Übung zur Vorlesung

Multiples Testen

Jun.-Prof. Dr. Thorsten Dickhaus
RUD25, Raum 1.203
www.math.hu-berlin.de/~dickhaus/
dickhaus@math.hu-berlin.de
Tel.: 030/2093-5841

Humboldt-Universität zu Berlin
Institut für Mathematik
Sommersemester 2010

Blatt 1

Aufgaben

1. **Strukturierte Hypothesensysteme**

   Konstruieren Sie je ein Hypothesensystem, in welchem

   (a) die Menge der Maximalhypothesen ungleich der Menge der Globalhypothesen ist.

   (b) die Menge der Minimalhypothesen ungleich der Menge der Elementarhypothesen ist.

2. **Kohärenz multipler Tests**

   (a) Beweisen Sie Lemma 1.17 aus der Vorlesung.

   (b) Beweisen Sie Teil (a) von Lemma 1.18 aus der Vorlesung.

3. **Erwartete Anzahl von Typ I Fehlern**

   Für einen multiplen Test φ für das multiple Testproblem $(\Omega, \mathcal{A}, \mathcal{P}, \mathcal{H})$ mit $\mathcal{H} = \{H_i, i \in I = \{1, \ldots, m\}\})$ bezeichne die Zufallsgröße

   $$V_\varphi(\vartheta) := \sum_{i \in I_0(\vartheta)} \varphi_i$$

   die Anzahl der fälschlicherweise verworfenen Nullhypothesen. Beweisen Sie die folgenden Ungleichungen:

   $$\forall \vartheta \in \Theta : \frac{\mathbb{E}_\vartheta[V_\varphi(\vartheta)]}{m_0} \leq \mathbb{P}_\vartheta\Big(\bigcup_{i \in I_0(\vartheta)} \{\varphi_i = 1\}\Big) \leq \mathbb{E}_\vartheta[V_\varphi(\vartheta)].$$

   Dabei sei wie üblich $m_0 = |I_0(\vartheta)|$ die Anzahl wahrer Nullhypothesen.

4. **Durchführung eines multiplen Tests in Handrechnung**

   In einer Untersuchung zum Fettgehalt in sogenannten „Light“-Butterprodukten wird für drei Buttersorten der Fettgehalt Y in g pro 100g festgestellt. Von jeder Sorte werden vier Proben betrachtet. Zusätzliche Effekte auf den Fettgehalt der Proben bleiben unberücksichtigt. Zur Analyse wird das lineare Modell $y_{ij} = \beta_i + \varepsilon_{ij}$, $i = 1, 2, 3$, $j = 1, \ldots, 4$ mit $\varepsilon_{ij} \sim \mathcal{N}(\mu, \sigma^2)$ iid., $\sigma^2 > 0$ unbekannt, unterstellt. Folgende Ergebnisse werden notiert:

   Fettgehalt in g je 100g Butter

Sorte 1	Sorte 2	Sorte 3
61	62	65
58	59	62
60	61	63
60	61	62

(a) Stellen Sie das durchschnittsabgeschlossene Hypothesensystem zu der Fragestellung auf, ob zwischen den Sorten bezüglich des Fettgehalts Unterschiede bestehen. Testen Sie jede der resultierenden Nullhypothesen zum lokalen Niveau $\alpha = 0.05$.

(b) Treffen Sie eine Aussage zur Widerspruchsfreiheit der erzielten Testentscheidung.

(c) Was lässt sich über Kohärenz und Konsonanz dieser Entscheidung sagen?

5. **Programmieraufgabe**

In dem Artikel Notterman, D. A., Alon, U., Sierk, A. J. (2001). Transcriptional Gene Expression Profiles of Colorectal Adenoma, Adenocarcinoma, and Normal Tissue Examined by Oligonucleotide Arrays. Cancer Research 61, 3124-3130, finden sich publizierte Daten aus einem Krebs-Forschungsprojekt. Das Ziel der Studie war, differentiell exprimierte Gene und R(D)NA-Profile in Tumorgewebe im Vergleich mit normalem (gesundem) Gewebe zu finden.

Dazu wurde eine klinische Studie mit $n = 22$ Krebspatienten durchgeführt. Wir betrachten hier nur den Teildatensatz der 18 Patienten mit Adenokarzinom. Von diesen 18 Individuen wurden Genexpressionsdaten („Intensitäten“) für 7457 verschiedene RNA-, DNA- und Genorte erhoben, und zwar jeweils einmal in einer Gewebeprobe mit Tumor und einmal in einer gesunden Gewebeprobe. Die zugehörigen (leicht aufbereiteten) Daten sind Bestandteil des `mutoss`- Zusatzpakets für `R`.

In solchen QTL- (quantitative trait loci) Analysen wird typischerweise eine Log-Normalverteilung für die Intensitätsquotienten angenommen. Nach einigen Vorverarbeitungsschritten (siehe „Materials and Methods“ in dem o.a. Artikel) wurden daher zum Vergleich der beiden Gruppen gepaarte t-Tests für verbundene Stichproben auf den durch den natürlichen Logarithmus transformierten Daten vorgeschlagen.

(a) Laden Sie die Nutzdaten in `R`. Vollziehen Sie die p-Werte für die zweiseitigen t-Tests nach.

(b) Welche RNA-, DNA- bzw. Genorte zeigen nach Bonferroni-Adjustierung eine zum multiplen Niveau $\alpha = 0.05$ signifikante differentielle Expression in Tumor- im Vergleich zu gesundem Gewebe? Geben Sie die entsprechenden Indizes an.