

Parametrische und nichtparametrische simultane Konfidenzintervalle für multiple Kontraste

Edgar Brunner

und

Frank Konietzschke

Abteilung Medizinische Statistik

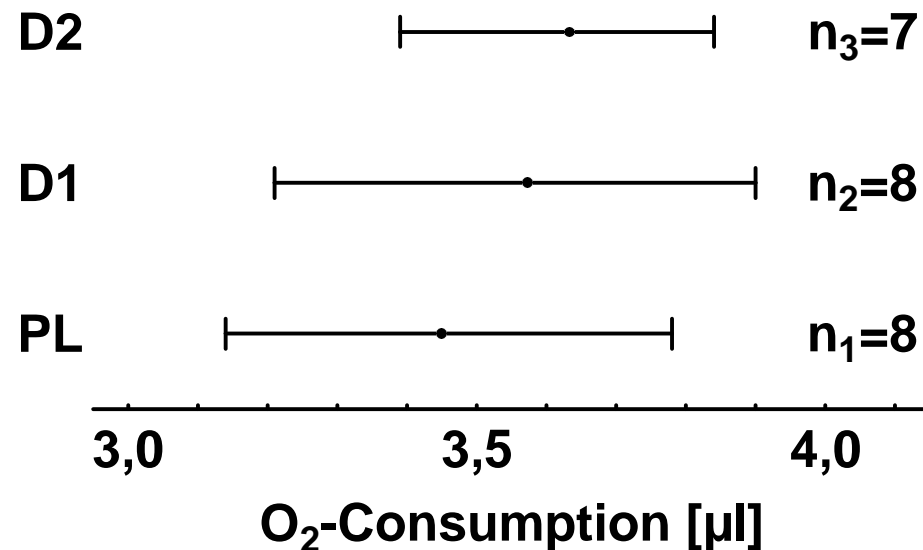
Universität Göttingen

Inhalt

- **Parametrische Methoden**
 - ▷ Motivierendes Beispiel
 - ▷ SKI Methode
 - ▷ Analyse des Beispiels
- **Nichtparametrische Methoden**
 - ▷ Motivierendes Beispiel
 - ▷ SKI Methode
 - ▷ Analyse des Beispiels
 - ▷ Spezielle Schwierigkeiten
- **Literatur**

I Parametrische Methoden

- Motivierendes Beispiel
- O₂-Verbrauch von Leukozyten
 - ▷ Linien stellen min | ... | max der Daten dar



- Fragestellung
 - ▷ Welche Dosis bringt einen Unterschied zur Kontrolle?

Motivierendes Beispiel

- **Klassische Analyse**

- (1) ANOVA / $H_0 : \mu_P = \mu_1 = \mu_2$
- (2) H_0 abgelehnt \rightarrow multiple Vergleiche ($FWE_s = 0.05$)
- (3) Konfidenzintervalle für $\mu_1 - \mu_P$ und $\mu_2 - \mu_P$
 - müssen kompatibel zu den Entscheidungen der multiplen Vergleiche sein
 - d.h. Konfidenzintervall (KI) für $\mu_i - \mu_P$ darf nicht die 0 enthalten $\iff H_0 : \mu_i - \mu_P = 0$ abgelehnt wird, $i = 1, 2$

- **Statistische Methoden / Verfahren**

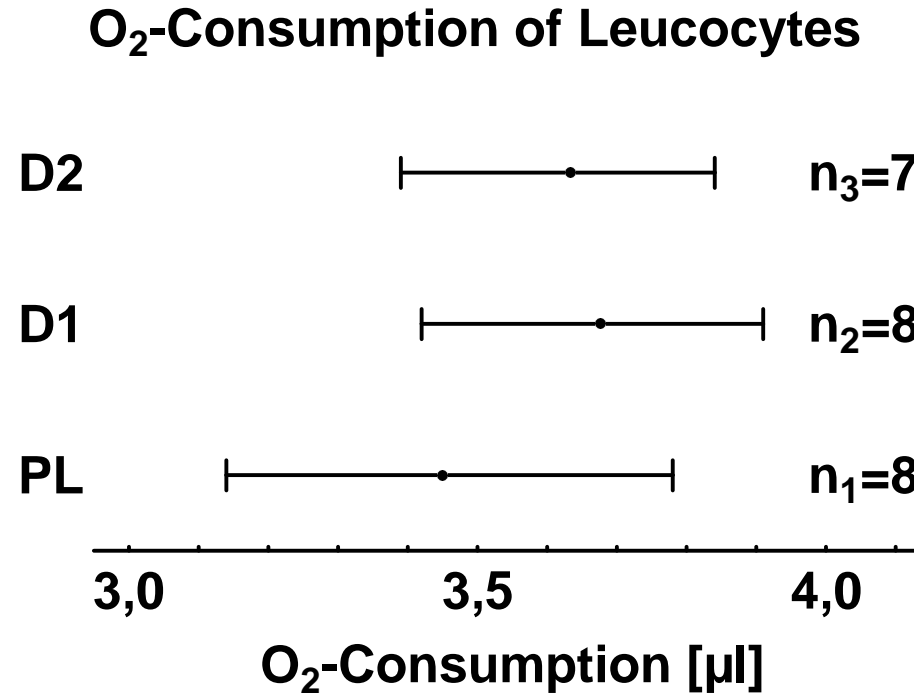
- ▷ ANOVA (F -Test)
- ▷ multiple Vergleiche, z.B: Abschlusstest (AT)
- ▷ Bonferroni Konfidenzintervalle ($1 - \alpha = 0.975$)

