

# **Multiples Testen**

## **-Simultane Konfidenzintervalle-**

Dr. Martin Scharpenberg

MSc Medical Biometry/Biostatistics

WiSe 2019/2020

## Konfidenzintervall für einzelnen Kontrast

## Konfidenzintervall für einzelnen Kontrast

- Betrachten  $k$  Gruppen mit Erwartungswert  $\mu = (\mu_1, \dots, \mu_k)$  und Varianz  $\sigma^2$
- $\bar{Y} = (\bar{Y}_1, \dots, \bar{Y}_k)$  Gruppenmittelwerte
- Sei  $\mathbf{c} = (c_1, \dots, c_k)$  mit  $c_1 + \dots + c_k = 0$  ein Kontrast
- Der Parameter  $\mathbf{c}^T \mu = \sum_{i=1}^k c_i \mu_i$  hat Schätzer  $\mathbf{c}^T \bar{Y} = \sum_{i=1}^k c_i \bar{Y}_i$
- Wir wissen, dass

$$\mathbf{c}^T \bar{Y} = \sum_{i=1}^k c_i \bar{Y}_i \sim N(\mathbf{c}^T \mu, \sigma^2 \sum_{i=1}^k \frac{c_i^2}{n_i})$$

- Damit folgt

$$\mathbf{c}^T \bar{Y} - \mathbf{c}^T \mu \sim N(0, \sigma^2 \sum_{i=1}^k \frac{c_i^2}{n_i})$$

**Test für  $H_0^\delta : \mathbf{c}^T \mu = \delta$** 

- Wenn  $\mathbf{c}^T \mu = \delta$ :

$$T_{\mathbf{c}}(\delta) = \frac{\mathbf{c}^T \bar{Y} - \delta}{\hat{\sigma} \sqrt{\sum_{i=1}^k \frac{c_i^2}{n_i}}} \sim t_{N-k}$$

- Verwerfen  $H_0^\delta : \mathbf{c}^T \mu = \delta$  falls  $|T_{\mathbf{c}}(\delta)| \geq t_\alpha = Q_{N-k}^t(1 - \alpha/2)$
- Nicht-verwerfen von  $H_\delta$  ist gleichbedeutend mit

$$|\mathbf{c}^T \bar{Y} - \delta| < t_\alpha \sqrt{\sum_{i=1}^k \frac{c_i^2}{n_i}} \hat{\sigma}$$

**Test für  $H_0^\delta : \mathbf{c}^T \mu = \delta$** 

- Dies lässt sich wie folgt umformen:

$$|\mathbf{c}^T \bar{Y} - \delta| < t_\alpha \sqrt{\sum_{i=1}^k \frac{c_i^2}{n_i} \hat{\sigma}}$$

$$\Leftrightarrow -t_\alpha \sqrt{\sum_{i=1}^k \frac{c_i^2}{n_i} \hat{\sigma}} < \mathbf{c}^T \bar{Y} - \delta < t_\alpha \sqrt{\sum_{i=1}^k \frac{c_i^2}{n_i} \hat{\sigma}}$$

$$\Leftrightarrow \mathbf{c}^T \bar{Y} - t_\alpha \sqrt{\sum_{i=1}^k \frac{c_i^2}{n_i} \hat{\sigma}} < \delta < \mathbf{c}^T \bar{Y} + t_\alpha \sqrt{\sum_{i=1}^k \frac{c_i^2}{n_i} \hat{\sigma}}$$

## Konfidenzintervall für einzelnen Kontrast

- Wir „sammeln“ alle  $\delta$  für die  $H_0^\delta$  auf dem Niveau  $\alpha$  nicht verworfen werden kann
- Dies führt zu dem Konfidenzintervall

$$CI = \left( \mathbf{c}^T \bar{Y} - t_\alpha \sqrt{\sum_{i=1}^k \frac{c_i^2}{n_i} \hat{\sigma}^2}, \mathbf{c}^T \bar{Y} + t_\alpha \sqrt{\sum_{i=1}^k \frac{c_i^2}{n_i} \hat{\sigma}^2} \right)$$

