche

Um zu überprüfen woran ein signifikanter Haupteffekt liegt nimmt man geschickterweise alle möglichen Paarvergleiche. Diese werden über t-Tests für verbundene Stichproben mit Bonferroni-adjustiertem Alpha geprüft.

Interessant ist hier eigentlich nur die dritte Tabelle zu [Tests bei gepaarten Stichproben](#). Die ersten beiden enthalten jede Menge Deskriptiva und Korrelationen.

Paired Samples Test

	Paired Differences					T	df	Sig. (2-tailed)			
	Mean	Std. Error Mean	95% CI (Difference)								
			Lower	Upper							
Vorher - Nachher	-4.40000	1.34164	-6.06587	-2.73413	-7.333	4	.002				
Vorher - Später	.40000	1.34164	-1.26587	2.06587	.667	4	.541				
Nachher - Später	4.80000	1.34164	3.18107	6.41893	8.232	4	.001				

Hier liegt die Signifikanz der ANOVA wohl eindeutig am hohen Wert direkt nach dem Treatment „Zwicken, Kitzeln, Beissen“.

3.5.1.2. Nichparametrische Äquivalente

Die nichparametrischen Äquivalente für den Omnibus-Test sind die [Friedman-\(Rang-\)Varianzanalyse](#) oder auch der [Kendall-W-Test](#). Beide berechnen ein χ^2 und prüfen dieses auf Signifikanz.

Für nonparametrische Einzelvergleiche stehen zwei Alternativen zur Verfügung: Wilcoxon-Rangsummentest und Vorzeichentest. Der [Wilcoxon-Test](#) gibt eine Tabelle zu den Rangsummen der einzelnen Variablen aus (blau markiert), die verwendet werden kann, um in den exakten Tabellen für kleine Stichprobenumfänge nachzuschlagen (Überschreitet die jeweilige Rangsumme den kritischen Wert?).

Ranks – Wilcoxon Test

		N	Mean Rank	Sum of Ranks
Nachher - Vorher	Negative Ranks	0(a)	.00	.00
	Positive Ranks	5(b)	3.00	15.00
	Ties	0(c)		
	Total	5		

Später - Vorher	Negative Ranks		3(d)	3.33	10.00
	Positive Ranks		2(e)	2.50	5.00
	Ties		0(f)		
	Total		5		
Später - Nachher	Negative Ranks		5(g)	3.00	15.00
	Positive Ranks		0(h)	.00	.00
	Ties		0(i)		
	Total		5		

a Nachher < Vorher

d Später < Vorher

g Später < Nachher

b Nachher > Vorher

e Später > Vorher

h Später > Nachher

c Nachher = Vorher

f Später = Vorher

i Später = Nachher

Ansonsten gibt es natürlich auch noch den asymptotischen Test:

Test Statistics(c) – Wilcoxon Test

	Nachher - Vorher	Später - Vorher	Später - Nachher
Z	-2.041(a)	-.707(b)	-2.060(b)
Asymp. Sig. (2-tailed)	.041	.480	.039

a Based on negative ranks.

b Based on positive ranks.

c Wilcoxon Signed Ranks Test

Der **Vorzeichentest** als Binomialtest mit $p = .5$ wertet aus, wie viele Wertedifferenzen über 0 liegen (Treffer). Bei kleinen Stichprobenumfängen wird die Testwahrscheinlichkeit direkt aus der Binomialverteilung bestimmt, bei größeren über die Normalverteilung genähert.

Frequencies

		N
Nachher - Vorher	Negative Differenc- es(a,b,c)	0
	Positive Differenc- es(d,e,f)	5
	Ties(g,h,i)	0
	Total	5
Später - Vorher	Negative Differenc- es(a,b,c)	3
	Positive Differenc- es(d,e,f)	2
	Ties(g,h,i)	0
	Total	5
Später - Nachher	Negative Differenc- es(a,b,c)	5
	Positive Differenc- es(d,e,f)	0
	Ties(g,h,i)	0
	Total	5

Test Statistics(b)

	Nachher - Vorher	Später - Vorher	Später - Nachher
Exact Sig. (2-tailed)	.063(a)	1.000(a)	.063(a)

a Binomial distribution used.

b Sign Test

3.5.2. Kontraste

Eigentlich interessieren im Normalfall nicht alle mögliche Einzelvergleiche mit adjustiertem Alpha-Niveau, sondern die zugrunde liegende Fragestellung sollte mit geplanten Kontrasten zu beantworten sein. Die Kodierung von Kontrasten in MWH-Designs wird hier genauer betrachtet.

SPSS bietet zwar einige Kontrasttypen an (siehe 2.4.3.3), die jedoch nicht immer der Fragestellung angemessen sind. Einfache Vergleiche (Mittelwert 1 gegen Mittelwert 2) lassen sich dabei theoretisch auch mit t-Tests berechnen. Dieses Tests liefern unter Berücksichtigung von $F = t^2$ exakt dieselben Ergebnisse wie die entsprechenden Kontraste.

Wenn es etwas komplexer wird bleibt allerdings häufig nur noch der Ausweg der [SPSS-Syntax](#). Die Syntax wird entweder im Output angezeigt (einstellbar unter EditViewer) oder lässt sich direkt in den Syntax-Editor einfügen. Syntax lässt sich im Syntax-Editor über das Menü Ausführen auf den gerade aktiven Datensatz anwenden.

Eine Beschreibung der Funktionen findet sich auch in der Hilfe unter Command Syntax Reference.

Im Folgenden soll das Kontrastbeispiel: „Person bleibt nach dem Ärgern sauer“ überprüft werden, also der [Kontrastvektor \(1 | -0.5 | -0.5 \)](#). Zwar wäre eine solche Prüfung auch über den SPSS-Kontrast Helmert möglich, aber andere Fälle (etwa der [Vektor \(-1|2|-1\)](#) – der Ärger kommt schnell und verfliegt auch schnell wieder) verlangen zwingend nach der Syntax.

Das relevante Unterkommando heißt

`/mmatrix=`

Damit lassen sich Kontrastvektoren in relativ natürlicher Weise eingeben, und zwar nach dem Schema: „Name des Kontrasts“ Var_1 Gewicht_1 Var_2 Gewicht_2 usw. Mehrere Kontrastvektoren werden mit einem Semikolon getrennt. Der Output findet sich schließlich unter „Benutzerdefinierte Hypothesentests“. Es werden t-Tests und F-Tests berechnet:

t-Test: Contrast Results (K Matrix) –

Cont- rast(a)		Transformed Variable	
		Ärger hält an	Ärger nur direkt danach
L1	Contrast Estimate	-2.000	9.200
	Hypothesized Value	0	0
	Difference (Estimate - Hypothesized)	-2.000	9.200
	Std. Error	.524	1.020
	Sig.	.019	.001
	95% Confidence Interval for Difference	Lower Bound Upper Bound	3.456 -544
			6.369 12.031

a Estimable Function for Intercept

F-Test: Univariate Test Results

Source	Transformed Variable	Sum of Squares	df	Mean Square	F	Sig.
Contrast	Ärger hält an	20.000	1	20.000	14.545	.019
	Ärger nur direkt danach	423.200	1	423.200	81.385	.001
Error	Ärger hält an	5.500	4	1.375		
	Ärger nur direkt danach	20.800	4	5.200		

3.5.2.1. L- und MMATRIX

Kontraste auf einem MWH-Faktor werden über die [M-Matrix](#) kodiert. Kontraste auf einem Between-Faktor hingegen werden über die sogenannte [L-Matrix](#) eingegeben (`/LMATRIX=`).