- **Ergebnisse**

- ▷ Globalhypothese: $F = 2.53$ p -Wert 0.1056 - (n.s.)
- ▷ multiple Vergleiche
 - PL - D1: $p = 0.1424$ - (n.s.) / PL - D2: $p = 0.0488$ - (n.s.)

Motivierendes Beispiel

- Verschiebung der Daten für D1



- **Ergebnis**

- ▷ Globalhypothese: $F = 4.06$ p -Wert 0.0355 - (*)
- ▷ Abschlusstest
 - PL - D1: $p = 0.0256$ - (*) / PL - D2: $p = 0.0488$ - (*)

Folgerungen aus dem motivierenden Beispiel . . .

- **Konfidenzintervalle (Bonferroni)**
 - ▷ PL - D1: $[-0.024, 0.557]$ - enthält 0 / nicht kompatibel zum AT
 - ▷ PL - D2: $[-0.063, 0.538]$ - enthält 0 / nicht kompatibel zum AT
- **Konsequenzen (unerwünschte Eigenschaften)**
 - ▷ Entscheidung über den Effekt PL - D2 abhängig vom beobachteten Effekt PL - D1
 - ▷ Konfidenzintervalle nicht kompatibel zum Abschlusstest
 - ▷ Abhängigkeiten der Statistiken $\bar{X}_1 - \bar{X}_P$ and $\bar{X}_2 - \bar{X}_P$ nicht berücksichtigt - (Information verschwendet)
 - ▷ andere Methode wird benötigt

Andere Methode

• Idee

- ▷ Statistisches Modell
 - adaptiert und reduziert
 - auf die spezielle Fragestellung des Anwenders
- ▷ Abhängigkeit der Statistiken berücksichtigen
 - bei völliger Abhängigkeit → keine α -Adjustierung nötig
 - Unabhängigkeit ist ungünstigster Fall
- ▷ Beispiel: O₂-Verbrauch

- $\mathbf{C} = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} = (-\mathbf{1}_2 : \mathbf{I}_2)$ and $\bar{\mathbf{X}} = (\bar{X}_{P.}, \bar{X}_{1.}, \bar{X}_{2.})'$

- Kontraste $\mathbf{C}\bar{\mathbf{X}} = \begin{pmatrix} \bar{X}_{1.} - \bar{X}_{P.} \\ \bar{X}_{2.} - \bar{X}_{P.} \end{pmatrix}$, $\boldsymbol{\mu}_\delta = \begin{pmatrix} \mu_1 - \mu_P \\ \mu_2 - \mu_P \end{pmatrix}$

- betrachte Verteilung von $\mathbf{C}\bar{\mathbf{X}} \sim N(\boldsymbol{\mu}_\delta, \boldsymbol{\Sigma})$

- $\boldsymbol{\Sigma} = \sigma^2 \left[\begin{pmatrix} n_1^{-1} & 0 \\ 0 & n_2^{-1} \end{pmatrix} + n_P^{-1} \mathbf{J}_2 \right] = (s_{ij})_{i,j=1,2}$

Andere Methode

• Herleitung der Statistik

▷ $s_{ii} = \sigma^2 \cdot (n_i + n_P) / (n_i \cdot n_P)$, $i = 1, 2$ - Diagonalelemente von Σ

▷ \hat{s}_{ii} : LS-Schätzer für s_{ii} in σ^2 einsetzen

$$\bullet \hat{\sigma}_N^2 = \frac{1}{N-3} \sum_{i=P,1,2} \sum_{k=1}^{n_i} (X_{ik} - \bar{X}_{i.})^2, \quad N = n_1 + n_2 + n_P$$

