Analyse von Messwiederholungsdaten

Kolloquium Forschungsmethodik

Psychologisches Institut, Uni Mainz

19.6.2008

Dr. Daniel Oberfeld-Twistel

Allgemeine Experimentelle Psychologie

Psychologisches Institut

Johannes Gutenberg – Universität Mainz

- ANOVA
- (Regression)

- alternative Verfahren
- Vor- und Nachteile je nach Datensituation

- Empfehlungen und Rechenbeispiele

Messwiederholungsdaten

- *between subjects/completely randomized*: jede Vp wurde in genau **einer** Bedingung untersucht

- *within-subjects/repeated measures*: Vp wurde in **mehr** als einer Bedingung untersucht

ANOVA: between subjects

- *Verkostung von Wein, 4 Umgebungsfarben, pro Farbe 30 Vpn, jede Vp verkostet einen Wein unter* ***einer*** *Farbe und nennt maximalen Kaufpreis*

  Modell: $Y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$

  - Y_{ij} vorhergesagter Wert der AV für Vp *i* in Bedingung/Gruppe *j*
  - μ Populations-„grand mean"
  - α_j Populationseffekt von Bedingung *j*
  - ε_{ij} Fehlerterm für Vp *i* in Bedingung *j*

- Annahmen:
  - Normalverteilung der AV (*Y*) in der Population
    - **unkritisch**
  - Populationsvarianzen in jeder der *a* Gruppen identisch
    - **unkritisch**, falls Gruppengrößen identisch
    - ansonsten: **kritisch** -> Brown-Forsythe / Welch
  - Y-Werte **unabhängig**
    - **kritisch**

  → ε_{ij} i.i.d. $N(0, \sigma_1 = \sigma_2 = \ldots = \sigma_a = \sigma)$

ANOVA: within-subjects

- *Jede Vp verkostet 4 Weine (einen unter jeder Umgebungsfarbe), nennt maximalen Kaufpreis*

Modell: $Y_{ij} = \mu + \alpha_j + \pi_i + (\pi\alpha)_{ij} + \varepsilon_{ij}$

  - Y_{ij} Wert auf AV für Vp *i* in Bedingung/Gruppe *j*
  - μ Populations-„grand mean"
  - α_j Populationseffekt von Bedingung *j* (**fixed factor**)
  - π_i Effekt von Person *i* (**random factor**)
  - $(\pi\alpha)_{ij}$ Effekt der Interaktion von Bedingung & Person (im Test = 0 gesetzt)
  - ε_{ij} Fehlerterm für Vp *i* in Bedingung *j*

- Also: zwei-faktorielles Design, *Farbe* = fixed factor, *Vp* = random factor (*mixed model*)
- Vorteil ggü. between-subjects Design: Power

Within-subjects: Test

$$F = \frac{SS_A/(a-1)}{SS_{A\times S}/[(n-1)(a-1)]}$$

- Zähler: Varianz durch Bedingungen (*Farbe*)
- Nenner: Ausmaß, in dem der Effekt von Faktor *A* von Person zu Person variiert

ANOVA within-subjects: Annahmen

1. Normalverteilung der AV (*Y*) in der Population
   - **unkritisch**
2. Y-Werte der *verschiedenen* Vpn **unabhängig**
   - **kritisch**, aber normalerweise gewährleistet

3. Homogenität der Populationsvarianzen aller Differenzen zwischen der Bedingungen
   - z.B. $Var(Y_{grün} - Y_{rot}) = Var(Y_{grün}) + Var(Y_{rot}) - 2\, \rho_{\text{Grün-rot}}\, \sigma_{\text{Grün}}\, \sigma_{\text{rot}}$
     - $\rho_{\text{Grün-rot}}$ Populations-Korrelationskoeffizient für Scores unter grünem vs. rotem Licht
   - *Äquivalent*: Populations Varianz-Kovarianzmatrix ist **sphärisch (sphericity)**
   - Annahme **kritisch!!!!** und in der Praxis *fast immer* verletzt