Auch die vordefinierten Kontraste in SPSS verwenden bestimmte L- und M-Matrizen. Im Output steht dabei beispielsweise `/WSFACTOR= Ärgerwerte_Zeit 3 Helmert` – was bedeutet, dass ein MWH-Faktor mit 3 Stufen im Design steckt und dieser mit einem Helmert-Kontrast geprüft werden soll.

Nach dem Kommando `/WSFACTOR` lassen sich jedoch auch eigene Kontrastmatrizen definieren, indem der Name des Kontrasts durch [Special \(\)](#) ersetzt wird. In der Klammer wird der entsprechende Kontrast definiert (was auf F 49 genauer erklärt wird).

3.5.2.2. Spezialfall: Polynomialer Kontrast

Bisher wurde der standardmäßig eingestellte polynomiale Kontrast als uninterpretierbar deklariert und ausgelassen. Es ist allerdings möglich, Trendanalysen mit ungleichen Zeitabständen zu rechnen (Command Syntax Reference).

Dazu erweitert man die Syntax beispielsweise um `/WSFACTOR= Ärgerwerte_Zeit 3 POLYNOMIAL(1,10,70)` um zu signalisieren, dass die Differenz zwischen Zeitpunkt 1 und Zeitpunkt 2 10 Minuten beträgt und dass zwischen Zeitpunkt 2 und 3 60 Minuten liegen.

3.5.3. Split-Plot-Analyse

Beispiel ist wie gehabt die Prüfung verschiedener Rehamaßnahmen (Between-Faktor: Nur Reha vs. Reha + Standardpsychotherapie vs. Reha + Novum) zu verschiedenen Zeitpunkten (Within-Faktor: Vorher, Nachher, Später).

Hinsichtlich [der multivariaten Tests](#) sollten wie gehabt Pillai-Spur oder Wilks-Lambda verwendet werden, wobei die Pillai-Spur insbesondere bei ungleichen Zellbesetzungen zu empfehlen ist. Sonst: Wilks Lambda.

[Mauchly's Test auf Sphärität](#) und die Ergebnisse der [univariaten Tests](#) sind exakt so wie bei der bisher besprochenen MWH-ANOVA. Auch die [Tests der Innersubjektkontraste](#) sind so zweckfrei wie gehabt, es sei denn sie werden entsprechend über die Syntax umkodiert.

Die einzige Änderung ergibt sich für die [Tests der Zwischensubjekteffekte](#). Hier wird jetzt der Between-Faktor aufgelistet und geprüft.

Bzgl. der [Plots](#) ist wieder der Wertebereich zu beachten sowie die vorgespielte Gleichabständigkeit der Zeitpunkte.

3.5.3.1. Traditionelle ANOVA-Tabelle

Das Zusammensetzen der traditionellen ANOVA-Tabelle (wie im Bortz und anderen Büchern) ist etwas mühsam. Die beiden Hauptbestandteile der QS_{Total} lassen sich über die Tabellen zu Innersubjekt- und Zwischensubjekteffekten rekonstruieren ($F64+$).

Hier werden einfach die Quadratsummen und Freiheitsgrade der entsprechenden Faktoren aufsummiert und in die Tabelle eingetragen. Bei den Zwischensubjekteffekten ist nur der Faktor Treatment sowie der Fehler interessant. Die Prüfung des Konstanten Terms (Intercept) ist wie bereits angemerkt ziemlich zweckfrei.

3.5.3.2. Doppelte MWH-Designs

Bisher war immer ein Faktor between, der andere within. Für die Auswertung mit 2 MWH-Faktoren müssen jeweils andere Prüfvarianzen im Nenner stehen – aber das wird getrost SPSS überlassen.

4. Multiple Regression

Bisher wurden Verfahren betrachtet, bei denen eine (oder mehrere) AV und der Einfluss mehrerer kategorialer UVn betrachtet wurde. Nun soll mit der multiplen Regressionsanalyse ein Verfahren beschrieben werden, mit dem der Einfluss von mehreren nicht manipulierten, stetigen UVn analysiert werden kann.

Dabei wird die AV häufig als **Kriterium**, die UVn als **Prädiktoren** bezeichnet. Die für das Kriterium vorhergesagten Werte werden auch als **Schätzwerte**, die Differenz zwischen Schätzwerten und dazugehörigen Ausprägungen des Kriteriums als **Residuen** bezeichnet.

4.1. Modell

Das Datenmodell der multiplen Regression sieht dabei eine (gewichtete) Linear-kombination der Prädiktoren vor:

$$y = a + b_1 x_{m1} + b_2 x_{m2} + \dots + b_k x_{mk} + \varepsilon_m,$$

wobei jede der n Versuchspersonen durch einen entsprechenden Merkmalsvektor beschrieben werden kann. Der Fehlerterm ε wird dabei als additiv, normalverteilt und unabhängig angenommen.

Die Spezifikation der Modellgleichung (nach Datenerhebung; spezifiziert werden also a und b_k) erfolgt dabei so, dass die **Residuenquadratsumme** (Abweichungs-QS) minimal wird. Es liegt demnach eine KQ-Schätzung vor.

Üblicherweise werden die Originalvariablen für diese Schätzung z-transformiert, sodass der Achsenabschnittsparameter a entfällt. Die Regressionsgewichte bei standardisierten Variablen werden dabei normalerweise mit β bezeichnet:

$$\sum (z(y) - z(\hat{y}))^2 = \sum (z(y) - (\beta_1 z(x_1) + \beta_2 z(x_2) + \dots + \beta_k z(x_k)))^2 \rightarrow \min$$

Da hier ein Minimum der Funktion in Abhängigkeit mehrerer Parameter gesucht wird, muss diese partiell abgeleitet (und nullgesetzt) werden. Dadurch erhält man p Gleichungen für p unbekannte Parameter (siehe F11-13).

4.1.1. Roh- und Standardwert-Regressionsgleichung

Die Standardwert-Regressionsgleichung hat den großen Vorteil, dass alle Variablen die gleiche SD aufweisen (1). Die Einflussstärke verschiedener Prädiktoren ist also direkt vergleichbar.

Allerdings lässt sich die Standardwert-Regressionsgleichung wieder in eine Rohwert-Regressionsgleichung überführen (β heißen dann wieder b). Hier ist zwar kein direkter Vergleich der Prädiktoren mehr möglich, allerdings lässt sich aus den Testergebnissen verschiedener Prädiktoren ohne Umwege ein Kriteriumswert vorhersagen.

4.1.2. Interpretation von Koeffizienten

Die Interpretation der einzelnen Gewichtungsfaktoren hat noch zwei Haken: Die Werte sind nur dann eindeutig interpretierbar, wenn die Prädiktoren nicht miteinander korrelieren und beim Schätzen wird auch immer ein Teil der Feh-

lervarianz mit betrachtet (Overfitting). Selbst wenn die Gleichung in Wahrheit gilt ergeben sich bei mehreren Schätzungen also zwangsläufig leicht verschiedene Ergebnisse.

4.1.2.1. Unabhängigkeit der Prädiktoren

In der linearen Einfachregression gibt das standardisierte Regressionsgewicht die Korrelation zwischen Kriterium und Faktor an. Bei der multiplen Regression hingegen zeigen die β -Gewichte den spezifischen Zusammenhang zwischen einem Prädiktor und dem Kriterium an, der um den Einfluss der anderen Prädiktoren bereinigt ist.