▷ studentisiere jede Zeile in $\mathbf{C}\bar{\mathbf{X}} = \begin{pmatrix} \bar{X}_{1.} - \bar{X}_{P.} \\ \bar{X}_{2.} - \bar{X}_{P.} \end{pmatrix}$ mit $\sqrt{\hat{s}_{ii}}$

▷ t -Statistiken für $H_0 : \boldsymbol{\mu}_\delta = \mathbf{0}$

$$\bullet T_i = \sqrt{\frac{n_i n_P}{n_i + n_P}} (\bar{X}_{i.} - \bar{X}_{P.}) / \hat{\sigma}_N \overset{\cdot}{\sim} N(0, 1), \quad N \rightarrow \infty,$$
$$N/n_i < N_0 < \infty, \quad i = 1, 2$$

▷ multivariate Statistik

$$\bullet \mathbf{T} = (T_1, T_2)' \overset{\cdot}{\sim} N(\mathbf{0}, \mathbf{R}), \quad \mathbf{R}: \text{Korrelationsmatrix}$$

Andere Methode

- **Herleitung der $(1 - \alpha)$ -Quantile**

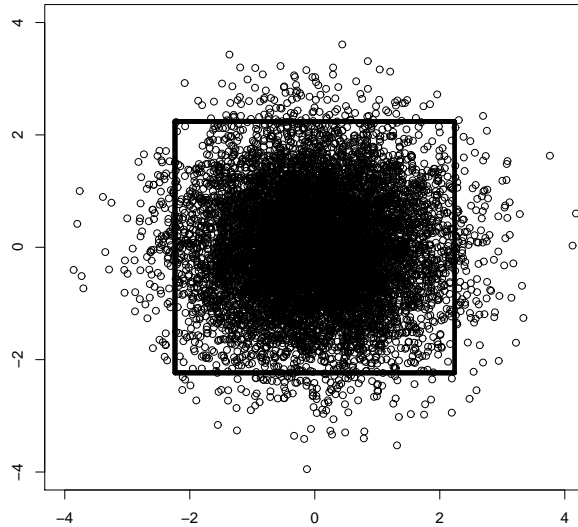
- ▷ ein kritischer Wert $z_{1-\alpha,2,\mathbf{R}}$ für alle Komponenten, so dass

$$\int_{-z_{1-\alpha,2,\mathbf{R}}}^{z_{1-\alpha,2,\mathbf{R}}} \int_{-z_{1-\alpha,2,\mathbf{R}}}^{z_{1-\alpha,2,\mathbf{R}}} dN(\mathbf{0}, \mathbf{R}) = 1 - \alpha$$

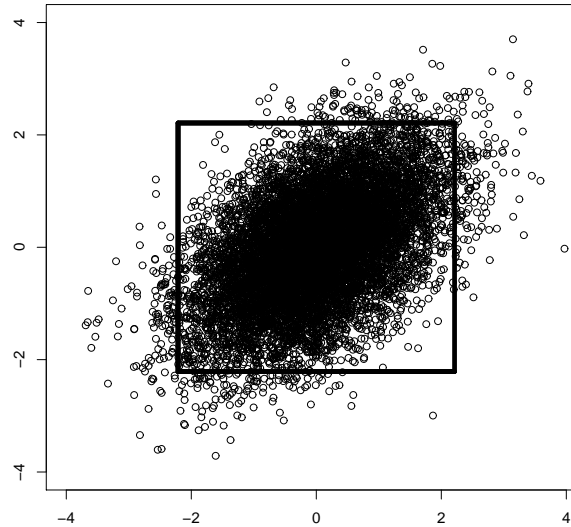
- ▷ bessere Approximation: multivariate t -Verteilung: $t_{1-\alpha,2,\mathbf{v},\hat{\mathbf{R}}_N}$
- ▷ $\hat{\mathbf{R}}_N$: LS-Schätzer für \mathbf{R}
 - ersetze σ^2 mit $\hat{\sigma}_N^2$
 - Diagonalelemente = 1
 - alle anderen Elemente nur von n_i und $\hat{\sigma}_N^2$ abhängig
 - $\mathbf{T} \sim$ multivariate t -Verteilung
- ▷ Literatur
 - ursprüngliche Arbeit: Bretz, Genz and Hothorn (2001)
 - multivariate Integration: Genz and Bretz (2009)
 - heteroskedastischer Fall: Hasler and Hothorn (2008)
- allgemein: \mathbf{C} kann irgendeine geeignete Kontrastmatrix sein

