

Multiples Testen

-p-Wert basierte Verfahren-

Dr. Martin Scharpenberg

MSc Medical Biometry/Biostatistics

WiSe 2019/2020

Set-Up (Annahmen; typischerweise erfüllt)

- $h \geq 2$ Nullhypothesen: H_0^1, \dots, H_0^h
- α Gesamtniveau (z.B. $\alpha = 0.025$ einseitig)
- Möchten familienweisen Fehler 1. Art (FWER) kontrollieren, d.h.

$$P(\text{Verwerfung von mindestens einem wahren } H_0^i) \leq \alpha$$

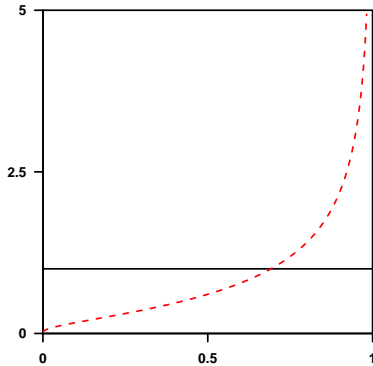
- Haben für jedes H_0^i einen p-Wert p_i , mit der Eigenschaft:

$$P(p_i \leq u | H_0^i) \leq u \quad \text{für alle } u \in [0, 1]$$

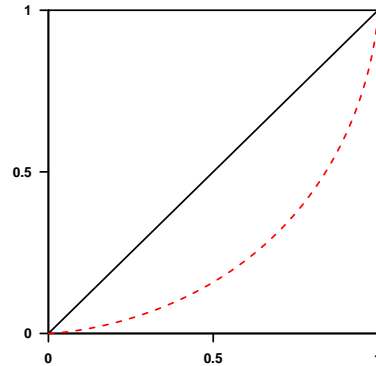
(*“konservative p-Werte”*)

Konservativer p-Werte

Dichte (unter H_0^i)

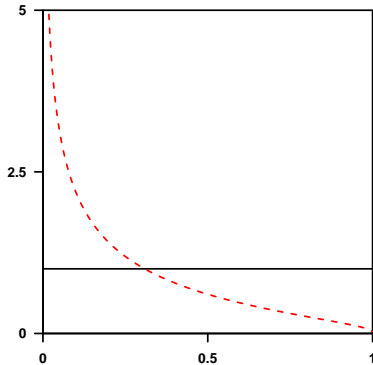


Verteilungsfunktion (unter H_0^i)

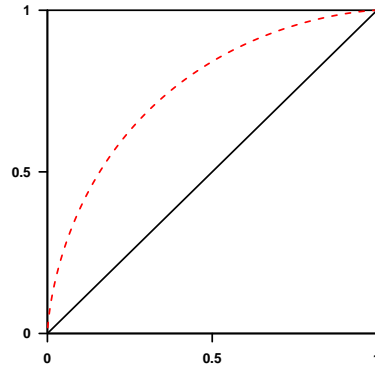


Antikonservativer p-Werte

Dichte (unter H_0^i)



Verteilungsfunktion (unter H_0^i)



Bsp. 1 (Many-to-one comparison)

Randomisierter Vergleich von 4 Dosen Eniporide zu Placebo; für akuten Myokardinfarkt mit insgesamt 430 Patienten.

- **Gruppe 0:** Placebo (88 Pat.)
- **Gruppe 1:** 50 mg Eniporide (86 Pat.)
- **Gruppe 2:** 100 mg Eniporide (91 Pat.)
- **Gruppe 3:** 150 mg Eniporide (74 Pat.)
- **Gruppe 4:** 200 mg Eniporide (91 Pat.)

Primärer Endpunkt:

α -HBDA, beschreibt Ausmaß des akuten Herzinfarkts

μ_i Erwartungswert des α -HBDA unter Dosis $i = 0, 1, \dots, 4$

$$H_0^1 : \mu_1 \geq \mu_0, \quad H_0^2 : \mu_2 \geq \mu_0, \quad H_0^3 : \mu_3 \geq \mu_0, \quad H_0^4 : \mu_4 \geq \mu_0$$

Möchten für mind. eine Dosis Überlegenheit zu Placebo zeigen

Bsp. 1 (Many-to-one comparison)

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	34.6
STD:	26.0	31.8	22.5	20.5	27.0

Mittelwerte und Standardabweichungen für α -HBDA

Bsp. 1 (Many-to-one comparison)