Ob ein Prädiktor ein hohes Gewicht hat, hängt hier also neben dem Zusammenhang von Prädiktor und Kriterium auch von den sonstigen Prädiktoren in der Gleichung ab.

Für eine vernünftige Bewertung der Prädiktoren braucht man also beide Informationen: Den inkrementellen Beitrag aus der Regressionsgleichung und den alleinstehenden Zusammenhang der bivariaten PM-Korrelation.

4.1.2.2. Overfitting der Fehlerterms

In der Schätzung der Modellparameter wird, wie oben erwähnt, auch der Fehlerterm angepasst (Overfitting). Dieser Verzerrung kann mithilfe einer Kreuzvalidierung begegnet werden.

4.1.3. Beurteilung der Schätzung: R^2

Die Korrelation der wahren Werte mit den auf Basis der Modellgleichung geschätzten Werten wird als **multiple Korrelation** bezeichnet:

$$R = r(y, \hat{y}),$$

wobei die multiple Korrelation meist größer ist als die größte Einzelkorrelation (naheliegend).

Das Quadrat der multiplen Korrelation, der **multiple Determinationskoeffizient R^2** , gibt an, welcher Anteil der Varianz des Kriteriums durch die Schätzgleichung erklärt wird:

$$R^2 = \frac{1/n \cdot \sum (\hat{y} - \bar{y})^2}{1/n \cdot \sum (y - \bar{y})^2} = \frac{SS_{\text{Regress}}}{SS_{\text{Total}}}$$

Allerdings ist R^2 kein erwartungstreuer Schätzer der wahren multiplen Korrelation, sondern es überschätzt diese systematisch. Daher existieren verschiedene Korrekturformeln, die von SPSS als „**corrected R^2** “ oder „**adjusted R^2** “ ausgegeben werden.

4.2. Prüfung der Schätzung

Bei der einfaktoriellen ANOVA wurde die Quadratsumme in zwei Teile gegliedert:

$$SS_{\text{Total}} = SS_{\text{Between}} + SS_{\text{Within}}$$

wobei SS_{Between} die systematische Varianz (PV), SS_{Within} die Restvarianz (FV) als Streuung der Werte um die Stichprobenmittelwerte darstellt.

Ähnlich lässt sich die Varianz der Kriteriumswerte zerlegen in Regressionsvarianz und Residualvarianz:

$$SS_{\text{Total}} = SS_{\text{Regress}} + SS_{\text{Residual}}$$

Die systematische Varianz geht auf die Regressionsgleichung zurück und kann durch diese berechnet werden, während die Restvarianz das Rauschen der Werte um die Regressionsgerade darstellt (Varianz der Residuen).

Eine Abweichung eines manifesten Kriteriumswertes besteht also aus der Abweichung des Schätzwertes (Vorhersage) vom Grand Mean und der Abweichung des Kriteriums vom Schätzwert.

Daher lässt sich auch für die multiple Korrelation ein F-Test berechnen. Seine H_0 ist immer „Das Kriterium lässt sich nicht aus den Prädiktoren vorhersagen. Jegliche Abweichungen des R^2 -Werts von Null sind Zufall“, während die H_1 besagt, dass ein Zusammenhang zwischen Prädiktoren und Kriterium besteht.

$$F = \frac{SS_{\text{Regress}} / k}{SS_{\text{Residual}} / (n - k - 1)} \quad \text{bzw.}$$
$$F = \frac{R^2 / k}{(1 - R^2) / (n - k - 1)}$$

Ähnlich wie bei einer herkömmlichen ANOVA werden auch die Ergebnisse des F-Tests der multiplen Regression in einer [ANOVA-Tabelle](#) angegeben.

4.2.1. Test der Koeffizienten

Zusätzlich lässt sich überprüfen, ob die einzelnen Prädiktoren einen bedeutsamen Beitrag zur Vorhersage leisten. Die „Signifikanz“ eines einzelnen Prädiktors ist dabei nichts anderes als die statistische Absicherung, dass dieser von Null verschieden ist:

$$t = \frac{\beta_i - 0}{SE_i} \quad \text{mit } SE_i = \text{Standardfehler des Regressionkoeffizienten.}$$

Theoretisch kann der Wert auch gegen eine andere Schranke getestet werden, was allerdings in der Praxis selten gemacht wird.

4.2.2. SPSS-Outputs

...finden sich ab F34.

Anmerkung: Multiple Regressionsanalysen könnten theoretisch auch über die Analysefunktion „Regression“ in Excel gerechnet werden. Zumindest im Excel XP wurde in der ANOVA-Tabelle zur Regression F_{krit} statt p geschrieben, aber p ganz normal angegeben.

4.3. Voraussetzungen

Voraussetzungen der multiplen Regression sind:

- Normalverteilte Fehler: $\epsilon \sim N(0; \sigma)$
- Varianzhomogenität (auf allen Prädiktorwerten)
- Prädiktoren sind selbst keine Zufallsvariablen sondern werden exakt gemessen
- Alle relevanten Prädiktoren wurden berücksichtigt

Aus mathematischer Sicht ist es dabei wünschenswert, dass die Prädiktoren nicht korreliert sind, dies ist aber keine Voraussetzung der Regressionsanalyse. Sind die Prädiktoren jedoch korreliert, so ist die Analyse empfindlich gegen die Reihenfolge der Tests (s.u.).

Eine ähnliche Voraussetzung ist die der **Multikollinearität**: Jeder Prädiktor muss zumindest einen kleinen eigenen Beitrag leisten und darf sich nicht perfekt aus den anderen vorhersagen lassen. Also: Die Prädiktoren dürfen nicht linear abhängig sein, was sich über k neue Regressionen prüfen lässt. Ein Kriterium dabei ist die **Toleranz (1-R²)** in den Einzelregressionen.

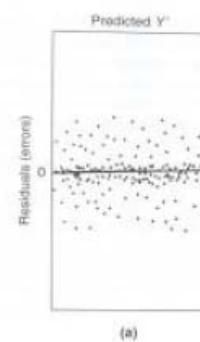
Viel wichtiger hingegen ist eine entsprechende **Fallzahl**. Man benötigt immer mehr Fälle als Variablen, um bedeutungsvolle Aussagen treffen zu können. Für die statistische Absicherung der Regressionsgleichung werden dabei mindestens $N \geq 50 + 8k$ Probanden, für die Testung der einzelnen Regressionskoeffizienten mindestens $N \geq 104 + k$ Probanden benötigt. Also: Unter $N = 100$ braucht man sowieso nicht anzufangen.

Anmerkung: Für mehr Power, bei kleineren Effektstärken oder schießen Kriteriendistributions werden mehr Probanden benötigt. Bei einer Fishing-Excursion mit schrittweiser Regression wird sogar ein VP/Variablen-Verhältnis von 40:1 vorgeschlagen (vgl. 4.3.2).

4.3.1. Prüfung der Voraussetzungen: Residuenplots

Zwar existieren auch statistische Tests, allerdings erfolgt die Prüfung der Voraussetzungen (normalverteilte Fehler, Varianzhomogenität, Prädiktoren exakt gemessen und alle relevanten Prädiktoren berücksichtigt) normalerweise über **Residuenplots**.

