Multiples Testen

Wolfgang Mader

12. Juli 2007

Gliederung

Problemstellung

- eine Fragestellung
- einfaches Datenmaterial
- mehrere **unabhängige** Tests zum Niveau α
- versuchsbezogene Irrtumswahrscheinlichkeit: $\alpha_{\mathrm{v}} > \alpha$

  $\Rightarrow$ Fehler erster Art

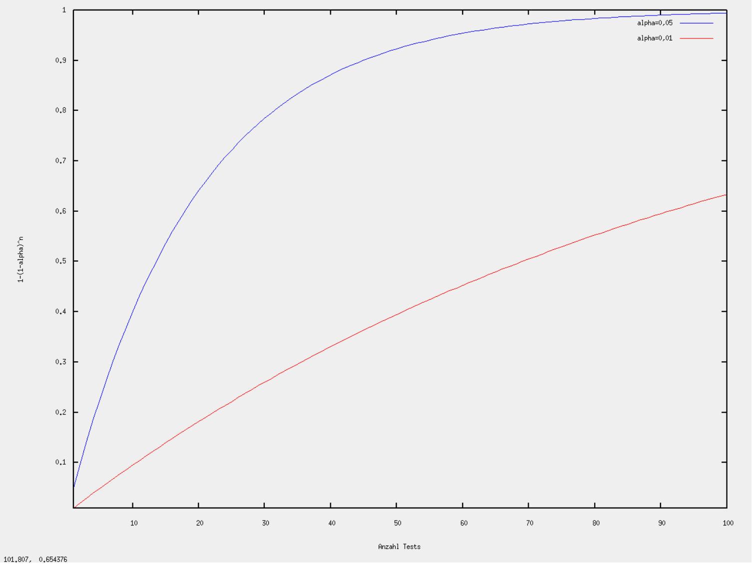
Änderung des Signifikanzniveaus

n	α	α_v
1	0,05	0,05
2	0,05	0,10
20	0,05	0,64
100	0,05	0,99

- $P(\text{korrekt abzulehnen}) \begin{cases} \text{für einen Test:} & 1-\alpha \\ \text{für n Tests:} & (1-\alpha)^n \end{cases}$
- $\alpha_\mathrm{v} = 1 - (1-\alpha)^n$

Verlauf von α_V mit n

$\alpha = 0,05$ $\alpha = 0,01$



Bonferroni–Methode

Signifikanzniveau

Ziel:

- Kontrolle der versuchsbezogenen Irrtumswahrscheinlichkeit

  $\Rightarrow \alpha_{\mathrm{v}} \overset{!}{=} \alpha_{\mathrm{ad}}$

Methode:

- setze $\alpha = 1 - (1 - \alpha_{\mathrm{ad}})^{\frac{1}{n}}$
- oder genähert: $\alpha = \alpha_{\mathrm{ad}}/n$

  $\Rightarrow \alpha_{\mathrm{v}} = 1 - (1 - 1 + (1 - \alpha_{\mathrm{ad}})^{\frac{1}{n}})^{n} = \alpha_{\mathrm{ad}}$

Bonferroni–Methode

p-Wert

- $p = 1 - F_{H^0}(X)$
- rechtsseitiger Test
- falls $p \leq \alpha$ verwerfe H^0

Methode

- adjustiere p-Wert: $p_{\text{ad}} = n \cdot p$
- multiples Signifikanzniveaus bleibt lokales Signifikanzniveau

Probleme der Bonferroni–Methode

Relevanz

- Menge von Hypothesen $\{H_i^0 : i = 1, \ldots, n\}$
- Globalhypothese: $H^0 = \bigcap_{i=1}^n H_i^0$
- lehne H^0 ab, falls ein H_i^0 abgelehnt
- keine Kenntnis, welche H_i^0 abgelehnt
- keine Kenntnis, wieviele H_i^0 abgelehnt

- Wer möchte sowas wissen?