- Für die Überdeckungswahrscheinlichkeit gilt dann:

$$P(\mathbf{c}^T \mu \in CI) = P(|T_{\mathbf{c}}(\mathbf{c}^T \mu)| < t_\alpha) = 1 - \alpha$$

## Konfidenzintervall für einzelnen Gruppenvergleich

- Der Vergleich zweier Gruppen ist ein Spezialfall eines Kontrasttests ( $c_i = 1, c_j = -1, c_l = 0$  sonst)
- $(1 - \alpha)$ -Konfidenzintervall für  $\mu_i - \mu_j$ :

$$CI = \left( \bar{Y}_i - \bar{Y}_j - t_\alpha \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \hat{\sigma}, \bar{Y}_i - \bar{Y}_j + t_\alpha \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \hat{\sigma} \right)$$

- Für die Überdeckungswahrscheinlichkeit gilt wieder

$$P(\mu_i - \mu_j \in CI) = P(|T_{ij}(\mu_i - \mu_j)| < t_\alpha) = 1 - \alpha$$

## Simultane Konfidenzintervalle



## Simultane Konfidenzintervalle - Definition

- Untersuchen  $h$  Kontraste gleichzeitig

$$\mathbf{c}_j = (c_{j,1}, \dots, c_{j,k}), \quad j = 1, \dots, h$$

- Wollen  $h$  Konfidenzintervalle  $CI_1, \dots, CI_h$  berechnen, so dass für die simultane Überdeckungswahrscheinlichkeit gilt:

$$P\left(\mathbf{c}_j^T \mu \in CI_j \text{ für ALLE } j = 1, \dots, h\right) \geq 1 - \alpha$$

## Bonferroni Methode

## Bonferroni Methode

- Testen die Nullhypothesen

$$H_{\delta_j,j} : \mathbf{c}_j^T \mu_j = \delta_j, \quad j = 1, \dots, h$$

mit Kontrasttests auf dem Niveau  $\alpha/h$

- D.h., wir verwerfen  $H_{\delta_j,j}$  wenn

$$|T_{\mathbf{c}_j}(\delta_j)| = \frac{|\mathbf{c}_j^T \bar{Y} - \delta_j|}{\hat{\sigma} \sqrt{\sum_{i=1}^k c_{j,i}^2 / n_i}} \geq t_{\alpha/h}$$

mit  $t_{\alpha/h} = Q_{N-k}^t(1 - \frac{\alpha}{2h})$

## Bonferroni Methode

- Die Intervalle

$$CI_j^{Bonf} = \left( \mathbf{c}_j^T \bar{Y} \pm t_{\alpha/h} \hat{\sigma} \sqrt{\sum_{i=1}^k \frac{c_{j,i}^2}{n_i}} \right)$$

haben simultane Überdeckungswahrscheinlichkeit  $\geq 1 - \alpha$ , d.h.

$$P \left( \mathbf{c}_j^T \mu \in CI_j^{Bonf} \quad \text{für alle } j \right) \geq 1 - \alpha$$

## Beispieldatensatz 'cholesterol' - Wiederholung

- Datensatz `cholesterol` im R-Package `multcomp`:
- Vergleich von 5 Behandlungen zur Cholesterol-Reduktion in randomisierter Studie mit 10 Patienten pro Gruppe ( $N=50$ )
  - **Gruppe 1:** Neues Medikament, 1 mal 20 mg pro Tag
  - **Gruppe 2:** Neues Medikament, 2 mal 10 mg pro Tag
  - **Gruppe 3:** Neues Medikament, 4 mal 5 mg pro Tag
  - **Gruppe 4:** Kontrollmedikament D
  - **Gruppe 5:** Kontrollmedikament E.
- Wir interessieren uns nun für alle  $h = 5 \cdot 4/2 = 10$  Gruppenvergleiche

## Bonferroni-Methode – Beispiel 'Cholesterol' mit SAS

```
proc glm data=grvgl.cholest;  
class trt;  
model response = trt;  
means trt / cldiff Bon;  
run;
```