SKI-Methode / Quantile

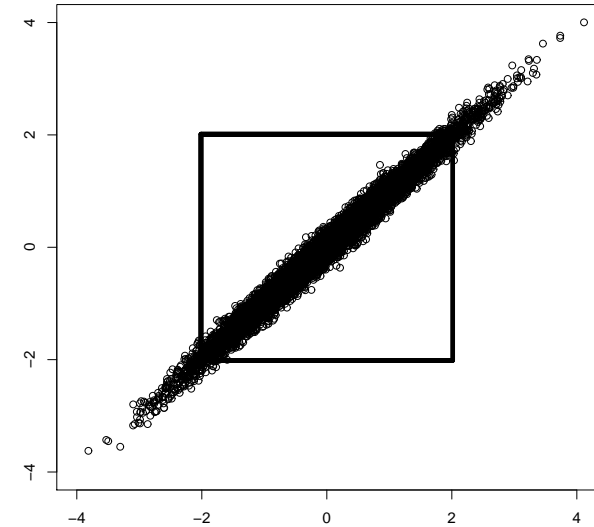
Korrelation = 0, Quantil= 2.2365



Korrelation = 0.5, Quantil= 2.2121



Korrelation = 0.99, Quantil= 2.0133



- simultane Quantile für verschiedene bivariate Normalverteilungen
- Quadrate enthalten die Masse $1 - \alpha$ der bivariate Normalverteilungen
- Berechnung mithilfe des R-Paketes „mvtnorm“
- SAS-Makro: muss noch entwickelt werden
- oder Input von R-Code in SAS/IML Studio 3.2

SKI-Methode / Verfahren

- **Multiple Vergleiche**

- ▷ $H_0^{(i)} : \delta_i = \mu_i - \mu_P = 0$ verwerfen, falls

- $|T_i| \geq z_{1-\alpha,2,\mathbf{R}}$ - or $|T_i| \geq t_{1-\alpha,2,\mathbf{v},\hat{\mathbf{R}}_N}$

- **Globalhypothese**

- ▷ $H_0 : \mathbf{C}\boldsymbol{\mu} = \boldsymbol{\mu}_\delta = \mathbf{0}$ verwerfen, falls

- $\max\{T_1, T_2\} \geq z_{1-\alpha,2,\mathbf{R}}$ - oder $\max\{T_1, T_2\} \geq t_{1-\alpha,2,\mathbf{v},\hat{\mathbf{R}}_N}$

- **Simultane Konfidenzintervalle**

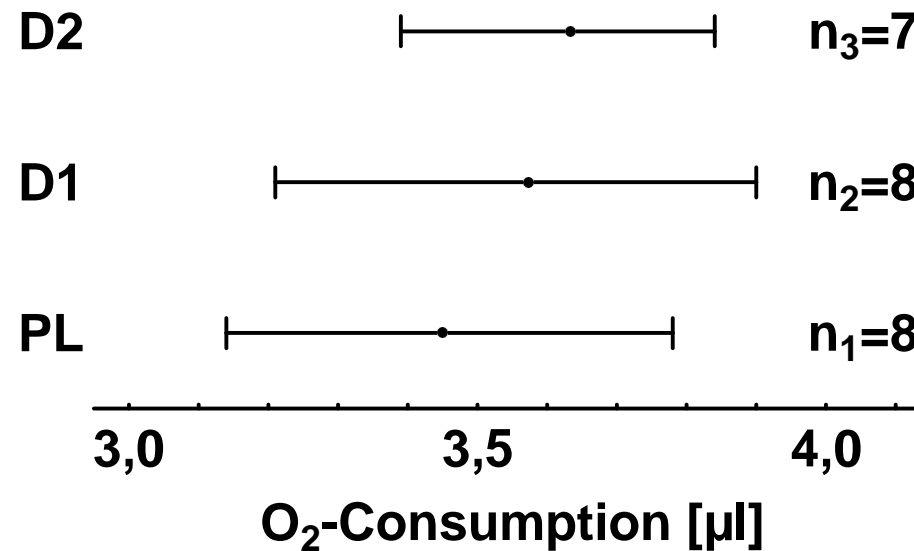
- ▷ $P \left(\bigcap_{i \in I} \left\{ \delta_i \in \left[\bar{X}_{i\cdot} - \bar{X}_{P\cdot} \pm \frac{z_{1-\alpha,2,\hat{\mathbf{R}}_N}}{\hat{\sigma}_N} \sqrt{\frac{n_i + n_P}{n_i n_P}} \right] \right\} \right) \doteq 1 - \alpha$

- **Fehlerkontrolle?**

- ▷ FWE_s (aufgrund des Satzes von Gabriel, 1969)