Dabei werden die Schätzwerte aus der Regressionsgleichung auf der x-Achse gegen die jeweiligen Residuen auf der y-Achse angetragen. Da die Fehlerterme unabhängig von den Prädiktoren bzw. dem Wert des Kriteriums normalverteilt sein sollten mit $m = 0$ und konstanter Varianz, erwartet man im Plot eine symmetrische Verteilung mit einer Häufung der Punkte um die Nulllinie.



Ist die Verteilung nicht symmetrisch (wie im Beispielbild) sondern schief, so ist eine entsprechende Transformation des Kriteriums angebracht. In anderen Plots könnte ein deutlicher funktioneller Zusammenhang zu erkennen sein (etwa quadratisch), der darauf hindeutet, dass die Modellgleichung nicht korrekt ist. Hier kann eine Transformation der Prädiktoren helfen. Welcher Prädiktor hier in Frage kommt lässt sich prüfen, indem man die Prädiktorwerte gegen die jeweiligen Residuen anträgt.

In wieder einem anderen Plot könnte die Fehlerstreuung mit der Größe der Schätzwerte steigen, also die Homoskedastizitätsannahme verletzt sein. Dies ist insbesondere dann kritisch, wenn sich Minimal- und Maximalvarianz um mehr als den Faktor 3 unterscheiden. Hier hilft evtl. auch wieder eine Transformation von Prädiktoren oder aber die Verteilung kann ein Hinweis darauf sein, dass ein nicht enthaltener Prädiktor mit den enthaltenen interagiert.

4.3.2. Optimaler Variablenatz

Wie unter den allgemeinen Voraussetzungen bemerkt sollte eine Regressionsgleichung alle relevanten Variablen enthalten, sowohl aus statistischen Gründen (Unabhängigkeit der Fehlerterme) und praktischen Gründen (gute Vorhersage).

Allerdings sollten auch so wenig Variablen wie möglich enthalten sein. Auch dies hat mehrere Gründe: statistische (Gefahr der Multikollinearität), wissenschaftstheoretische (Ockhams Skalpell) und praktische (Variablen kosten Zeit und Geld).

Es bietet sich also an, einige Gedanken darauf zu verwenden, einen optimalen Variablenatz zu finden. Bisher wurden dabei immer alle Prädiktoren in die Vorhersagegleichung aufgenommen (Enter) und nachträglich auf Signifikanz geprüft. Werden jetzt nicht signifikante Prädiktoren entfernt, so ändern sich zwangsläufig alle β -Gewichte, da in der Gleichung der spezifische Beitrag einer Variable beschrieben wird.

Geschickter ist es deswegen, die Variablen Stück für Stück in das Modell aufzunehmen. Dabei lassen sich zwei Vorgehensweisen unterscheiden: Theoretische und statistische Selektion. Bei der **theoretischen Selektion** hat der VL (hoffentlich) gute Gründe, die Variablen in einer bestimmten Reihenfolge zu untersuchen, seien es theoretische⁶ oder praktische Gründe (Geld). Bei der **statistischen Selektion** sucht der VL einfach nur nach einer möglichst guten Gleichung, ohne theoretische Restriktionen anzuwenden (**fishung excursion**).

4.3.2.1. Allgemeines Selektionsprinzip

Bei beiden Vorgehensweisen werden so lange Variablen in die Gleichung eingefügt, bis sich die Vorhersage nicht mehr signifikant ändert (Fall 1: Der VL bestimmt die Reihenfolge, Fall 2: rein statistische Kriterien).

$$F = \frac{(R^2_{\text{mehr Variablen}} - R^2_{\text{weniger Variablen}}) / (k_{\text{mehr}} - k_{\text{weniger}})}{(1 - R^2_{\text{mehr Variablen}}) / (n - k_{\text{mehr}} - 1)}$$

Eine feste Reihenfolge wird dabei als sequentielle oder hierarchische Regression bezeichnet, wobei die Sequenz/Hierarchie der Prädiktoren wie erwähnt vom VL festgelegt wird. Die Realisation dieser Sequenz in SPSS wird in der Übung betrachtet.

Das Vorgehen der statistischen Regression wird auch als schrittweise (stepwise) Regression bezeichnet, wobei diese streng genommen ein Spezialfall ist. Details siehe unten.

⁶ Z.B. in der Pädagogischen Psychologie: Dort wird bei der Vorhersage von Schulleistungen immer der IQ zuerst aufgenommen. Dessen Einfluss wird dabei leicht überschätzt, aber dafür sind alle anderen Variablen sicher um den Einfluss des IQs bereinigt.

4.3.2.2. Automatische Variablenelektion

SPSS bietet 4 Mechanismen an:

- **Einschluss:** Langweiliges Standardvorgehen.
- **Rückwärts:** Alle Variablen werden aufgenommen und dann so lange wieder gestrichen, bis die Vorhersage deutlich schlechter wird.
- **Vorwärts:** Variablen werden so lange eingebaut, bis sich die Vorhersage nicht mehr verbessert.
- **Schrittweise:** Rein statistische Suche nach der „besten“ Variablenmenge durch geschicktes Ausprobieren.

Bei den Methoden Vorwärts und Rückwärts bleibt ein einmal aufgenommener Prädiktor auf jeden Fall im Modell enthalten, während bei der Methode Schrittweise Variablen später wieder rausfliegen können. Dieses Verfahren ist daher extrem empfindlich gegen die Reihenfolge.

Bei der Vorwärts-Selektion werden dabei liberale Einschlusskriterien empfohlen ($\alpha = 0.15-0.20$), um zu vermeiden, dass relevante Variablen nicht in die Gleichung aufgenommen werden. Stepwise-Regressionen dagegen sind sehr empfindlich gegen Zufallsfehler (Overfitting), sodass eine Kreuzvalidierung empfohlen wird (normalerweise an Teilstichproben).

Die einzelnen Verfahren führen dabei zu **unterschiedlichen Regressionsgleichungen**, da die Anteile gemeinsamer Varianz auf der Prädiktorseite unterschiedlich behandelt werden. Bei Einschluss werden nur die Unique-Beiträge berücksichtigt, während bei Vorwärts- und Rückwärts-Selektion gemeinsame Varianzanteile immer dem erstbesten Prädiktor zugeschrieben werden. Bei schrittweiser Regression wird die Reihenfolge nach der Höhe der Einzelkorrelation festgelegt (sonst wie Vorwärts). Details siehe F62+.

4.3.2.3. Schlussfolgerung

Eine Regression kann man nur verstehen und bewerten, wenn

- die standardisierten Koeffizienten
- die Einzelkorrelationen der Prädiktoren mit dem Kriterium
- und das gewählte Selektionsverfahren

bekannt sind.

4.4. Ausblick

In der Regressionsanalyse können auch Interaktionseffekte kodiert werden. Dabei entspricht der Interaktionsterm dem Produkt der Einzelvariablen. Allerdings ist das nur bei wenigen Variablen sinnvoll möglich. Auch sind nichtlineare Modellgleichungen denkbar, was etwa bei Trendtests verwendet wird.

Schließlich lässt sich die bisher betrachtete Regressionsanalyse auf kategoriale Prädiktoren (ANOVA), kategoriale und stetige Prädiktoren (ANCOVA) und kategoriale Kriterien (logistische Regression und loglineare Modelle) erweitern. Auch lässt sich eine Erweiterung auf mehrere Kriterien formulieren (Kanonische Korrelation, Strukturgleichungsmodelle).