## Bonferroni-Methode – Beispiel 'Cholesterol' mit SAS

Comparisons significant at the 0.05 level  
are indicated by \*\*\*.

trt Comparison	Difference		Simultaneous 95%		
	Between Means		Confidence Limits		
drugE - drugD	5.586	1.325	9.847	***	
drugE - 4times	8.573	4.312	12.834	***	
drugE - 2times	11.723	7.462	15.984	***	
drugE - 1time	15.166	10.905	19.427	***	
drugD - 4times	2.986	-1.275	7.247		
drugD - 2times	6.136	1.875	10.397	***	
drugD - 1time	9.579	5.318	13.840	***	
4times - 2times	3.150	-1.111	7.411		
4times - 1time	6.593	2.332	10.854	***	
2times - 1time	3.443	-0.818	7.704		

## Bonferroni Methode – Vor- und Nachteile

- **Vorteile**
  - Kann auf beliebige (gültige) Konfidenzintervalle angewendet werden
  - Man kann  $h$  beliebige Kontraste untersuchen
- **Nachteile**
  - Man muss die zu testenden Kontraste vorab festlegen
  - Die multiple Überdeckungswahrscheinlichkeit ist meistens echt grösser  $1 - \alpha \rightarrow$  Es gibt oft andere Verfahren, mit engeren Konfidenzintervallen



## Scheffé Methode

## Scheffé-Konfidenzintervalle

- Mit  $d^{\text{Scheffé}} = \sqrt{(k-1)Q_{k-1, N-k}^F(1-\alpha)}$  haben alle Intervalle

$$CI_{\mathbf{c}}^{\text{Scheffé}} = \left( \mathbf{c}^T \bar{Y} \pm d^{\text{Scheffé}} \hat{\sigma} \sqrt{\sum_{i=1}^k \frac{c_i^2}{n_i}} \right)$$

- simultane Überdeckungswahrscheinlichkeit  $\geq 1 - \alpha$ , d.h.

$$P \left( \mathbf{c}^T \mu \in CI_{\mathbf{c}}^{\text{Scheffé}} \quad \text{für alle Kontraste } \mathbf{c} \right) \geq 1 - \alpha$$

## Scheffé-Methode – Beispiel 'Cholesterol' mit SAS

```
proc glm data=grvgl.cholest;  
class trt;  
model response = trt;  
means trt / cldiff Scheffe;  
run;
```

## Scheffé-Methode – Beispiel 'Cholesterol' mit SAS

Comparisons significant at the 0.05 level  
are indicated by \*\*\*.

trt Comparison	Difference				Bonferroni Confidence Lim.	
	Between Means	Simultaneous 95% Confidence Limits				
drugE - drugD	5.586	0.951	10.222	***	1.325	9.847
drugE - 4times	8.573	3.937	13.208	***	4.312	12.834
drugE - 2times	11.723	7.087	16.358	***	7.462	15.984
drugE - 1time	15.166	10.530	19.801	***	10.905	19.427
drugD - 4times	2.986	-1.649	7.622		-1.275	7.247
drugD - 2times	6.136	1.501	10.772	***	1.875	10.397
drugD - 1time	9.579	4.944	14.215	***	5.318	13.840
4times - 2times	3.150	-1.486	7.785		-1.111	7.411
4times - 1time	6.593	1.957	11.228	***	2.332	10.854
2times - 1time	3.443	-1.193	8.079		-0.818	7.704

## Tukey Methode

## Tukey Methode

- Suchen simultanes Konfidenzintervall für alle Paarvergleiche
- Mit entsprechender krit. Grenze  $d_{k,\alpha}^{Tukey}$  haben die Intervalle

$$CI_{ij}^{Tukey} = \left( \bar{Y}_i - \bar{Y}_j \pm d_{k,\alpha}^{Tukey} \hat{\sigma} \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \right)$$

simultane Überdeckungswahrscheinlichkeit  $\geq 1 - \alpha$ , d.h.,

$$P \left( \mu_i - \mu_j \in CI_{ij}^{Tukey} \text{ für alle } i, j \right) \geq 1 - \alpha$$