Beispiel: Analyse mit der SKI-Methode

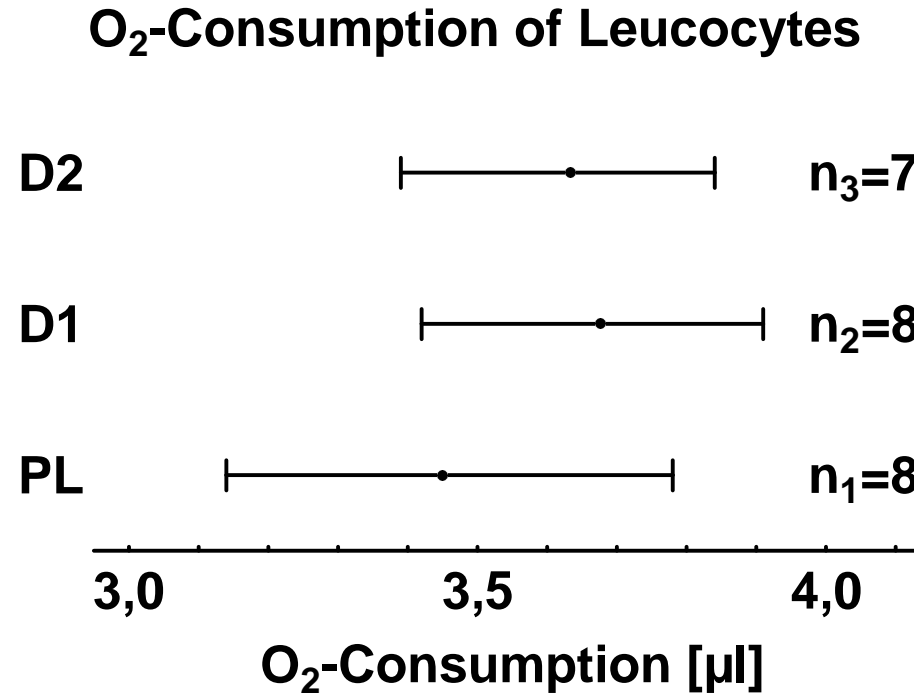
- Ursprünglicher Datensatz (O₂-Verbrauch von Leukozyten)



SKI			klassisch
PL - D1	$t = 2.10$	p -Wert 0.0965 - n.s.	n.s.
PL - D2	$t = 2.18$	p -Wert 0.0864 - n.s.	n.s.

Beispiel: Analyse mit der SKI-Methode

- Verschiebung der Daten für D1



		SKI			klassisch
PL - D1	$t = 2.53$	p -Wert	0.0460	- (*)	(*)
PL - D2	$t = 2.18$	p -Wert	0.0864	- n.s.	(*)

Folgerungen aus dem Beispiel

- **Konfidenzintervalle (Daten für D1 verschoben)**

- ▷ PL - D1: $[0.0049, 0.5276]$ - enthält nicht 0 / kompatibel
- ▷ PL - D2: $[-0.0324, 0.5074]$ - enthält 0 / kompatibel

- **Folgerungen**

- ▷ Entscheidung über den Effekt PL - D2 nicht abhängig vom beobachteten Effekt PL - D1
- ▷ Konfidenzintervalle sind kompatibel zu den Testergebnissen
- ▷ Abhängigkeit der Statistiken $\bar{X}_1 - \bar{X}_P$ and $\bar{X}_2 - \bar{X}_P$ berücksichtigt

Erweiterungen / Verallgemeinerungen

- **Faktorielle Designs**

- ▷ Biesheuvel and Hothorn (2002) / stratifizierte Stichproben
- ▷ allgemeiner Fall: derzeit in einer Diplomarbeit untersucht

- **große Anzahl von Dimensionen**

- ▷ $\widehat{\Sigma}_N$ kann singulär werden (bricht was zusammen?)

- **Repeated Measures**

- ▷ $n \geq d$ and $n < d$ (bricht was zusammen?)
- ▷ hoch-dimensionale Daten / Froemke, Hothorn and Kropf (2008)
- ▷ gibt es eine Grenzverteilung?

- **dichotome Daten**

- ▷ Schaarschmidt, Sill and Hothorn (2008)

- **nichtparametrische Effekte**

- ▷ nicht-normalverteilte Daten (Konietschke, 2009)
- ▷ ordinale Daten: ordinales Effektmaß (Ryu and Agresti, 2008)

II Nichtparametrische Methoden

- **Motivierendes Beispiel**
- **Toxizitätsversuch (60 Wistar Ratten)**
 - ▷ Schädigung der Nasenschleimhaut durch eine inhalierbare Substanz
 - ▷ 3 Konzentrationen (2[ppm], 5[ppm], 10[ppm])
 - ▷ Score ($0 \hat{=}$ „keine Schädigung“, ..., $3 \hat{=}$ „schwere Schädigung“)