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho & \rho & \rho \\ & 1 & \rho & \rho & \rho \\ & & 1 & \rho & \rho \\ & & & 1 & \rho \\ & & & & 1 \end{bmatrix}$$

Univariater Ansatz mit df-Korrektur

- Lösung:
  - Ausmaß der Abweichung von Spherizität aus den Daten schätzen
    - Wert **ε** (Box, 1954), $\varepsilon \leq 1.0$
  - *F*-Test rechnen mit um ε **reduzierten** Freiheitsgraden $F[\varepsilon\,(a-1), \varepsilon\,(n-1)(a-1)]$
- Zwei Varianten für ε: Greenhouse-Geisser und Huynh-Feldt
  - HF: weniger konservativ

Multivariater Ansatz

- *MBP für Wein unter 3 Farben*
- Univariate Analyse: H_0: alle 3 Populationsmittelwerte gleich
- Äquivalent: Differenz-Scores zwischen jeweils zwei Bedingungen bilden (z.B. $Y_{blau} - Y_{rot}$ und $Y_{blau} - Y_{grün}$)
  - H_0: alle D-Variablen = 0

⇨ **multivariater** Test

  - bei *a* Faktorstufen: (a – 1) D-Variablen = linear unabhängige Kontraste
  - *F*-Test mit (a – 1) und n - (a – 1) *dfs*

- **keine Annahmen** über die Var-Covar-Matrix notwendig!

$$\mathrm{UN} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ & \sigma_2^2 & \sigma_{23} & \sigma_{24} \\ & & \sigma_3^2 & \sigma_{34} \\ & & & \sigma_4^2 \end{bmatrix}$$

Empirical-Bayes Ansatz

- Univariat: Kovarianzstruktur **Sphärizität**
- Multivariat: Kovarianzstruktur **Unstrukturiert**
- Alternative: Empirical-Bayes Ansatz (Boik, 1997)
  - in den Daten vorhandene Evidenz für die beiden alternativen Kovarianzstrukturen → "Kombination" der beiden Schätzer

"Mixed model"/"multilevel" Analysen

- Grundlage: allgemeinere Formulierung des linearen Modells
  - insbesondere: Kovarianzstruktur wird *explizit* modelliert
    - univariater und multivariater Ansatz = Spezialfälle
  - Lit.: Keselman, Algina, & Kowalchuk (2001)
- **Maximum-likelihood** statt least-squares Optimierung
- *Vorteile*:
  - **Falls** man die tatsächliche Kovarianzstruktur kennen würde, hätte man die größte Power…
    - in der Praxis: "beste" Struktur anhand von goodness-of-fit Maßen auswählen (Akaike oder Schwarz information criteria)
  - Modell kann mit missing values umgehen (falls MCAR oder MAR)
- *Nachteil*: viele Optionen, evtl. Probleme bei der Modellanpassung
- **Wichtig**: adjustierten *F*-Test nach Kenward & Roger (1997) verwenden
  - SAS PROC Mixed: `/ddfm=KR`

Between-subjects Faktoren

- bislang: nur **within-subjects** Faktoren im Design
- falls **zusätzlich** between-subjects Faktor(en): “split-plot” design
  - *Jede Vp nennt MBP für Wein unter 4 Farben, aber es gibt mehrere Gruppen mit verschiedenen Reihenfolgen (Latin Squares)*
- **Zusätzliche** Annahme notwendig!
  - *univariater Ansatz*: Var-Covar Matrizen sphärisch **und** für alle Gruppen identisch
  - *multivariater Ansatz*: Var-Covar Matrizen für alle Gruppen identisch
  - ⇨ “**multi-sample sphericity**”
- Simulationsstudien: entscheidend, ob Gruppengrößen **identisch** oder nicht (Keselman, Algina, & Kowalchuk, 2001)
  - balanciertes Design: gleich viele Vpn pro Gruppe = Stufe des between-subjects Faktors
  - unbalanciert: Gruppengrößen nicht identisch