4.5. Übung IV: Multiple Regression

4.5.1. Excel

Auf den Folien 4+ ist eine multiple Regression über die entsprechende Analysefunktion „Regression“ beschrieben. Achtung: Hier steht in der ANOVA-Tabelle F_{krit} statt p , angegeben wird aber natürlich p .

Achtung 2: In der SPSS-Tabelle des Tests der einzelnen Koeffizienten stehen sowohl Roh-Wertkoeffizienten (B) als auch standardisierte Koeffizienten (Beta). Excel hingegen berechnet für die Einzeltests nur die Rohkoeffizienten.

4.5.2. SPSS-Einstellungen

4.5.2.1. Einzelkorrelationen

Beta-Gewichte geben die Stärke des Einflusses eines Prädiktors in Abhängigkeit der Zusammensetzung der Prädiktorenmenge aus. Daneben sollten jedoch auch die Einzelkorrelationen der Prädiktoren mit dem Kriterium in die Bewertung mit einbezogen werden.

Das geht in SPSS erstaunlicherweise über das Kommando „Korrelationen“.

4.5.2.2. Regressionsgleichung

Aus der Tabelle der Koeffizienten lässt sich auch die Regressionsgleichung ableiten. Hier ein Beispiel für die Rohwert-Regression:

Modell		Koeffizienten ^a				
		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	-12.794	17.750		-.721	.475
	x1	.007	2.223	.000	.003	.997
	x2	2.333	1.574	.205	1.482	.145
	x3	-3.891	1.732	-.315	-2.246	.030
	x4	1.285	1.869	.095	.688	.495

a. Abhängige Variable: komplexesY

Aus der markierten Spalte lässt sich folgende Gleichung ableiten:

$$\hat{y}_m = -12.794 + 0.007x_{m1} + 2.333x_{m2} - 3.891x_{m3} + 1.285x_{m4}$$

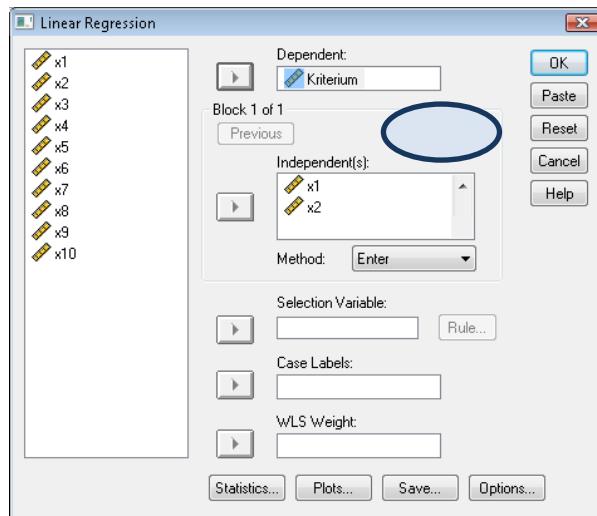
4.5.2.3. Prüfung der Voraussetzungen

Wie unter 4.3.1 beschrieben, wird hierfür am geschicktesten ein Plot erzeugt, in dem Schätzwerte (ZPRED) gegen Residuen (ZRESID) angetragen werden.

Zudem kann SPSS Einzelplots zu dem Zusammenhang zwischen Schätzung und Kriterium ausgeben. Hier können Ausreißerwerte identifiziert werden, deren Eliminierung zwar meist keinen großen Einfluss auf die Güte der gesamten Schätzung hat. Aber hierdurch werden die Gewichte der einzelnen Prädiktoren teilweise deutlich verändert.

4.5.2.4. Schrittweise Regression

Unter 4.3.2.1 wurden die in SPSS verfügbaren Standardverfahren beschrieben. Es ist jedoch wie erwähnt auch möglich, die Reihenfolge der Aufnahme in das Modell festzulegen. Das geht über den markierten Button.



Bei einer sequentiellen Regression kann man noch prüfen, ob die Hinzunahme der Variablen die Vorhersage signifikant verbessert. Die hierfür nötigen Kennwerte stehen in der Tabelle *Modellzusammenfassung* und können über den erwähnten F-Bruch geprüft werden.

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	.263 ^a	.069	.066	108.95117
2	.443 ^b	.196	.191	101.38374
3	.459 ^c	.211	.204	100.59804

a. Einflußvariablen : (Konstante), x3, x1, x2

b. Einflußvariablen : (Konstante), x3, x1, x2, x4, x6, x5

c. Einflußvariablen : (Konstante), x3, x1, x2, x4, x6, x5, x7, x8, x9

Beispiel für den Vergleich von Modell 2 und 3: Aus (b) und (c) geht hervor, dass in Modell 2 6, in Modell 3 9 Variablen enthalten sind. Damit ergibt sich:

$$F = \frac{(0.211 - 0.196) / (9 - 6)}{(1 - 0.211) / (1000 - 9 - 1)} = 6.27$$

Diesen Wert kann man in Excel über die Funktion FVERT prüfen. In SPSS ist dies mit etwas Umstand auch möglich, indem man die entsprechende Formel unter *Transformieren\Berechnen* eingibt und sich den Wert in eine neue Variable eingeben lässt. Die Formel lautet hier $1 - \text{CDF-F}(F\text{-Bruch}; df_{\text{Zähler}}, df_{\text{Nenner}})$.

4.5.3. Nichtlineare Modelle

Werden ab F48 behandelt. Im Prinzip wird die vermutete Gleichung aufgestellt, berechnet und ergibt eine neue Schätzgleichung mit veränderten Gewichten.

Ob ein nichtlineares Modell eine signifikant bessere Vorhersage erlaubt, lässt sich wieder mit dem bekannten F-Wert aus 4.3.2.1 berechnen.

5. ANCOVA, MANOVA und MANCOVA

In diesem Kapitel sollen Kovarianzanalyse (ANCOVA), multivariate Varianzanalyse (MANOVA) und multivariate Kovarianzanalyse (MANCOVA) kurz angerissen werden.

5.1. Kovarianzanalyse (ANCOVA)

Die Kovarianzanalyse (ANCOVA) bietet sich für Datensituationen an, bei denen der Einfluss mehrerer kategorialer UVn auf eine metrische AV untersucht *und um den Einfluss metrischer Kontrollvariablen bereinigt werden soll*.

Dabei ist das Vorgehen relativ einfach und erfolgt in zwei Schritten:

- 1.) Zunächst erfolgt [eine lineare Regression](#) mit den Kontrollvariablen als Prädiktoren und der UV als Kriterium. Als Ergebnis dieser Rechnung werden die Residuen gespeichert.
- 2.) Dann wird eine [ANOVA mit den Residuen](#) aus dem ersten Schritt als AV durchgeführt.

Da durch die ANCOVA die Fehlervarianz verringert wird (Streuung innerhalb der Gruppen wird verringert), wird das Ergebnis leichter signifikant. Dabei reduziert das Herauspartialisieren nicht nur die Fehlervarianz sondern kostet auch einen Freiheitsgrad.

5.1.1. Annahmen

Neben den Annahmen der ANOVA (Normalverteilung, Varianzhomogenität) kommt hinzu, dass die Regressionsgerade für die Beziehung zwischen UV und Kovariate für alle Gruppen dieselbe Steigung haben muss. Ist dies nicht der Fall, liegt eine Interaktion zwischen Kovariate und unabhängiger Variable vor.