Konzentration	Anzahl der Ratten mit Score			
	0	1	2	3
2 [ppm]	18	2	0	0
5 [ppm]	12	6	2	0
10 [ppm]	3	7	6	4

- ▷ ordinale Daten

Motivierendes Beispiel

• Klassische Analyse Strategie

▷ statistisches Modell $X_{ik} \sim F_i(x)$, $i = 1, 2, 3$; $k = 1, \dots, 20$

▷ Hypothese

- $H_0^{(1)} : F_1 = F_2 = F_3$

- $H_0^{(2)} : F_1 = F_2$ - relativer Effekt: $p_{12} = \int F_1 dF_2$

- $H_0^{(3)} : F_1 = F_3$ - relativer Effekt: $p_{13} = \int F_1 dF_3$

- $H_0^{(4)} : F_2 = F_3$ - relativer Effekt: $p_{23} = \int F_2 dF_3$

▷ relativer Effekt p_{ij} - Interpretation

- $p_{ij} = \int F_i dF_j = P(X_{i1} < X_{j1}) + \frac{1}{2}P(X_{i1} = X_{j1})$

- W', dass die Beobachtungen in Gruppe i zu kleineren Werten tendieren als in Gruppe j

- ordinale Daten: ordinales Effektmaß (Ryu and Agresti, 2008)

▷ benötigt: Konfidenzintervalle für $p_{ij} = \int F_i dF_j$, $i \neq j = 1, 2, 3$

▷ Fehlerkontrolle: FWE_s

SKI-Methode

- **interessierende Hypothesen**

- ▷ $H_0^{(1)} : p_{12} = \frac{1}{2}, \quad H_0^{(2)} : p_{13} = \frac{1}{2}$

- **Schätzer für den relativen Effekt p_{ij}**

- ▷ $\hat{p}_{ij} = \int \hat{F}_i d\hat{F}_j = \frac{1}{n_i} \left(\bar{R}_{j\cdot}^{(ij)} - \frac{n_{j+1}}{2} \right) \rightarrow \hat{\mathbf{p}} = \begin{pmatrix} \hat{p}_{12} \\ \hat{p}_{13} \end{pmatrix}$

- ▷ asymptotische Verteilung von $\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p}) \sim N(\mathbf{0}, \mathbf{V}_N)$

- ▷ hängt von unbekanntem Parametern ab (Elemente von \mathbf{V}_N)

- ▷ keine Pivot-Größe

- **Statistiken**

- ▷ studentisiere jede Zeile (i, j) von $\hat{\mathbf{p}}$ mit $\sqrt{\hat{v}_{(ij)}}$

- ▷ $\hat{v}_{(ij)}$: Varianzschätzer für \hat{p}_{ij} (Diagonalelemente von \mathbf{V}_N)

- ▷ Schätzung mittels Rängen $R_{ik}^{(ij)}, R_{jk}^{(ij)}, R_{ik}^{(i)}$, und $R_{jk}^{(j)}$

- Literatur: Brunner, Munzel und Puri (2002)

SKI-Methode

- **Asymptotische Verteilung der Statistiken**

- ▷ asymptotische Verteilung unter $H_0^{(ij)} : p_{ij} = \frac{1}{2}$ von

$$T_{ij} = \sqrt{N} \cdot (\hat{p}_{ij} - \frac{1}{2}) / \sqrt{\hat{v}_{ij}} \overset{\cdot}{\sim} N(0, 1)$$