Gruppe	MW-Diff	Std. Error	t-Wert	p-Wert
50 mg	1.1	3.9	0.28	0.610
100 mg	-4.0	3.9	-1.03	0.152
150 mg	-10.3	4.1	-2.52	0.006
200 mg	-9.6	3.9	-2.46	0.007

t-Tests zum Vergleich zur Kontrollgruppe

Bsp. 2a (Multiple Endpunkte)

- Randomisierter Vergleich zweier Therapien zur Senkung des Blutdrucks bei Hypertonie-Patienten
- **Zwei gleichwertige Endpunkte:**
Systolischer und diastolischer Blutdruck,
 $Y_{sys,j}$ und $Y_{dia,j}$, $j = 1, 2$ die Therapie
- Erwartungswerte: $\mu_{sys,j} = E(Y_{sys,j})$ und $\mu_{dia,j} = E(Y_{dia,j})$
- $H_0^{sys} : \Delta_{sys} = \mu_{sys,2} - \mu_{sys,1} \leq 0$, $H_0^{dia} : \Delta_{dia} = \mu_{dia,2} - \mu_{dia,1} \leq 0$
- Haben zwei p-Werte: p_{sys} für H_0^{sys} und p_{dia} für H_0^{dia}
- Es genügt H_0^{sys} oder H_0^{dia} zu verwerfen.

Bsp. 2b (Multiple Endpunkte)

- Randomisierte Phase III Studie zum Nachweis der Wirksamkeit einer neuen Therapie
- **Ein primärer und ein wichtiger sekundärer Endpunkt:**
z.B. Lebensqualität (primär) und Überlebenszeit (sekundär)
- $Y_{p,j}$ und $Y_{s,j}$, $j = 1, 2$ die Therapie.
- Erwartungswerte: $\mu_{p,j} = E(Y_{p,j})$ und $\mu_{s,j} = E(Y_{s,j})$
- $H_0^p : \Delta_p = \mu_{p,2} - \mu_{p,1} \leq 0$, $H_0^s : \Delta_s = \mu_{s,2} - \mu_{s,1} \leq 0$
- Haben zwei p-Werte: p_p für H_0^p und p_s für H_0^s
- Wir möchten eine konfirmatorische Aussage auch für den sekundären Endpunkt machen können

Bsp. 3 (Nicht-Unterlegenheit und Überlegenheit)

- Randomisierte Studie zum Nachweis der Nicht-Unterlegenheit einer neuen Therapie zum Standard
- möchten zusätzlich auch Überlegenheit nachweisen können
- Y_j primärer Endpunkt in Behandlungsgruppe $j = 1, 2$
- Erwartungswerte: $\mu_j = E(Y_j)$, $j = 1, 2$, und $\Delta = \mu_2 - \mu_1$
- δ_{NI} der (vorab festgelegte) Nicht-Unterlegenheitsmargin

$$H_0^{NI} : \Delta \leq \delta_{ni}, \quad H_0^{Sup} : \Delta \leq 0$$

- Haben zwei p-Werte: p_{NI} für H_0^{NI} und p_{Sup} für H_0^{Sup}
- Es macht nur Sinn H_0^{Sup} zu testen, wenn H_0^{NI} bereits verworfen wurde (ansonsten logischer Widerspruch)

Bonferroni-Test

- h -Hypothesen H_0^i ($i = 1, \dots, h$) mit p -Werten p_i

Bonferroni-Test: Entscheidungsregel

Verwerfe H_0^i , falls $p_i \leq \alpha/h$

- **Beispiel 1:** $h = 4$, $p_1 = 0.61$, $p_2 = 0.152$, $p_3 = 0.006$, $p_4 = 0.007$, $\alpha = 0.025$, $\alpha/h = 0.00625$
- Können nur H_0^3 verwerfen, d.h. Wirksamkeit der 150 mg Dosis nachgewiesen.
- **Kontrolle des Fehlers 1. Art:**

$$P(V > 0) = P\left(\bigcup_{i \in W} \{p_i \leq \alpha/h\}\right) \leq \sum_{i \in W} P(p_i \leq \alpha/h) \leq |W| \cdot \alpha/h \leq \alpha,$$

wobei W die Indexmenge der wahren Nullhypothesen ist

Bonferroni-Holm-Test

- Ordne die p-Werte der Größe nach: $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(h)}$

Bem.: “(i)” steht für den Index des i größten p-Werts.