Beispiel Simulationsstudie

- Grundidee: Daten anhand des Modells zufällig aus Normalverteilung generieren, **keine** Mittelwertsunterschiede -> wie viele Tests werden **trotzdem** signifikant?
  - “empirical rate of Type I error”

- Variablen wie Stichprobengröße, Abweichung von Sphärizität, Ungleichheit der Gruppen bei between subjects etc. variieren

Table 4. Empirical Type I error rates (%), **Z** lognormal distributed

	Main effect test					Interaction effect test					
C	IGA	WJ	$\bar{\varepsilon}$	T^2	EB	IGA	WJ	$\bar{\varepsilon}$	HL(EB)	P(EB)	W(EB)
Response variable normally distributed											
$f = 3,\ N = 18,\ \varepsilon = 1.00$											
a	4.90	–	4.68	4.50	4.42	6.22	–	5.10	5.56	3.98	4.64
b	5.38	–	5.48	6.18	5.52	6.72	–	6.44	**7.94**	5.10	6.68
c	5.14	–	3.04	2.70	2.64	5.86	–	3.42	3.54	**1.86**	2.82
d	6.72	–	**10.14**	**14.08**	**13.38**	**7.64**	**–**	**10.58**	**15.76**	**12.16**	**14.46**
c′	5.18	–	**1.40**	**1.22**	**1.08**	6.16	–	**2.46**	**1.80**	**1.08**	**1.44**
d′	**8.96**	**–**	**14.22**	**19.54**	**19.06**	**9.68**	**–**	**15.12**	**21.68**	**17.52**	**20.60**
$f = 6,\ N = 45,\ \varepsilon = 0.40$											
a	5.14	4.48	5.12	4.66	4.66	6.38	5.54	4.66	5.14	4.38	4.80
b	4.82	5.08	4.92	5.88	5.48	5.94	4.74	5.50	6.04	4.96	5.58
c	4.34	4.80	3.30	3.70	3.64	6.60	5.40	4.80	4.54	3.60	4.06
d	5.66	5.40	7.28	**8.38**	**8.20**	6.52	6.20	**7.70**	**9.92**	**8.94**	**9.52**
c′	4.68	4.88	2.90	3.04	2.86	5.94	5.06	3.98	3.70	3.00	3.36
d′	5.46	5.38	**8.00**	**10.02**	**9.70**	6.72	6.34	**8.50**	**10.94**	**9.56**	**10.36**
Response variable χ^2_3 distributed											
$f = 3,\ N = 18,\ \varepsilon = 1.00$											
a	6.44	–	6.56	**10.24**	**8.84**	3.94	–	3.96	4.62	3.54	4.26
b	6.20	–	6.48	**11.46**	**10.26**	5.50	–	6.28	**8.20**	5.54	7.02
c	6.32	–	3.88	6.18	4.78	4.80	–	3.54	3.64	**2.12**	3.02
d	**8.66**	**–**	**12.16**	**19.58**	**18.42**	7.10	–	**10.52**	**14.36**	**10.98**	**13.02**
c′	6.32	–	2.56	4.46	3.26	4.98	–	2.50	2.58	**1.60**	**2.28**
d′	**10.96**	**–**	**15.62**	**27.50**	**22.48**	**9.36**	**–**	**14.42**	**19.90**	**15.88**	**18.60**
$f = 6,\ N = 45,\ \varepsilon = 0.40$											
a	5.40	–	5.46	6.78	6.56	5.86	–	4.88	4.18	3.80	4.06
b	5.10	–	5.30	**8.26**	**8.10**	5.78	–	5.70	6.46	5.68	6.10
c	5.40	–	4.40	7.42	7.12	5.94	–	4.60	5.78	4.78	5.30
d	5.76	–	7.44	**11.22**	**11.00**	6.14	–	7.36	**9.32**	**7.62**	**8.56**
c′	5.24	–	3.62	4.94	4.78	5.62	–	3.92	3.84	2.96	3.54
d′	5.32	–	**7.84**	**12.88**	**12.62**	6.10	–	**7.96**	**10.76**	**9.64**	**10.42**