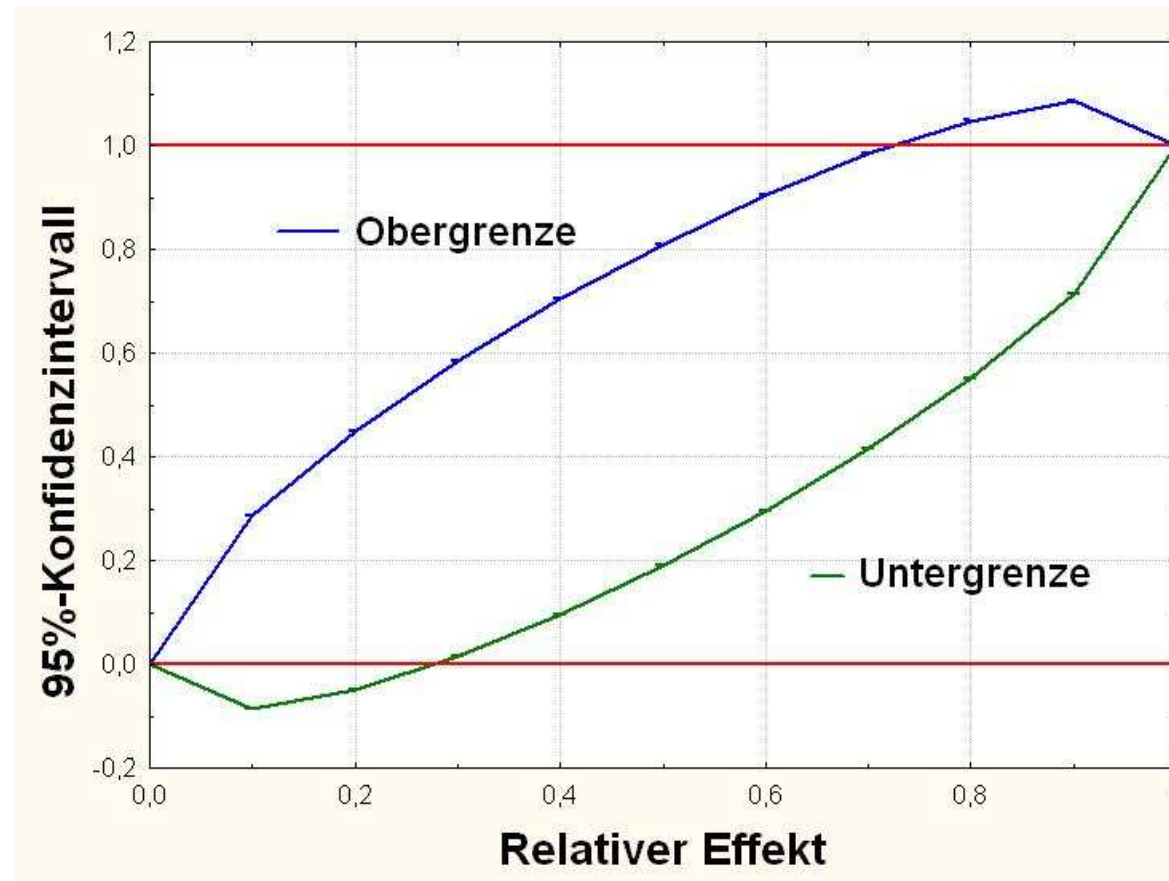
- ▷ $\mathbf{T} = (T_{12}, T_{13})' \overset{\cdot}{\sim} N(\mathbf{0}, \mathbf{R})$, \mathbf{R} : Korrelationsmatrix

- verwende dasselbe Verfahren wie im parametrischen Fall
- Fehlerkontrolle: FWE_s
- Problem: Konfidenzintervall kann über das $[0, 1]$ -Intervall hinausgehen \rightarrow nicht bereichserhaltend

SKI-Methode / Eigenschaften

- **Problem**

- ▷ Konfidenzintervalle sind nicht bereichserhaltend



- ▷ untere und obere 95% Konfidenzgrenzen ($n = 10$)

- **Lösung** multivariate δ -Methode

Bereichserhaltende Intervalle

- **Verfahren**

- ▷ stetige Transformation $G(\hat{p}_{ij}) \rightarrow (-\infty, \infty)$
- ▷ $\mathbf{G} : (G_1, \dots, G_q) : (0, 1)^q \rightarrow \mathbb{R}^q$
 - strikt monoton, d.h. $G'_\ell(p_{ij}) \neq 0$
 - differenzierbar, bijektiv, $G_\ell(\frac{1}{2}) = 0$, $\ell = 1, \dots, q$
- im Beispiel: $q = 2$
- asymptotische Verteilung von \mathbf{G} : Cramer's δ -Satz
 - ▷ transformierte Schätzer sind auch multivariat normalverteilt
 - ▷ Elemente v_{ij}^* der Kovarianzmatrix von \mathbf{G}
 - multivariater δ -Satz: $v_{ij}^* = [G'(p_{ij})]^2 \cdot v_{ij}$
- Rücktransformation der Grenzen $\rightarrow [0, 1]$ - bereichserhaltend

Beispiel: Analyse mittels SKI-Methode

- Toxizitätsversuch (60 Wistar Ratten)

Konzentration	Anzahl der Ratten mit Score			
	0	1	2	3
2 [ppm]	18	2	0	0
5 [ppm]	12	6	2	0
10 [ppm]	3	7	6	4

- Ergebnis (Probit-Transformation)

Vergleich	Effekt	Intervall	<i>p</i> -Wert
2 vs. 5	0.66	0.5 \notin [0.501;0.787]	0.049
2 vs. 10	0.90	0.5 \notin [0.753;0.970]	< 0.0001