Zur Überprüfung dieser Annahme kann eine sequentielle ANOVA verwendet werden, bei der zunächst Kovariate, dann Gruppenvariable und schließlich in einem dritten Schritt die Interaktion Kovariate*Gruppe eingefügt werden. Das wird in der Übung nochmal genauer gemacht.

5.1.2. Weitere Anwendungen

Neben der Auspartialisierung von Kovariaten (→ leichtere Signifikanz), wird die ANCOVA auch für Post-Hoc-Tests im Rahmen einer multivariaten ANOVA (MANOVA) verwendet.

5.2. Multivariate Varianzanalyse (MANOVA)

Die MANOVA stellt eine Erweiterung der ANOVA auf mehrere metrische abhängige Variablen dar. Hier sind die unabhängigen Variablen weiterhin kategorial.

Statt einem einzigen Messwert hat man es jetzt also mit [Messwert- und Parametervektoren](#) zu tun.

$$\vec{y}_{mij} = \vec{\mu} + \vec{\alpha}_i + \vec{\beta}_j + \vec{\alpha\beta}_{ij} + \vec{\varepsilon}_{mij}$$

An den Verfahren ändert sich sonst relativ wenig, nur dass die Varianzschätzer nun für Varianz-Kovarianz-Matrizen berechnet werden.

Der multivariate Test betrachtet also den Unterschied (Abstand) der beiden Zentroiden (bzw. Mittelwertsvektoren). Die Frage ist also nun, ob der Unterschied in einer Linearkombination der verschiedenen Vektoren für die einzelnen AVn so groß ist, dass die H_0 verworfen werden kann.

5.2.1. Interpretation der Ergebnisse

Wie üblich, sagt ein signifikantes Ergebnis zunächst nur aus, dass irgendwo ein Unterschied vorliegt, aber nicht in welcher Richtung, zwischen welchen Gruppen und auf welchen Variablen.

Üblicherweise werden daher anschließend univariate ANOVAs auf den einzelnen AVn berechnet und ggf. mit Einzelvergleichen geprüft. Den ersten Schritt übernimmt SPSS dabei automatisch.

Allerdings führt man in diesem Fall viele Einzeltests durch (MANOVA → viele ANOVAs), was eine α -Adjustierung notwendig macht. Allerdings sind die AVn im Normalfall korreliert und damit nicht unabhängig...und für diesen Fall gibt es keine Standardadjustierung.

Daher wurde mit der [Roy-Bargmann Stepdown Analysis](#) eine alternative Post-Hoc-Auswertung entwickelt. Die zugrunde liegende Idee entspricht einem sequentiellen/hierarchischen Regressionsmodell:

- 1.) Auf Basis von theoretischen/praktischen Überlegungen werden die [AVn in eine Reihenfolge](#) gebracht.
- 2.) Die wichtigste Variable wird dann mit einer [univariaten ANOVA](#) geprüft (Alpha-Adjustiert mit $\alpha_{neu} = \alpha_{alt} / k$).
- 3.) Anschließend werden die [nächsten Variablen mit ANCOVAs](#) geprüft, wobei die bereits getesteten Variablen als Kovariaten dienen.

Natürlich bleibt bei diesem Vorgehen für die späteren Variablen immer weniger Varianz zum Aufklären übrig. Dafür sind die Tests dann unabhängig von den vorhergehenden Variablen. Da unabhängige Tests vorliegen, können die Alpha-Adjustierungen nach dem bekannten Bonferroni-Schema (α/k) erfolgen.

5.2.2. Bewertung

Gerade bei multivariaten Forschungsfragen ist die MANOVA das natürliche Auswertungsinstrument. Auch kann es in seltenen Fällen vorkommen, dass die durchgeführten univariaten Tests nicht signifikant werden, die multivariaten dagegen schon. Zudem eignen sich MANOVAs zur Analyse von Messwiederholungsdesigns.

Allerdings ist die MANOVA im Allgemeinen weniger teststark als die univariaten Analysen (mit Ausnahme des seltenen Spezialfalls). Zudem können wie bei der multiplen Regression korrelierte Variablen Probleme bei der Interpretation bereiten. Nur handelt es sich hier dann um korrelierte AVn statt UVn. Eine mögliche Lösung bietet die Roy-Bargmann Stepdown Analysis.

5.3. Multivariate Kovarianzanalyse (MANCOVA)

Die MANCOVA ist die direkte Erweiterung der ANCOVA auf den multivariaten Fall. Auch hier gibt es mehrere metrische AVn, kategoriale UVn und metrische KVn (Kontrollvariablen).

Die MANCOVA untersucht also Unterschiede von [Messwertvektoren](#) zwischen Gruppen, wobei die Messwerte um den Einfluss der Kontrollvariablen bereinigt werden. Interpretationen und Folgetests verlaufen dann wie bei MANOVA bzw. univariater ANCOVA.

5.4. Profilanalyse: MANOVA bei Split-Plot

Im Kontext der ANOVAs im Messwiederholungsdesign wurde die Profilanalyse als multivariate Auswertung von MWH-Designs erwähnt (vgl. 3.3).

Im Beispiel aus dem MWH-Kapitel sollte eine neue Rehamaßnahme evaluiert werden. Dabei wurden 3 Gruppen gemessen (Between-Faktor: Kontrolle, Standard, Novum) und jeweils 3 Zeitpunkte analysiert (Within-Faktor: vorher, nachher, nach 6 Wochen). Als AV wurde ein Scherzrating verwendet.

Der Between-Haupteffekt wird getestet durch eine einfaktorielle ANOVA der individuellen Mittelwerte (gemittelt über die Zeitpunkte).

Eigentlich interessiert die Interaktion Gruppe*Zeitpunkt, da sich die Patienten mit neuem Treatment stärker verbessern sollten als diejenigen mit dem Standard-treatment.

Zwar wird die Interaktion auch von SPSS automatisch berechnet, allerdings lässt sich die Vorgehensweise auch einfach illustrieren. Hierfür lassen sich die beiden Differenzen Vorher-Nachher und Nachher-Nach6Wochen als zwei AVn auffassen und in eine einfaktorielle MANOVA stecken.

Dabei wird also folgendes Datenmodell zugrunde gelegt:

$$\vec{d}_{mi} = \vec{\mu} + \vec{\alpha}_i + \vec{\varepsilon}_{mi}$$

Der Vektor der Differenzen (hier: 2-dimensional) entsteht aus einem konstanten Vektor und einem Vektor für gruppenspezifische Abweichungen sowie einen Fehlerterm. Der Faktor α_i beschreibt dabei die Interaktion Gruppe*Zeitpunkt (er enthält die Abweichungen jedes Gruppenvektors von μ), während μ (in SPSS: Konstanter Term) das Mittel aller Gruppenabweichungen enthält: Der Haupteffekt Within!

Da alle vorgestellten Analysen einfaktoriell waren (Between-ANOVA und einfaktorielle MANOVA der Differenzwerte) braucht man die Sphäritätsannahme nicht. Und genau so arbeiten die ominösen multivariaten Tests.

5.5. Übung V: ANCOVA und MANOVA

5.5.1. ANCOVA

Auf den Folien werden Kodierung, Interpretation und Einzelvergleiche beschrieben. Da das allerdings weniger spannend ist, soll hier vor allem die Prüfung der Voraussetzung homogener Regressionsgeraden betrachtet werden.