Note: C, Pairing of covariance matrices and group sizes condition; IGA, Huynh's (1978) Improved General Approximation Test; WJ, Welch–James test (Keselman *et al*., 1993), $\bar{\varepsilon}$, EBAR (Quintana & Maxwell, 1994); T^2, Hotelling's test; HL, Hotelling–Lawley trace; P, Pillai test; W, Wilks test; EB, empirical Bayes. – indicates that rates of Type I error were not collected since sample size was smaller than recommended values.

Keine between-subjects Faktoren *oder* balanciertes Design

- mehrere Ansätze **grundsätzlich** geeignet (Kontrolle der Type I Fehlerrate)
- **univariater** Ansatz mehr Power als **multivariat**, falls $\varepsilon > 0.85$ und $N < K + 30$ (K = Anzahl Faktorstufen within-subjects Faktor) (Algina & Keselman, 1997)
  - falls $N < K$: nicht genug *df*s für multivariaten Ansatz
- **EB**, **mixed model mit Kenward-Roger**: mehr Power als univariat und multivariat (Keselman, Algina, & Kowlachuk, 2001)
- **univariat**: Vorteile bei **nicht normalverteilten** Daten gegenüber multivariat und EB, besonders bei sehr kleinen Stichproben ($N < 15$) (Keselman, Kowalchuk & Boik, 2000; Berkovits et al., 2000)
  - mixed model: Befunde noch unklar
    - Gomez, Schaalje, & Fellingham (2005): sehr kleine Stichproben (3-5 Vpn pro Gruppe) -> liberale Tests
    - Muller, Edwards, Simpson, & Taylor (2007): bei $N = 10$ univariat okay, PROC MIXED aber nicht
- Mein Fazit: univariat mit *df*-Korrektur, multivariat falls großes N, PROC MIXED nur bei missing data

Software

- **Rechenbeispiel:** Oberfeld & Hecht (2008, JEP: HPP), Exp. 1
  - Kontaktzeitschätzung
    - Auto kommt in VR auf Vp zu
    - wird kurz vor Vp abgeblendet
    - Vp gibt an, ob das Auto *spät* oder *früh* bei ihr angekommen wäre
  - 2 within-subjects factors:
    - Distraktorbedingung
      - nur Target
      - LKW, der früher als das Auto ankommt
      - LKW, der später als das Auto ankommt
    - Geschwindigkeit Target (3 Stufen)
  - AV: Proportion der “Target spät” Antworten (30 Trials pro Datenpunkt)
  - *N* = 11

Split-plot Design, *unbalanciert*

- Univariat/multivariat: Paarung Cov-Struktur / Gruppengröße
  - Test **konservativ**, wenn Cov-Matrix der kleinsten Gruppe die kleinsten Elementwerte hat (*positive* Paarung)
  - Test **liberal**, wenn *negative* Paarung

- Alternativen:
  - **Improved General Approximation Test** (IGA; Algina, 1997; basiert auf Huynh, 1978)
    - SAS Macro verfügbar
  - **Welch-James Test** (Johansen, 1980; Keselman, Carriere, & Lix 1993)
    - größere Samples notwendig
      - bei 3 Gruppen: Haupteffekt (n_{min}) > 3(K - 1), Interaktion (n_{min}) > 5(K - 1)
    - SAS Macro verfügbar
  - **Mixed model**: mit Kenward Roger adjusted F-Tests, UN-H Cov-Struktur / AIC
    - SAS PROC MIXED
  - **Brown-Forsythe** (Vallejo Seco et al., 2006)