- **Step-Down-Algorithmus:**

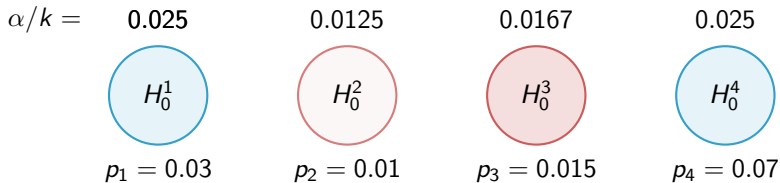
1. Verwerfe $H_0^{(1)}$, falls $p_{(1)} \leq \alpha/h$, sonst akzeptiere H_0^1, \dots, H_0^h und stopp. (Wie Bonferroni)
2. Verw. $H_0^{(2)}$, falls $p_{(2)} \leq \alpha/(h-1)$, sonst akzeptiere $H_0^{(2)}, \dots, H_0^{(h)}$ und stopp.
.....
- k. Verw. $H_0^{(k)}$, falls $p_{(k)} \leq \alpha/(h-k+1)$, sonst akzeptiere $H_0^{(j)}, j \geq k$, und stopp.
.....
- h. Verwerfe $H_0^{(h)}$, falls $p_{(h)} \leq \alpha$.

Bonferroni-Holm-Test

Bonferroni-Holm-Test: Entscheidungsregel (äquiv. zum Algorithmus)

verwerfe alle $H_0^{(k)}$ für $k \leq r_{BH}$ wobei

$$r_{BH} = \max\{l : p_{(j)} \leq \alpha/(h - j + 1) \text{ für alle } j \leq l\}$$

Bonferroni-Holm-Test / Beispiel ($\alpha = 0.05$, $h = 4$)

Wir verwerfen H_0^2 und H_0^3 und akzeptieren H_0^1 und H_0^4 .

Wir verwerfen immer mind. genauso viele Hypothesen wie mit dem Bonferroni-Test,

→ Der Bonferroni-Holm verbessert den Bonferroni gleichmäßig!

Kontrolliert die Bonferroni-Holm-Prozedur die FWER?

Fehlerkontrolle mit dem Bonferroni-Holm-Test

- Sei $p_{(k)}$ kleinster p-Wert aller wahren Hypothesen H_0^i , $i \in W$
- Aus $W \subseteq \{(k), (k+1), \dots, (h)\}$ folgt $|W| \leq (h - k + 1)$
- Weitere Beobachtungen:
 - Um mindestens ein wahres H_0^i zu verwerfen, muss $H_0^{(k)}$ im k -ten Schritt verworfen werden.
 - Schritte davor schränken Verwerfung eines wahren H_0^i nur ein.
 - $\{p_{(k)} = \min_{i \in W} p_i \leq \alpha / (h - k + 1)\} = \bigcup_{i \in W} \{p_i \leq \alpha / (h - k + 1)\}$

$$\begin{aligned} P(V > 0) &\leq P\left(p_{(k)} \leq \frac{\alpha}{(h - k + 1)}\right) = P\left(\bigcup_{i \in W} \left\{p_i \leq \frac{\alpha}{(h - k + 1)}\right\}\right) \\ &\leq \sum_{i \in W} P\left(p_i \leq \frac{\alpha}{(h - k + 1)}\right) \leq |W| \cdot \frac{\alpha}{h - k + 1} \leq \alpha \end{aligned}$$

Beispiel 1 mit Bonferroni-Holm-Test

- $h = 4$, $p_1 = 0.61$, $p_2 = 0.152$, $p_3 = 0.006$, $p_4 = 0.007$, $\alpha = 0.025$, $\alpha/h = 0.00625$
- Geordnete p-Werte:

$$p_{(1)} = 0.006 (H_0^3), \quad p_{(2)} = 0.007 (H_0^4), \quad p_{(3)} = 0.15 (H_0^2), \quad p_{(4)} = 0.6 (H_0^1)$$

- **Step-Down-Algorithmus:**

1. $p_{(1)} \leq \alpha/4 = 0.00625 \longrightarrow$ verwirfe H_0^3
2. $p_{(2)} \leq \alpha/3 = 0.00833 \longrightarrow$ verwirfe H_0^4
3. $p_{(3)} > \alpha/2 = 0.0125 \longrightarrow$ akzeptiere H_0^2 und H_0^1

- Können nun H_0^3 und H_0^4 verwerfen, d.h. Wirksamkeit von 150 mg und 200 mg Dosis nachgewiesen

Hochberg-Test

Ordne die p-Werte der Größe nach: $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(h)}$

Step-Up-Algorithmus:

1. Akzeptiere $H_0^{(h)}$, falls $p_{(h)} > \alpha$, sonst verw. alle H_0^1, \dots, H_0^h und stopp.
2. Akz. $H_0^{(h-1)}$, falls $p_{(h-1)} > \alpha/2$, sonst verw. alle $H_0^{(1)}, \dots, H_0^{(h-1)}$ u. stopp.
-
- k. Akz. $H_0^{(k)}$, falls $p_{(k)} > \alpha/(h - k + 1)$, sonst verw. alle $H_0^{(j)}, j \leq k$, und stopp.
-
- h. Verwerfe $H_0^{(1)}$, falls $p_{(1)} \leq \alpha/h$.