Die Bedingung „Gleiche Steigungen der Regressionsgeraden in allen Gruppen“ ist wie erwähnt gleichbedeutend mit „Keine Interaktion zwischen Gruppenzugehörigkeit und Kovariate“. Dieser Test entspricht einer sequentiellen ANOVA in die nacheinander Kovariate, Gruppenvariable sowie die Interaktion Kovariate*Gruppenvariable eingefügt werden.

Diese Spezifikation erfolgt in SPSS unter *Model\Custom*. Wichtig ist, dass hierbei **Quadratsummen des Typ I** eingestellt werden (sequentiell), damit korrekt auspartialisiert wird.

Wird der Interaktionseffekt nicht signifikant, kann H_0 beibehalten werden. H_0 heißt in diesem Fall: „Die Steigungen der Regressionsgeraden sind gleich“.

Die Regressionsgeraden lassen sich nun in einem Scatterplot darstellen (unter *Graphs\Legacy Dialogs\Scatter/Dot...* oder über den Chart Builder). Dabei wird die Kovariate auf der x-Achse gegen die AV auf der y-Achse angetragen. Die Gruppenzugehörigkeit wird als Label eingefügt [Anmerkung: Ggf. muss diese Variable als nominalskaliert definiert werden].

Die Regressionsgeraden werden über einen Rechtsklick in das Diagramm eingefügt (*Add Reference Line*). Die Gleichungen muss man sich leider selbst ausrechnen (über *Analyze\Regression*) – Statistica kann das auch so.

5.5.2. MANOVA

SPSS berechnet automatisch multivariate Tests und nachgeschaltete einzelne ANOVAs (Tests der Zwischensubjekteffekte). An sich wenig spektakulär.

Interessanter wird's bei der Stepdown-Analyse nach Roy-Bragmann. Wie unter 5.2.1 beschrieben, werden die AVn zunächst in eine Reihenfolge gebracht. Die „wichtigste“ AV wird dann α -adjustiert mit einer univariaten ANOVA geprüft. Danach folgen ANCOVAs, bei denen gegen die bereits geprüften Effekte kontrolliert wird.

Anmerkung: Stepdown-Analysen lassen sich in SPSS auch über die Syntax kodieren. Das rentiert sich aber nur, wenn man viele solche Designs rechnen muss und wird auf den Folien 29 und 30 beschrieben.

6. Diskriminanzanalyse

Bisher war immer die Frage: Kann man aus der Kenntnis der Stichprobenzugehörigkeit einen Messwert vorhersagen? Die Diskriminanzanalyse geht die entgegengesetzte Richtung und fragt: „Kann man aus der Kenntnis der Messwerte die Stichprobenzugehörigkeit vorhersagen?“. Ein klassisches Beispiel ist die Krankheitsdiagnose in Abhängigkeit vom verwendeten Fragebogeninventar.

Es wird also gefragt, in welche Gruppe eine Versuchsperson mit einem bestimmten Messwert(-profil,-vektor) gehört. Die Diskriminanzanalyse (DKA) dient also der Klassifikation von Objekten anhand von Merkmalen bzw. Merkmalsvektoren. Mathematisch gesehen sind DKA und MANOVA dabei identische Verfahren.

Anmerkung: Bei sehr ungleichen Zellbesetzungen kommt auch die logistische Regression in Frage (die eine dichotome Stichprobenzugehörigkeit vorhersagt). Auch besteht eine große Ähnlichkeit zur Faktorenanalyse (PCA ohne Rotation).

Grenzen der DKA sind mehrfaktorielle Designs (nur in Spezialfällen möglich) und Messwiederholungen (gehen gar nicht). Ein großes Problem sind zudem inhomogene Varianzen: Je größer die Varianz einer Gruppe, desto wahrscheinlicher wird ihr ein Fall zugeordnet!

6.1. Grundidee

In der DKA werden Merkmale (Prädiktoren) linear kombiniert, und zwar so, dass sie optimal zwischen den Gruppenmittelwerten trennen. Diese Linearkombinationen werden auch als **Diskriminanzfunktion** bezeichnet.

Beispiel für 3 Gruppen:



Bei drei Gruppen kann es dabei maximal zwei Diskriminanzfunktionen geben (allgemein: **Anzahl Gruppen – 1**), es kann aber auch sein, dass eine Diskriminanzfunktion ausreicht. [Anmerkung: Eigentlich ist $n_{\max}(\text{DKF}) = \min(n(\text{Gruppen}) - 1, n(\text{Prädiktoren}) - 1)$].

Die Idee der DKA ähnelt also der Hauptkomponentenanalyse als Basisverfahren der Faktorenanalyse. Es wird eine Linearkombination von Variablen (hier: Diskriminanzfunktion, dort: Faktor) gesucht, so dass die Projektionen der Datenpunkte auf diese neue Achse maximale Varianz aufweisen.

Die nächste Linearkombination wird dann orthogonal zur ersten gewählt und klärt wieder die maximale Varianz auf, die noch nicht vom ersten Faktor bzw. der ersten DKF erfasst ist. In der Praxis finden sich daher meist maximal zwei signifikante DKFs.

6.2. Modell

Die Diskriminanzfunktion einer DKA ist einem Faktor der Faktorenanalyse sehr ähnlich. Der Wert einer VP auf einer Diskriminanzfunktion entspricht dabei exakt dem Faktorwert:

$$D_i = d_{i1}z_1 + d_{i2}z_2 + \dots + d_{ip}z_p$$

Dabei kann man die Lage der VP-Werte bzw. der Gruppenzentroide in dem Raum darstellen, der von den Diskriminanzfunktionen aufgespannt wird. Da meist nur 2 DKF signifikant werden, ist das nicht zu kompliziert.

Dabei wird automatisch getestet, ob der Unterschied zwischen den Gruppenzentroiden auf den Achsen signifikant wird. Wie der Test funktioniert wird hier allerdings nicht beschrieben.

Die Diskriminanzfunktionen können dann ähnlich wie die Faktoren der Faktorenanalyse interpretiert werden, indem der Zusammenhang zwischen DKF und den einzelnen Merkmalen betrachtet wird.

6.3. Anwendung und SPSS

Die Diskriminanzanalyse findet sich unter *Analysieren\Klassifizieren\Diskriminanzanalyse*. Wichtig ist vor allem die Output-Tabelle zu standardisierten kanonischen **Diskriminanzkoeffizienten**, die den Gewichten d_i des Modells entsprechen. Aus diesen lässt sich also auch der Wert einer Person auf der Funktion berechnen.

Die Tabelle zur **Struktur-Matrix** enthält sozusagen die Ladungen der Merkmale auf den Dimensionen. Dabei gibt es keinen statistischen Grund, wie groß eine Ladung sein muss, um „groß genug“ zu sein. Konventionen sind $r > .33$ und $r > .5$.