- Die krit. Grenze des zweiseitigen Tukey-Tests erfüllt

$$P_0 \left( \max_{1 \leq i < j \leq k} |T_{ij}| \geq d_{k,\alpha}^{Tukey} \right) = \alpha$$

## Tukey-Methode – Beispiel 'Cholesterol' mit SAS

```
proc glm data=grvgl.cholest;  
class trt;  
model response = trt;  
means trt / cldiff tukey;  
run;
```

## Tukey-Methode – Beispiel 'Cholesterol' mit SAS

Comparisons significant at the 0.05 level  
are indicated by \*\*\*

trt Comparison	Difference Between Means	Simultaneous 95% Confidence Limits			
drugE - drugD	5.586	1.485	9.688	***	
drugE - 4times	8.573	4.471	12.674	***	
drugE - 2times	11.723	7.621	15.824	***	
drugE - 1time	15.166	11.064	19.267	***	
drugD - 4times	2.986	-1.115	7.088		
drugD - 2times	6.136	2.035	10.237	***	
drugD - 1time	9.579	5.478	13.680	***	
4times - 2times	3.150	-0.951	7.251		
4times - 1time	6.593	2.492	10.694	***	
2times - 1time	3.443	-0.658	7.544		



## Tukey-Methode – Beispiel 'Cholesterol' mit R

```
> confint(amod_glht)
```

Simultaneous Confidence Intervals

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = response ~ trt, data = cholesterol)

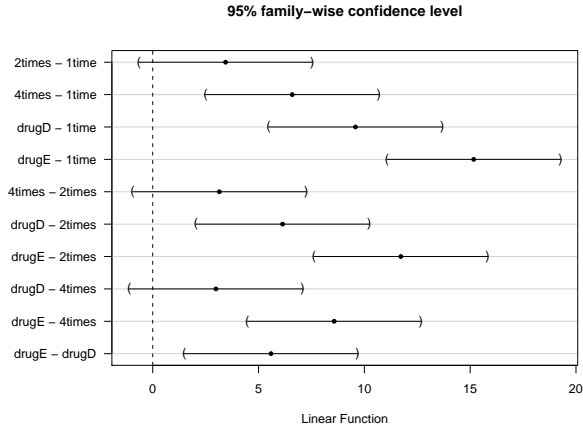
Estimated Quantile = 2.8417

95% family-wise confidence level

Linear Hypotheses:

	Estimate	lwr	upr
2times - 1time == 0	3.4430	-0.6586	7.5446
4times - 1time == 0	6.5928	2.4912	10.6944
drugD - 1time == 0	9.5792	5.4776	13.6808
drugE - 1time == 0	15.1656	11.0639	19.2672
4times - 2times == 0	3.1498	-0.9518	7.2514
drugD - 2times == 0	6.1362	2.0346	10.2378
drugE - 2times == 0	11.7226	7.6209	15.8242
drugD - 4times == 0	2.9864	-1.1152	7.0880
drugE - 4times == 0	8.5727	4.4711	12.6744
drugE - drugD == 0	5.5864	1.4847	9.6880

## Tukey's 95%-Konfidenzintervalle – Beispiel



## Tukey Methode – Vor- und Nachteile

- **Vorteile:**
  - Man bekommt immer engere Konfidenzintervalle als mit der Bonferroni-Methode
- **Nachteile:**
  - Tukey Methode kann nur bei homogener Varianz angewendet werden
  - Man kann nur die  $k$  Paarvergleichskontraste schätzen

## Vergleich von Bonferroni-, Tukey- und Scheffé-Methode

- Es gilt immer

$$CI_{ij}^{Tukey} \subseteq CI_{ij}^{Bonf} \quad \text{und} \quad CI_{ij}^{Tukey} \subseteq CI_{ij}^{Scheffé}$$

- Zwischen  $CI_{ij}^{Bonf}$  und  $CI_{ij}^{Scheffé}$  gibt es keine eindeutige Beziehung
- Die Beziehung ist abhängig von der Zahl der Gruppen und den Fallzahlen