Auswahlregel

- **kleine** Stichproben
  - IGA + BF besser als WJ (Kontrolle der Type I error rates)
  - PROC MIXED mit Kenward Roger und UN-H: auch bei relativ kleinen Samples (N = 30) und nicht normalverteilten Daten okay, insofern dem WJ-Test vorzuziehen (Kowalchuk et al., 2004)
    - allerdings: Vallejo & Livacic-Rojas (2005) etwas ungünstigere Resultate, besonders bei nicht-normalverteilten Daten
  - Simulationen für **sehr** kleine Stichproben noch rar
    - PROC MIXED Gomez, Schaalje, & Fellingham (2005): bei 3 Vpn pro Gruppe problematische Resultate
    - evtl. IGA besser, *aber unklar*
- **große** Stichproben: WJ hat mehr Power als IGA
- **[evtl. Alternative:** Bootstrap (Vallejo Seco et al., 2006)]

Beispiel

- Geschlecht ausgedacht für jede Vp…

Multiple comparisons

- Bislang: Omnibus-Tests
- Mehrere Kontraste (geplant, post-hoc) bzw. Paarvergleiche: Homogenität der Varianzen **noch** wichtiger als bei Omnibus-Tests (Mitzel & Games, 1981)
  - *df*-Korrektur reicht nicht aus

- **Lösung**: nicht-gepoolte Varianzen verwenden (nicht relevante Datenpunkte ignorieren, s. Maxwell & Delaney, 2004, S. 647ff, 695ff)

$$F = n\,\bar{D}^{2}/s_{\mathrm{D}}^{2}$$

  - 1 und n – 1 Freiheitsgrade
  - kompatibel mit **multivariatem** Ansatz
    - wenn Haupteffekt signifikant -> mindestens 1 der möglichen Kontraste signifikant
  - nicht unbedingt kompatibel mit univariatem Ansatz
- Zweifaktorielles Design: “simple main effects” als separate ANOVAs
  - z.B. unterscheidet sich der Effekt der Distraktorbedingung zwischen den beiden langsamsten Geschwindigkeiten? -> ANOVA, in der die Daten mit der höchsten Geschwindigkeit ignoriert werden

MCP

- Kontrolle der "familywise" Fehlerraten (nach Maxwell & Delaney, 2004):
  - planned comparisons: Bonferroni
  - Pairwise comp.: Bonferroni
  - Post-hoc complex comparisons: Roy-Bose

- Falls between-subjects Faktor im Design: wieder "multi-sample sphericity" kritisch
  - Überblick über valide Verfahren: Keselman (1994)

Missing data

- univariat / multivariat: alle Vpn mit mindestens einem missing value werden **komplett** ignoriert
  - Schätzer unverzerrt, falls MCAR -- jedoch nicht, falls MAR
  - auch bei MCAR: deutlicher Verlust an Power!!

- Rubin (1976): maximum-likelihood Analysen mit "Ignorieren" der missing data valide, falls MCAR **oder** MAR **oder** "*Covariate-dependent dropout*" (cf. Keselman, Algina, & Kowalchuk, 2001, 2002)

- Padilla & Algina (2004, 2007): PROC MIXED mit Kenward-Roger bei missing data okay

Fazit ANOVA

- Design: sicherstellen, dass **identische** Gruppengrößen
  - **univariat** mit Huynh-Feldt df Korrektur
  - bei großen Samples: **multivariat**
  - bei missing values: **PROC MIXED** mit Kenward-Roger adjusted *F*-Tests
- bei **ungleichen** Gruppengrößen: IGA, BF oder WJ (bei großen Samples), evtl. auch PROC MIXED
- **MCP**: nicht-gepoolte Tests verwenden
- **[Multivariat (mehrere AVs):** Vallejo Seco, G., Gras, J. A., & Garcia, M. A. (2007). Comparative robustness of recent methods for analyzing multivariate repeated measures designs. *Educational and Psychological Measurement, 67*(3), 410-432.)]