Hochberg-Test

Hochberg-Test: Entscheidungsregel (äquiv. zum Algorithmus)

verwerfe alle $H_0^{(k)}$ für $k \leq r_H$ wobei $r_H = \max\{l : p_{(l)} \leq \alpha/(h - l + 1)\}$

Es gilt immer: $r_H \geq r_{BH} = \max\{l : p_{(j)} \leq \alpha/(h - j + 1) \text{ für alle } j \leq l\}$

Beispiel 1 mit Hochberg-Test

- $h = 4$, $p_1 = 0.61$, $p_2 = 0.152$, $p_3 = 0.006$, $p_4 = 0.007$, $\alpha = 0.025$, $\alpha/h = 0.00625$
- Geordnete p-Werte:

$$p_{(1)} = 0.006 (H_0^3), \quad p_{(2)} = 0.007 (H_0^4), \quad p_{(3)} = 0.15 (H_0^2), \quad p_{(4)} = 0.6 (H_0^1)$$

- **Step-Up-Algorithmus:**

1. $p_{(4)} > \alpha = 0.025 \rightarrow$ akzeptiere H_0^1 ,
2. $p_{(3)} > \alpha/2 = 0.0125 \rightarrow$ akzeptiere H_0^2 ,
3. $p_{(2)} \leq \alpha/3 = 0.00833 \rightarrow$ verwirfe H_0^4 und H_0^3

- Können wieder H_0^3 und H_0^4 verwerfen
- **Bemerkung:** Wir hätten H_0^3 auch bei $p_3 = p_4 = 0.007$ verwerfen können (dann aber nicht mehr mit dem Bonferroni-Holm-Test)

Hochberg-Test für zwei Hypothesen

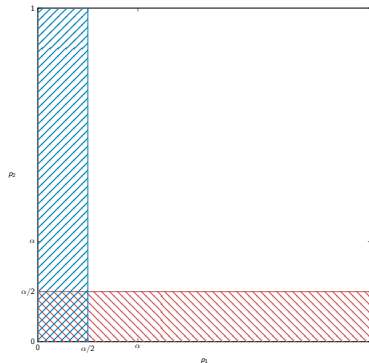
- Zwei Hypothesen H_0^1 und H_0^2 mit p-Werten p_1 und p_2 .
- **Verwerfungsregel: (äq. zum Step-Up-Alg.)**
 1. Verwerfe nur H_0^1 , falls $p_1 \leq \alpha/2$ (wie bei Bonferroni).
 2. Verwerfe nur H_0^2 , falls $p_2 \leq \alpha/2$ (wie bei Bonferroni).
 3. Verwerfe H_0^1 und H_0^2 , wenn $p_1 \leq \alpha$ und $p_2 \leq \alpha$ (zusätzl. zu Bonf.)

Hochberg-Test für zwei Hypothesen

Multipler Fehler bei unabhängigen und gleichverteilten p-Werten (beide Hypothesen wahr)

Bonferroni (und B.-Holm)

$$\begin{aligned} P(V > 0) &= P(p_1 \leq \alpha/2 \vee p_2 \leq \alpha/2) \\ &= P(p_1 \leq \alpha/2) + P(p_2 \leq \alpha/2) \\ &\quad - P(p_1 \leq \alpha/2 \wedge p_2 \leq \alpha/2) \\ &= \frac{\alpha}{2} + \frac{\alpha}{2} - \frac{\alpha^2}{4} < \alpha \end{aligned}$$

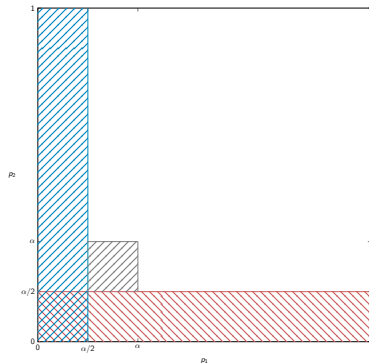


Hochberg-Test für zwei Hypothesen

Multipler Fehler bei unabhängigen und gleichverteilten p-Werten (beide Hypothesen wahr)

Hochberg

$$\begin{aligned} P(V > 0) &= P(p_1 \leq \alpha/2 \vee p_2 \leq \alpha/2) \\ &\quad + P(\alpha/2 < p_1 \leq \alpha \wedge \alpha/2 < p_2 \leq \alpha) \\ &= \frac{\alpha}{2} + \frac{\alpha}{2} - \frac{\alpha^2}{4} + \frac{\alpha^2}{4} = \alpha \end{aligned}$$



Fehlerkontrolle mit dem Hochberg-Test

Es gilt nicht immer $P(V > 0) \leq \alpha$, allerdings wenn z.B.