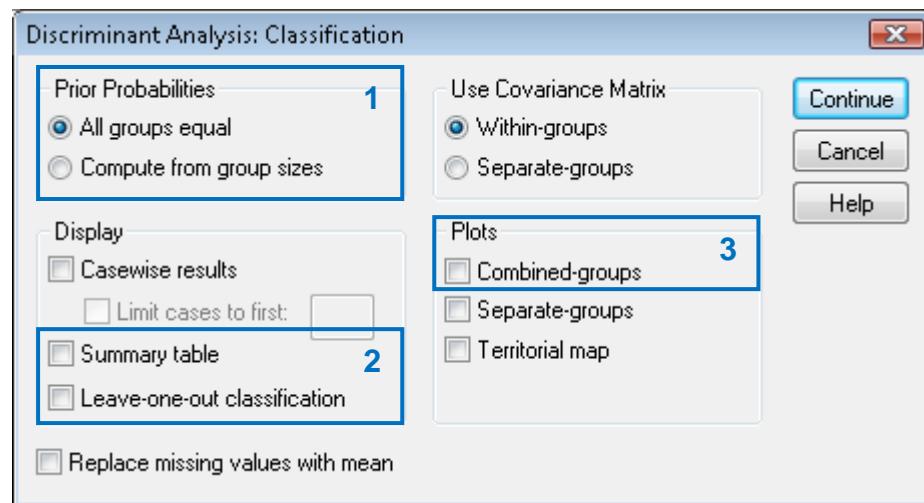
6.3.1. Klassifikation

Allerdings ist die Diskriminanzanalyse kein genuines Verfahren zur Dimensionsreduktion, sondern eigentlich ein Verfahren zur Klassifikation der Personen anhand der Merkmale. Das soll jetzt genauer unter die Lupe genommen werden.

SPSS kann nämlich neben den erwähnten Funktionen auch ausgeben, wie die Personen der Eichstichprobe auf Basis der **Klassifikationsfunktionen** klassifiziert werden würden. Dabei gibt es jedoch zwei Dinge zu beachten:

- Man kann berücksichtigen, dass die **Gruppen unterschiedlich häufig** in der Population auftreten.
- Außerdem erfolgt die Klassifikation auf denselben Daten wie die Schätzung der DKFs, was ein gewisses Overfitting impliziert. Daher müssen die Ergebnisse entweder anhand einer **Kreuzvalidierung (Teilung der Stichprobe oder komplett neue Stp)** oder über ein **Jackknife-Verfahren (SPSS: Klassifikation mit Fallauslassung)** geprüft werden.

Diesen Überlegungen kann in SPSS mit den entsprechenden Optionen des Menüs „Klassifizieren“ Rechnung getragen werden (siehe nächste Seite).



Unter (1) lässt sich einstellen, ob die A-priori-Wahrscheinlichkeit für die Zugehörigkeit zu einer Gruppe aus dem Stichprobenanteil der Gruppe geschätzt werden soll, oder ob alle Gruppen gleich behandelt werden sollen. Unter (2) hingegen lassen sich Kreuztabellen erzeugen, die als Zeile die tatsächliche Gruppenzugehörigkeit und als Spalten die vorhergesagte Gruppenzugehörigkeit enthalten. Auch wird der Anteil korrekter Klassifikation ausgegeben.

Die Option (3) gibt den Plot mit den Diskriminanzfunktionen als Achsen aus. Die obere Hälfte der Kreuztabelle enthält das Ergebnis, wenn man die Daten der Stichprobe mit den geschätzten DKFs klassifiziert. Die untere Hälfte ergibt sich, wenn man einen Fall aus der Stichprobe hinausnimmt, mit den n-1 Fällen die DKFs schätzt und damit den rausgenommenen Fall klassifiziert. Das wird für alle n Fälle gemacht.

6.3.2. Klassifikationsfunktionen

Wenn man die Diskriminanzfunktionen an einer Eichstichprobe ermittelt hat, kann auf Basis von vorhandenen Merkmalsvektoren einer Versuchsperson die jeweilige Gruppenzugehörigkeit geschätzt werden.

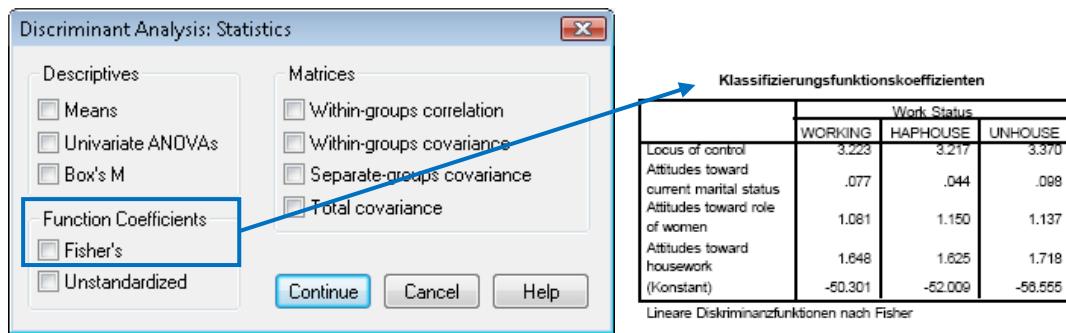
$$C_j = c_{j0} + c_{j1}x_1 + c_{j2}x_2 + \dots + c_{jp}x_p$$

Die Funktionen sind im Prinzip eine multiple Regressionsgleichung. Auf der linken Seite steht ein „Grad der Gruppenzugehörigkeit“ und es gibt eine Gleichung für jede Gruppe j. Eine Person wird dabei der Gruppe zugeordnet, für die ihre **Klassifikationsfunktion** am größten wird.

Klassifikationsfunktionen dienen also ähnlich wie die Rohwert-Regressionsgleichung der praktischen Aufgabe der Klassifikation, während die Diskriminanzfunktionen für theoretische Zwecke besser geeignet sind (Interpretation der zugrundeliegenden Unterschiedlichkeiten).

In SPSS geht das im Menü **Statistik** (siehe nächste Seite). Dort lässt sich der Menüpunkt **Funktionskoeffizient nach Fisher** auswählen (1), welcher im Output die Tabelle „Klassifizierungsfunktionskoeffizienten“ liefert.

Die entsprechende Output-Tabelle enthält die Koeffizienten der Klassifizierungsfunktionen für die einzelnen Gruppen.



6.4. Voraussetzungen und Ausblick

Prinzipiell hat die DKA dieselben Voraussetzungen wie die MANOVA. Ungleiche Zellbesetzungen machen hier insgesamt weniger Probleme, da nur univariate Designs untersucht werden.

Auch unkritisch ist die [multivariate Normalverteilung](#), welche auch nicht testbar ist. Lediglich [Varianzhomogenität](#) könnte bei ungleichen Zellgrößen ein Problem werden. Diese Voraussetzung ist hier deutlich wichtiger als bei der MANOVA, da Varianzhomogenität zu Fehlklassifikationen führt.

Was jetzt noch fehlt: Hierarchische und schrittweise DKA, die etwa hierarchischer und schrittweiser Regression entsprechen, die Berücksichtigung von Kosten (optimale Klassifikation) sowie die detaillierte Interpretation der Prädiktoren durch die Kontrastanalyse. Da hilft nur ein Blick ins Lehrbuch.

7. Fragen und Anmerkungen

Wichtige Folien:

- 01_a_uniANOVA: F35, 38, 42, 44, 62
- 01_b_Übung: F8 (so wird das Modell korrekt hingeschrieben)
- 02_a_2faktANOVA: F67 (bei ungleichen Stpn-Umfängen sind die Quadratsummen mit mehr additiv!)
- 04_Übung: Warum nimmt der F-Bruch bei einer Stepwise-Regression bei einem besseren Modell ab? (Teilaufgabe 1)
- 06_a_DKA: F15 Anzahl Gruppen und Anzahl Prädiktoren beide -1 oder nur Anzahl der Gruppen?