Bericht der Resultate

- Es gibt nicht **eine** RM-Analyse…
- Deshalb: genau sagen, **welche** Art von RM-Analyse verwendet wurde
  - *"The data were analyzed via a repeated-measures analysis of variance (ANOVA) using a univariate approach. The Huynh-Feldt correction for the degrees-of-freedom was used where applicable and the value of ε is reported."*
    - *There was a significant effect of physical height, $F(2, 22) = 19.46$, $p = .001$, $\varepsilon = .78$, $\eta^2 = .64$.*
  - *"The data were analyzed via a repeated-measures ANOVA using a multivariate approach."*
    - *There was a significant effect of physical height, $F(2, 10) = 10.76$, $p = .003$, $\eta^2 = .68$.*
  - *"The data were analyzed via a repeated-measures ANOVA using a mixed-model approach (SAS PROC MIXED) with Kenward-Roger's adjusted F-tests (Kenward & Roger, 1997). The "unstructured-heterogenous (UN-H)" was selected to model the covariance structure (Kowalchuk, Keselman, Algina, & Wolfinger, 2004)".*

Regression

- *Jede Vp verkostet Wein unter 4 Farben, nennt jeweils den maximalen Kaufpreis **und** schätzt die Fruchtigkeit des Weins auf einer Skala von 1-6 ein*
- **"Naive" Analyse**: Korrelation / Regression über alle Datenpunkte (15 Vpn x 4 Farben)
  - Signifikanztest für den Korrelationskoeffizient (*t* Test) nicht gültig, da Beobachtungen nicht unabhängig: (15 x 4) – 1 *df*s, obwohl es nur 15 Vpn gibt
    - Bland & Altman (1994)
- **"Data resolution"** (Burton et al, 1994):
  - Analyse der *Mittelwerte* jeder Vp: "Sind Vpn, die im Mittel die Weine fruchtig finden, auch bereit, viel Geld dafür auszugeben"?
    - Information über die verschiedenen Farbbedingungen geht verloren
  - separate Regressionen pro Vp, Intercepts & Regressionskoeffizienten mitteln
    - valide
    - aber uneffektiv, da wir Informationen über die "Masse" der Vpn verlieren

Regression für korrelierte Daten, 1

- **"Subject-specific":**
  - Zur Erinnerung: "naiv" wäre $Y_{ij} = \alpha + \beta\, x_{ij} + \varepsilon_{ij}$
    - Y_{ij} MBP von Vp *i* unter Farbe *j*
    - x_{ij} Fruchtigkeitsrating von Vp *i* unter Farbe *j*
  - Random effects model: $Y_{ij} = \alpha + a_i + (\beta + b_i)\, x_{ij}$
    - α „overall“ Populations-Intercept
    - a_i Diskrepanz zwischen „wahrem“ Intercept von Vp *i* und α (random effect Vp)
    - β „overall“ Populations-Slope
    - b_i Diskrepanz zwischen „wahrem“ Slope von Vp *i* und β (random effect Vp)
- Mixed model, MLE (SAS PROC MIXED)
  - liefert Koeffizienten und valide Tests (Hauptinteresse: fixed effects)
  - Annahme einer Varianz-Kovarianz Matrix notwendig
    - am „sichersten“: UN

Regression für korrelierte Daten, 2

- **“population averaged / marginal model”:**
  - $Y_{ij} = \alpha + \beta x_{ij} + \varepsilon_{ij}$
    - wie im „naiven“ Ansatz
  - Korrelationsstruktur der Antworten wird **separat** geschätzt
  - iteratives Verfahren

- Schätzung mittels Generalized Estimating Equations (GEE; Liang & Zeger, 1986)
  - liefert „robuste“ Schätzer der Koeffizienten & ihrer Standardfehler
  - Varianz-Kovarianz Matrix: „Störvariable“, aus Datenstruktur geschätzt
  - „working correlation structure“ muss angegeben werden
  - Modell ist auch bei Fehlspezifikation dieses Startwerts robust (Overall & Tonidandel, 2004)
- SAS: PROC GENMOD
- Nachteil: Goodness-of-fit Tests nicht trivial (Pan, 2001)

Beispiel

- Farbweinstudie 2: jede Vp verkostet 2 Weine (trocken/halbtrocken) jeweils unter 4 Farben

- Rating:
  - Geruchsaspekte (z.B. "Der Wein riecht nicht fruchtig … sehr fruchtig")
  - globales Urteil ("Ich mag den Wein gar nicht … sehr gerne")

- multiple Regression: kann man das **globale Urteil** aus den **Geruchsratings** vorhersagen?