- die p-Werte p_1, \dots, p_n stochastisch unabhängig sind,
- die p-Werte aus multivariat-normalverteilten Teststatistiken berechnet werden, deren paarweise Korrelationen alle nicht-negativ sind,
- im Beispiel und allgemeiner bei Many-to-One-Vergleichen, wenn die Annahmen einer ANOVA gelten
- unter einer allgemeineren (dennoch spezifischen) positiven Abhängigkeitsstruktur.

Gewichteter Bonferroni-Test

- Hypothesen H_0^1, \dots, H_0^h . Fixiere Gewicht $0 \leq w_i \leq 1$ für jedes H_0^i , wobei

$$\sum_{i=1}^h w_i = 1.$$

- Je größer w_i , desto mehr Niveau wird an H_0^i vergeben.
- Regel:** Verwerfe H_0^i , falls $p_i \leq \alpha_i = \alpha w_i$.
- Beispiel:**
 - $\alpha = 0.05, \quad w_1 = 0.4, w_2 = w_3 = w_4 = 0.2,$
 - $\alpha = 0.05, \quad \alpha_1 = 0.02, \alpha_2 = \alpha_3 = \alpha_4 = 0.01,$
 - $h = 4, \quad p_1 = 0.03, p_2 = 0.009, p_3 = 0.011, p_4 = 0.07$
 - Können nur H_0^2 verwerfen.

Gewichteter Bonferroni-Holm-Test für H_0^1, \dots, H_0^h

Fixiere $0 \leq w_1, \dots, w_h \leq 1$ mit $\sum_{i=1}^h w_i = 1$. Es sei nun $q_i = p_i/w_i$ und $q_{(1)} \leq \dots \leq q_{(h)}$. Seien $p_{(1)}, \dots, p_{(h)}$ und $w_{(1)}, \dots, w_{(h)}$ die zugehörigen p-Werte und Gewichte.

1. Wenn $p_{(1)} \leq \alpha w_{(1)}$ verwerfe $H_0^{(1)}$, sonst akz. alle H_0^i und stop. (Wie Bonf.)
2. Wenn $p_{(2)} \leq \alpha w_{(2)} / (\sum_{j=2}^h w_{(j)})$ verwerfe $H_0^{(2)}$, sonst akz. alle $H_0^{(j)}$, $j \geq 2$, und stop.
- ...
- k. Wenn $p_{(k)} \leq \alpha w_{(k)} / (\sum_{j=k}^h w_{(j)})$, verwerfe $H_0^{(k)}$ und akz. alle $H_0^{(j)}$, $j \geq k$.
- ...
- h. Wenn $p_{(h)} \leq \alpha$, dann verwerfe $H_0^{(h)}$, ansonsten akzeptiere $H_0^{(h)}$.

Übung: Zeige Kontrolle des Fehlers 1. Art (erst gewichteter Bonferroni-Test, dann gewichteter Holm-Test)

Gewichteter Bonferroni-Holm-Test

$$\alpha = 0.05, w_1 = 0.4, w_2 = w_3 = w_4 = 0.2,$$

$$\text{Notation: } \alpha_i^J = \alpha w_i / \sum_{j \in J} w_j \text{ für } i \in J \subseteq \{1, 2, 3, 4\}$$

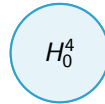
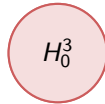
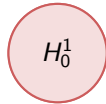
Vergleich zum gew. Bonferroni:

$$\alpha_i^{\{(k), \dots, (h)\}} = 0.0333$$

$$0.01$$

$$0.0125$$

$$0.05$$



$$p_1 = 0.03$$

$$p_2 = 0.009$$

$$p_3 = 0.011$$

$$p_4 = 0.07$$

Wir verwerfen H_0^1 , H_0^2 und H_0^3 und akzeptieren H_0^4 .

Wir verwerfen immer mind. genauso viele Hypothesen wie mit dem gew. Bonferroni-Test,

→ der gew. Bonferroni-Holm verbessert den gew. Bonferroni gleichmäßig!