Nichtparametrische Methoden / Probleme

- **Nicht-Transitivität**

- ▷ paarweise relative Effekte sind nicht transitiv
- ▷ z.B.: $p_1 < p_2 < p_3 < p_1$
- ▷ Gegenbeispiel: Efron's paradoxe Würfel (Rump, 2001)
 - Brown and Hettmansperger (2002) - Einweg-Klassifikation
 - Thangavelu und Brunner (2007) - stratifizierte Wilcoxon Tests

- **Neue Definition eines relativen Effektes für $a > 2$**

- ▷ z.B. $p_i = \int H dF_i$, $H =$ Mittelwert der F_i
- ▷ alle Verteilungen werden mit H verglichen
- ▷ oder alle Verteilungen werden mit derselben Referenz verglichen
- ▷ muss noch genauer ausgearbeitet werden
- ▷ Kovarianzmatrix von $\sqrt{N}(\hat{p}_1, \dots, \hat{p}_d)'$ ist sehr kompliziert
- ▷ erste Ergebnisse: Konietschke (2009)

- **Faktorielle Designs**

- ▷ jeden Faktor separat betrachten?
- ▷ alle Vergleiche in einem Vektor zusammenfassen?

Diskussion und Ausblick

- SKI-Methode fasst die 3 Schritte der klassischen Vorgehensweise
 - ▷ Globalhypothese (ANOVA)
 - ▷ multiple Vergleiche (unter FWE_s -Kontrolle)
 - ▷ Konfidenzintervalle für die Effekte - (kompatibel zur MCP!)
in *einem* Schritt zusammen
- weitere Fragen
 - ▷ genaue Ergebnisse zur Power
 - ▷ Erweiterung auf faktorielle Designs
 - ▷ Erweiterung auf Repeated Measures Designs
- sowohl für parametrische als auch nichtparametrische Modelle
- Software
 - ▷ bisher nur für unverbundene Stichproben (einfaktorielle Design)
 - ▷ parametrische Modelle: R-Paket: SimComp in CRAN
 - ▷ nichtparametrische Modelle: R-Paket: nparcomp in CRAN

Kooperation / Danksagung . . .

- Ludwig Hothorn und Assistenten
(Biostatistik, LU Hannover)
- Frank Konietzschke (Medizinische
Statistik, Universität Göttingen)

Literatur

- BIESHEUVEL, E. and HOTHORN, L.A. (2002). Many-to-one comparisons in stratified designs. *BIOMETRICAL JOURNAL* **44**, 101-116.
- BRETZ, F., GENZ, A., and HOTHORN, L.A. (2001). On the numerical availability of multiple comparison procedures, *Biometrical Journal* **43**, 645-656.
- BROWN, B. M. and HETTMANSPERGER, T. P. (2002). Kruskal-Wallis, Multiple Comparisons and Efron Dice. *Australian and New Zealand Journal of Statistics* **44**, 427-438.
- BRUNNER, E., MUNZEL, U., and PURI, M., (2002). The multivariate nonparametric Behrens-Fisher problem. *Journal of Statistical Planning and Inference* **108**, 37-53.
- FROEMKE C., HOTHORN L.A. and KROPF S. (2008). Nonparametric relevance-shifted multiple testing procedures for the analysis of high-dimensional multivariate data with small sample sizes. *BMC Bioinformatics*, 9:54 doi: 10.1186/1471-2105-9-54.

Literatur

- GABRIEL, K.R. (1969). Simultaneous Test Procedures - Some Theory of Multiple Comparisons. *The Annals of Mathematical Statistics* **40**, 224 - 250.
- GENZ, A. and BRETZ F. (2009). Computation of Multivariate Normal and t Probabilities. *Lecture Notes in Statistics* **195**. Springer, Heidelberg, New York.
- HASLER M. and HOTHORN L.A. (2008). Multiple Contrast Tests in the Presence of Heteroscedasticity. *Biometrical Journal* **50**, 793-800.
- KONIETSCHKE, F. (2009). Simultane Konfidenzintervalle für nichtparametrische relative Kontrasteffekte. *Dissertation, Georg-August-Universität Göttingen*
- RUMP, C. M. (2001). Strategies for Rolling the Efron dice. *Mathematics Magazine* **74**, 212-216.

Literatur

- RYU, E. and AGRESTI, A. (2008). Modeling and inference for an ordinal effect size measure. *Statistics in Medicine* **27**, 1703-1717.
- SCHAARSCHMIDT, F., SILL, M. and HOTHORN, L.A. (2008). Approximate Simultaneous Confidence Intervals for Multiple Contrasts of Binomial Proportions. *Biometrical Journal* **50**, 782-792.
- THANGAVELU, K. and BRUNNER, E. (2007). Wilcoxon Mann-Whitney Test for Stratified Samples and Efron's Paradox Dice. *Journal of Statistical Planning and Inference* **137**, 720-737.