Probleme der Bonferroni–Methode

Fehler zweiter Art

- vergleiche 20 unabhängige Testgrößen: $n = 20$
- multiples Signifikanzniveau: $\alpha_{\mathrm{v}} \stackrel{!}{=} 0.05$

  $\Rightarrow \alpha = 0,0026$

  $\Rightarrow$ Test sehr konservativ

  $\Rightarrow$ viele Fehler zweiter Art

Probleme der Bonferroni–Methode

Welche Tests sind betroffen?

- alle Tests zu einem Paper …

  …auch die nicht veröffentlichten?
- alle Tests des Journals?
- alle Tests des Autors?

- Erinnere: **viele** Tests **ein** Ergebniss

Wann macht Bonferroni–Methode Sinn?

- bei Interesse an Globalhypothese
- bei Wiederholungen in vielen Untergruppen

  z.B.: Chargen
- explorativer Versuch

  generiert Hypothesen

  Microarray

Bonferroni–Methode

Gewinne und Verluste

- Kontrolle über multiples Signifikanzniveau

- keine Aussage, welche H_i^0 verworfen
- Test wird mit n konservativer
- unklare philosophische Sachlage

Bonferroni–Holm–Methode

alternative Methode

- ordne p-Werte: $p_{(1)} \leq \cdots \leq p_{(n)}$
- adjustiere p-Werte: $p^{\text{ad}}_{(i)} = (n - i + 1) \cdot p_{(i)} \quad i = 1, \ldots, n$
- schrittweise Prozedur
  - $p^{\text{ad}}_{(k)} < \alpha \Rightarrow$ wiederhole mit $p^{\text{ad}}_{(k+1)}$ $\qquad 1 \leq k \leq n$
  - falls $p^{\text{ad}}_{(k)} \geq \alpha$

    $\Rightarrow$ verwerfe $H^0_{(1)}, \ldots, H^0_{(k)}$, behalte Rest

    $H^0_{(i)}$ sind geordnet wie $p_{(i)}$

Bonferroni–Holm–Methode

Gewinne und Verluste

- generell
- korrekt
- gleichmäßig besser als Bonferroni

- logische und stochastische Abhängigkeiten nicht beachtet
  - hier gibt es noch bessere Ansätze
  - Shaffer
  - Hommel und Bernhard
  - Westfall

False Discovery Rate
Überblick

- eingeführt von Benjamini und Hochberg (1995)
- kontrolliert FDR: $E\left(\frac{\text{\# falscher Ablehnungen}}{\text{\# aller Ablehnungen}}\right)$
- kontrolliert nicht versuchsbezogene Irrtumswahrscheinlichkeit (FWE)
- gilt für unabhängige Teststatistiken

Benjamini–Hochberg–Prozedur

Notation

- m Hypothesen: $\{H_1^0, H_2^0, \ldots, H_m^0\}$
  davon sind m_0 wahr
- Zufallsvektor: $\{X_1, X_2, \ldots, X_m\}$
- zugehörige p-Werte: $\{P_1, P_2, \ldots, P_m\}$

Benjamini–Hochberg–Prozedur

- Verfahren kontrolliert FDR zum Level $q\frac{m_0}{m} \leq q$
- ordne p-Werte: $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$
- definiere:

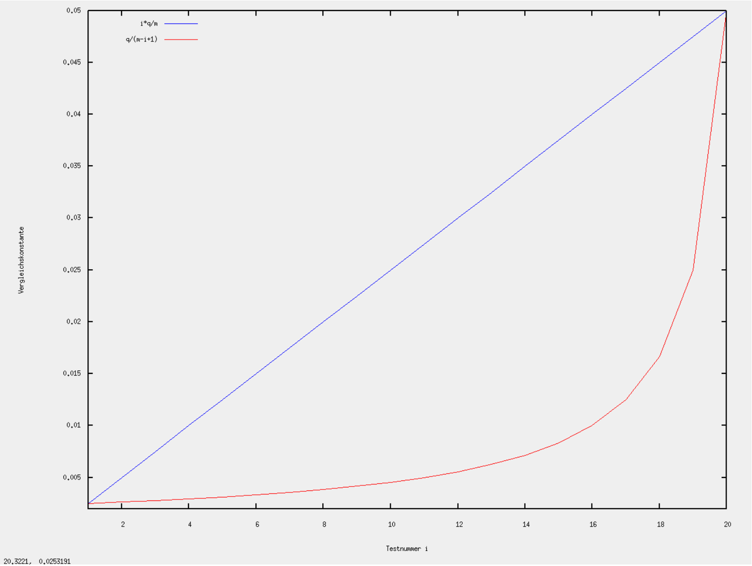
$$k = \max\left\{i : p_{(i)} \leq \frac{i}{m}q\right\}$$

- verwerfe $H^0_{(1)} \ldots H^0_{(k)}$

- ist Kontrolle der FWE erwünscht, benutze

$$k = \max\left\{i : p_{(i)} \leq \frac{q}{m-i+1}\right\}$$

FDR vs. FWE



i*q/m
q/(m-i+1)
0,05
0,045
0,04
0,035
0,03
0,025
0,02
0,015
0,01
0,005
Vergleichskonstante
2
4
6
8
10
12
14
16
18
20
Testnummer i
20,3221, 0,0253191

Die Wahl von q

- sei V Anzahl der falsch positiven Entscheidungen
- sei R Anzahl aller Ablehnungen
- definiere: $Q = \begin{cases} V/R, & \text{für } R > 0, \\ 0, & \text{sonst} \end{cases}$
- FDR ist $E(Q)$
- für $R > V \Rightarrow$ FWE konservativer als FDR
- für $R = V$, d.h. Globalhypothese ist wahr $\Rightarrow$ FDR $\hat{=}$ FWE

  $\Rightarrow$ wähle z.B. $q = \alpha$

Benjamini–Hochberg mit Abhängigkeit

Theorem

Falls die gemeinsame Verteilung der Teststatistiken der wahren H_0 PRDS ist, so kontrolliert die Benjamini–Hochberg–Prozedur die FDR zum Niveau $\leq \frac{m_0}{m} \cdot q$.

Definition

PRDS: Für jede zunehmende Menge D und für jedes $i \in I_0$ ist $P(\boldsymbol{X} \in D | X_i = x)$ monoton steigend in x.

PRDS: positive regression dependency on each one from a subset I_0

Verteilung mit PRDS–Eigenschaft

Bsp.: Multivariat normalverteilte Teststatistik

- sei $\boldsymbol{X} \sim N(\mu, \Sigma)$ Zufallsvektor der Teststatistik
- Test auf $H_i^0 : \mu_i = 0 \qquad i = 1, \ldots, m$
- $\mu_i = \begin{cases} 0, & i \in I_0, \\ > 0, & \text{sonst} \end{cases}$
- Falls für alle $i \in I_0$ und für alle $j \neq i$ gilt: $\Sigma_{ij} \geq 0$ ist X PRDS bezüglich I_0.

Anwendungsgebiete

Vergleich von Untergruppen

- Vergleiche neue Behandlung mit herkömmlicher
- Interessant: Unterschiede in Subkategorien?
- gewöhnlich Mittelwert innerhalb der Subkategorien
  $\Rightarrow$ unabhängige, normalverteilte Teststatistiken

- für einseitige t-Tests ex. Abb. so daß $\boldsymbol{X}$ PRDS

Anwendungsgebiete
explorative Experiment

- Aufgabe: suche Einflüsse auf eine Größe
- z.B. Pharmaindustrie oder Qualitätsmanagement
- hier kann q großzügig gewählt werden

