

Multiples Testen

-Einführung-

Dr. Martin Scharpenberg

MSc Medical Biometry/Biostatistics

WiSe 2019/2020

Multipler Fehler mit unabhängigen Tests

- Angenommen h unabhängige stat. Tests
(z.B. h exp. Behandlungsgruppen, jeweils mit eigener Kontrollgruppe)
- Jeder Test auf dem Niveau $\alpha = 0.05$.
- Angenommen alle h Nullhypothesen sind wahr

Multipler Fehler 1. Art

Wahrscheinlichkeit mindestens einen Test irrtümlich zu verwerfen?

Antwort: $1 - (1 - \alpha)^h$

h	1	2	3	4	6	10	50
$1 - (1 - 0.05)^h$	0.05	0.10	0.14	0.19	0.27	0.40	0.92

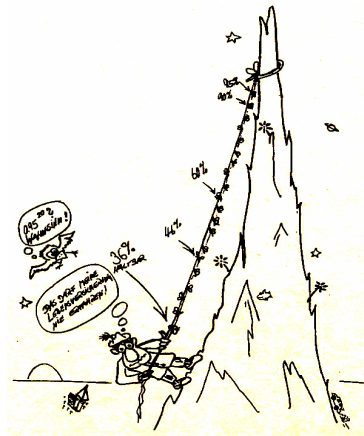
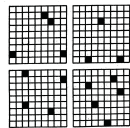
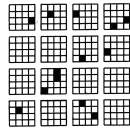
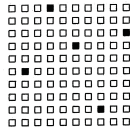


Abb. aus Beck-Bornholdt und Dubben, 1999

Typische Multiple Testprobleme (Gruppenvergleiche)

- Many-to-One Comparison (Vergleich mehrerer experimenteller Behandlungsgruppen zu einer Kontrolle)
- All-Pairwise Comparison (alle Paarvergleiche bei mehreren Behandlungsgruppen)
- Testen mehrerer Kontraste (lineare Kombinationen von Mittelwerten) im ANOVA Set-Up
- Testen mehrerer Kontraste im ANCOVA Set-Up
(z.B. Gruppenvergleiche unter Berücksichtigung von Kovariablen)

Typische multiple Testprobleme (Regressionsmodelle)

- Testen mehrerer Regressionskoeffizienten bzw. Linearkombinationen von Koeffizienten
- Testen mehrerer linearer Hypothesen
- Modell- und Variablenselektion

Typische multiple Testprobleme (klinische Studien)

- Multiple Endpunkte (“multiple endpoints”)
(mehrere primäre Endpunkte, ein primärer und wichtige sekundäre Endpunkte;
Genetik oder allg. “high throughput analysis”)
- Subgruppenanalysen, Enrichment Designs
- Gleichzeitiges Testen von Nichtunterlegenheit und Überlegenheit
- Dosisfindung (“dose finding studies”)
- Kombination aus den obigen Problemen

Charakteristika multipler Testprozeduren

- Single-Step-, Step-Down- und Step-Up-Prozeduren (meist basierenden auf p-Werten; genaueres später)
- Verfahren, die auf Abschlusstest basieren (Abschlusstests)
- Verfahren, die keine Annahmen über die Korrelationsstruktur benötigen
- Verfahren, die auf spezifischen Eigenschaften der Korrelationsstruktur beruhen
- Verfahren, die die Korrelationsstruktur komplett ausnutzen
- Resampling- und Permutationstests

Typische Fehlerkriterien

h	...	Zahl der Nullhypothesen (später kurz „Hypothesen“)
w	...	Zahl der wahren Nullhypothesen (i.d.R. unbekannt)
R	...	Zahl der verworfenen Nullhypothesen (Stichprobe)
V	...	Zahl der verworfenen wahren Nullhypothesen (Stichprobe; unbekannt)

Family-Wise-Error-Rate:

$$\text{FWER} = P(V > 0) = E(\mathbf{1}_{\{V > 0\}})$$

False-Discovery-Rate:

$$\text{FDR} = E(V/R)$$

Es gilt $\text{FWER} \geq \text{FDR}$, d.h. Kontrolle der FWER ist strenger.

Weitere Beispiel für Fehlerwahrscheinlichkeiten

h	...	Zahl der Nullhypothesen (später kurz „Hypothesen“)
w	...	Zahl der wahren Nullhypothesen (i.d.R. unbekannt)
R	...	Zahl der verworfenen Nullhypothesen (Stichprobe)
V	...	Zahl der verworfenen wahren Nullhypothesen (Stichprobe; unbekannt)

Verallgemeinerte Family-Wise-Error-Rate (lässt bis zu k Fehler zu):

$$\text{gFWER}(k) = P(V \geq k)$$

False-Discovery-Proportion:

$$\text{FDP} = V/R$$