Kategoriale Daten

- Binäre oder ordinale Daten: **logistische Regression**
  - wie oben: subject-specific oder marginal model
  - SAS PROC NLMIXED / SAS PROC GENMOD

Literatur

- Basistext für ANOVA

  - Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective* (2nd ed.). Mahwah, N.J.: Lawrence Erlbaum Associates.

- Reviews der verschiedenen ANOVA-Verfahren

  - Keselman, H. J., Algina, J., & Kowalchuk, R. K. (2001). The analysis of repeated measures designs: A review. *British Journal of Mathematical & Statistical Psychology, 54*, 1-20.
  - Keselman, H. J., Algina, J., & Kowalchuk, R. K. (2002). A comparison of data analysis strategies for testing omnibus effects in higher-order repeated measures designs. *Multivariate Behavioral Research, 37*(3), 331-357.

Literatur 2

- PROC MIXED

  - Kowalchuk, R. K., Keselman, H. J., Algina, J., & Wolfinger, R. D. (2004). The analysis of repeated measurements with mixed-model adjusted F tests. *Educational and Psychological Measurement, 64*(2), 224-242.
  - Gomez, E. V., Schaalje, G. B., & Fellingham, G. W. (2005). Performance of the Kenward-Roger method when the covariance structure is selected using AIC and BIC. *Communications in Statistics-Simulation and Computation, 34*(2), 377-392.
  - Littell, R. C., Henry, P. R., & Ammerman, C. B. (1998). Statistical analysis of repeated measures data using SAS procedures. *Journal of Animal Science, 76*(4), 1216-1231. (*Einführung in die Syntax*)

- Missing data und PROC MIXED

  - Padilla, M. A., & Algina, J. (2007). Type I Error Rates of the Kenward-Roger Adjusted Degree of Freedom F-test for a Split-Plot Design with Missing Values. *Journal of modern applied statistical methods, 6*(1), 66-80.
  - Keselman, H. J., Algina, J., & Kowalchuk, R. K. (2002), s.o.

Literatur 3

- MCP

  - Mitzel, H. C., & Games, P. A. (1981). Circularity and Multiple Comparisons in Repeated Measure Designs. *British Journal of Mathematical & Statistical Psychology, 34*(Nov), 253-259.
  - Keselman, H. J. (1994). Stepwise and Simultaneous Multiple Comparison Procedures of Repeated-Measures Means. *Journal of Educational Statistics, 19*(2), 127-162.

Literatur 4

- Regression for correlated data

  - Burton, P., Gurrin, L., & Sly, P. (1998). Extending the simple linear regression model to account for correlated responses: An introduction to generalized estimating equations and multi-level mixed modelling. *Statistics in Medicine, 17*(11), 1261-1291.
  - Bland, J. M., & Altman, D. G. (1994). Correlation, regression, and repeated data. *British Medical Journal, 308*(6933), 896.
  - Bland, J. M., & Altman, D. G. (1995). Statistics Notes .12. Calculating Correlation-Coefficients with Repeated Observations .1. Correlation within-Subjects. *British Medical Journal, 310*(6977), 446-446.
  - Hamlett, A., Ryan, L., Serrano-Trespalacios, P., & Wolfinger, R. (2003). Mixed models for assessing correlation in the presence of replication. *Journal of the Air & Waste Management Association, 53*(4), 442-450.
  - **Kategoriale Daten:** Pendergast, J. F., Gange, S. J., Newton, M. A., Lindstrom, M. J., Palta, M., & Fisher, M. R. (1996). A survey of methods for analyzing clustered binary response data. *International Statistical Review, 64*(1), 89-118.