- zweiseitige t-Tests führt zu $\boldsymbol{X}$ mit PRDS

Anwendungsgebiete

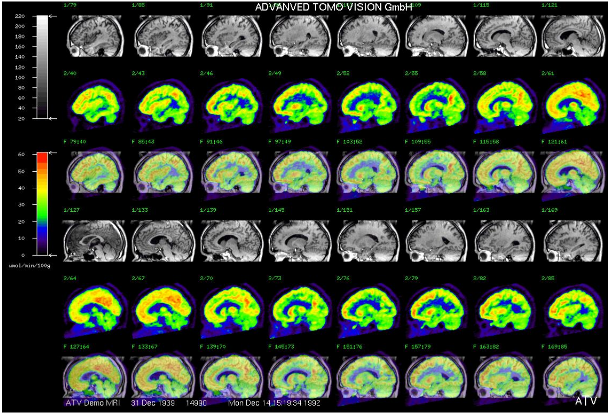
Many–to–one Vergleiche

- Vergleich von mehreren Behandlungen mit einer Kontrollgruppe
- suche Empfehlung für beste Behandlung $\Rightarrow$ FWE
- suche Anhalt für neue Forschung $\Rightarrow$ FDR

- Test ist einseitiger t-Test
  $\Rightarrow$ $\boldsymbol{X}$ ist PRDS

Übersicht

- Problem kommt aus der Medizin (z.B. Cerebral Blood Flow)
- Daten sind PET–Bilder
- von Interesse sind Unterschiede in den Bildern



Modellierung der Daten

- PET–Bilder von n Probanden in m Situationen
- $C_{ij}(x, y, z)$ ist CBF von Proband i in Situation j
- G_{ij} ist Mittelwert aller betrachteter Voxel
- $Y_{ij}(x, y, z) = \frac{C_{ij}}{G_{ij}}$

Bedeutung von $\Delta(x, y, z)$

- z.B. eine Zwei–Situationen–Subtraktion:

$$\Delta(x, y, z) = \sum_{i=1}^{n} \{Y_{i2}(x, y, z) - Y_{i1}(x, y, z)\}/\sqrt{2n}$$

- $\Delta(x, y, z)$ ist ein verallgemeinertes ΔCBF–Bild
- wie schätzt man die Varianz von $\Delta(x, y, z)$?

Varianz von Δ

- mehrere Möglichkeiten
- Summation über Situationen $\Rightarrow s^2(x, y, z)$
- Summation über Voxel $\Rightarrow S_c^2$
- Summation über beides

$\Rightarrow$ erhalte Schätzer für Varianz:

$$s_c^2(x, y, z) = \{S_c^2/S^2\}s^2(x, y, z)$$

$$S = \sum_{x,y,z} s^2(x, y, z)/N$$

Das statistische Feld

- $T_c(x, y, z) = \Delta(x, y, z)/s_c(x, y, z)$
- ist näherungsweise t-verteilt mit $\nu = (n-1)(m-1)$ Freiheitsgraden

- Aufgabe: finde signifikante Unterschiede der Signale im statistischen Feld

  $\Rightarrow$ finde Maxima

Euler–Charakteristik und excursion set

- excursion set: Menge der Pixel über Schwelle t
- EC: zählt zusammenhängende Gebiete auf excursion set
- M sei das Maximum im excursion set in der jeweiligen Suchregion
- wie ist t zu wählen um Fehler erster Art zu vermeiden?

Einheitsformel für p-Wert

- EC findet $\begin{cases} 1, & \text{für } t < M, \\ 0, & \text{sonst} \end{cases}$
- somit ist $E(EC)$ näherungsweise der p-Wert von M

$$P(M \geq t) \approx \sum_{d=0}^{3} R_d(V)\,\varrho_d(t)$$

- $R_d(V)$: einheitenlose Größe, abhänig von Dimension und Form der Suchregion ($\rightarrow$ Tabellen)
- $\varrho_d(t)$: EC Dichte, abhänig vom Schwellwert t ($\rightarrow$ Tabellen)

Schönes Wochenende

Danke für Eure Aufmerksamkeit