## Dunnett Methode

## Dunnett Methode

### Dunnett-Konfidenzintervalle

Mit krit. Grenze  $d_{k,\alpha}^{Dunnett}$  des Dunnett-Tests haben die Intervalle

$$CI_i^{Dunnett} = \left( \bar{Y}_i - \bar{Y}_1 \pm d_{k,\alpha}^{Dunnett} \cdot \hat{\sigma} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_i}} \right), i = 2, \dots, k$$

simultane Überdeckungswahrscheinlichkeit  $\geq 1 - \alpha$ , d.h.,

$$P \left( \mu_i - \mu_1 \in CI_i^{Dunnett} \quad \text{für } i \geq 2 \right) \geq 1 - \alpha$$

## Beispiel: Effektivität von Eniporide – Wiederholung

Vergleich von 4 Dosen Eniporide zu Placebo bei akutem Herzinfarkt in randomisierter Studie mit insgesamt 430 Patienten

- **Gruppe 1:** Placebo (88 Pat.)
- **Gruppe 2:** 50 mg Eniporide (86 Pat.)
- **Gruppe 3:** 100 mg Eniporide (91 Pat.)
- **Gruppe 4:** 150 mg Eniporide (74 Pat.)
- **Gruppe 5:** 200 mg Eniporide (91 Pat.)

Primärer Endpunkt:  $\alpha$ -HBDH AUC (0 bis 72 Stunden)

## Beispiel: Effektivität von Eniporide – Originaldaten

Mittelwerte und Standardabweichungen für  $\alpha$ -HBDA:

Dosis:	0 mg	50 mg	100 mg	150 mg	200 mg
n:	88	86	91	74	91
MW:	44.2	45.3	40.2	33.9	43.9
STD:	26.0	31.8	22.5	20.5	27.0



## Dunnett-Methode – Beispiel Eniporide mit SAS

```
proc glm data=grvgl.zeym;  
class group;  
model HBDH = group;  
means group / cldiff dunnett;  
run;
```

Comparisons significant at the 0.05 level  
are indicated by \*\*\*

group Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
1 - 0	1.100	-8.559	10.759	
2 - 0	-4.000	-13.523	5.523	
3 - 0	-10.300	-20.347	-0.253	***
4 - 0	-0.300	-9.823	9.223	

## Dunnett-Methode – Beispiel Eniporide mit R

```
> confint(bmod_glht)
```

Simultaneous Confidence Intervals

Multiple Comparisons of Means: Dunnett Contrasts

Fit: aov(formula = HBDH ~ group, data = zeymer2)

Estimated Quantile = 2.4531

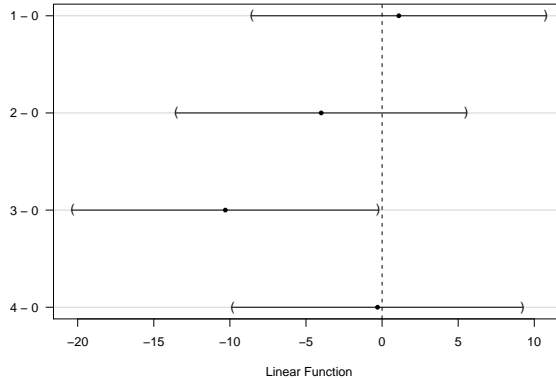
95% family-wise confidence level

Linear Hypotheses:

	Estimate	lwr	upr
1 - 0 == 0	1.1000	-8.5596	10.7596
2 - 0 == 0	-4.0000	-13.5245	5.5245
3 - 0 == 0	-10.3000	-20.3479	-0.2521
4 - 0 == 0	-0.3000	-9.8245	9.2245

## Dunnett's 95%-Konfidenzintervalle – Beispiel Eniporide mit R

95% family-wise confidence level



## Dunnett Methode – Vor- und Nachteile

- **Vorteile:**
  - Man bekommt immer engere Konfidenzintervalle als mit der Bonferroni- und Tukey-Methode
- **Nachteile:**
  - Dunnett Methode kann nur bei homogener Varianz angewendet werden.
  - Erlaubt keine anderen Vergleiche als mit  $\mu_1$  (oder einem anderen vorab gewählten  $\mu_i$ ).