Exercises for the lecture on

Statistical Methods

Humboldt-University Berlin
Department of Mathematics
Winter term 2013 / 2014

Prof. Dr. Vladimir Spokoiny
Vladimir.Spokoiny@wias-berlin.de
Dr. Thorsten Dickhaus
Thorsten.Dickhaus@wias-berlin.de
www.wias-berlin.de/people/dickhaus/

Sheet 9

Solutions are due on Monday, December 16th, 2013, 3:15pm.
Every completely and correctly solved exercise gives 4 points.

Exercises

33. **Global and local inference in the ANOVA1 model.**

(a) Prove Theorem 3.40. Hint: Use Theorem 3.32 and Corollary 3.34.

(b) Consider Model 3.38 (one-factorial analysis of variance) and assume that all pairwise mean differences shall be tested simultaneously (with two-sided "local" tests) for deviation from zero. Hence, we have a whole family $\mathcal{H} = \{H_{ij} : 1 \leq i < j \leq k\}$ of null hypotheses. Every single pair of null and alternative hypothesis in $\mathcal{H}$ is given by

$$H_{ij} : \{\mu_i = \mu_j\} \quad \text{versus} \quad K_{ij} : \{\mu_i \neq \mu_j\}, \quad 1 \leq i < j \leq k.$$

(i) Determine the cardinality of $\mathcal{H}$.

(ii) By making use of the general theory of multiple linear regression, represent the likelihood ratio test φ_{ij} at (local) significance level $\alpha_{\text{loc.}}$ for testing H_{ij} versus K_{ij} as an F-test and, equivalently, as a t-test.

(iii) How can $\alpha_{\text{loc.}}$ in part (b) be chosen such that for all possible parameter vectors $\vartheta = (\mu_1, \ldots, \mu_k, \sigma^2)^\top$ the probability for no type I error is at least $(1 - \alpha)$ for a given constant $\alpha \in (0, 1)$? Derive the resulting critical value for the local two-sided t-tests as a function of $n_\bullet$, k, and α.

Hint: Take another look at Exercise 22.(b).

34. **ANOVA2 by hand.** In a textbook, V. K. Rohatgi (1976, page 522) describes an agricultural trial. Four fertilizers A, B, C and D were tested on three wheat varieties. The dependent variable (i. e., response) was the yield on one respective parcel of land per factor level combination (first factor: fertilizer, second factor: variety). On a standardized scale, the collected response data are given in the following table.

Wheat variety:	I	II	III
Fertilizer A	8	3	7
Fertilizer B	10	4	8
Fertilizer C	6	5	6
Fertilizer D	8	4	7

(a) Why can interactions between the factors not be tested by making use of Theorem 3.50 and Remark 3.51?

(b) Ignore all possible interactions and consider a reduced model of the form

$$Y_{ij} = \mu_0 + \alpha_i + \beta_j + \varepsilon_{ij}; \quad i = 1, \dots, I,\ j = 1, \dots, J$$

with the usual constraints $\sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \beta_j = 0$. Test the (main) effects of both factors for significance.

Hint: The least squares estimators (or MLEs) for the (main) effects can easily be derived by taking partial derivatives of the sum of squares of residuals and by considering the given constraints.

35. **Programming exercise: Keuls (1952) as an `ANOVA2` model.**
We revisit the trial described by M. Keuls (1952, Euphytica 1, 112-122) from exercise 31. However, we now regard an `ANOVA2` model by additionally considering potential "block effects" on the (mean) gross yield per head of cabbage.

Obviously, we encounter the same problem with respect to testing for interactions as in exercise 34 above. However, John W. Tukey (1949) proposed a test for interaction effects which is also applicable in such cases. He assumed a specific multiplicative structure of the interaction effects, namely, $(\alpha\beta)_{ij} = G\alpha_i\beta_j$, where G denotes an (unknown) constant.

By making use of statistics software, carry out Tukey's test procedure on the Keuls (1952) dataset. To this end, the following quantities have to be computed.

(a)

$$SS_G = \frac{\left[\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{Y}_{i.} - \bar{Y}_{..})(\bar{Y}_{.j} - \bar{Y}_{..})Y_{ij}\right]^2}{\sum_{i=1}^{I}(\bar{Y}_{i.} - \bar{Y}_{..})^2 \sum_{j=1}^{J}(\bar{Y}_{.j} - \bar{Y}_{..})^2}$$

(b)

$$RSS = \sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2$$

(c) Under $H_{(\alpha\beta)} : (\alpha\beta)_{ij} = G\alpha_i\beta_j = 0\ \forall (i,j)$, it holds

$$F_{(\alpha\beta)} = \frac{SS_G}{(RSS - SS_G)/(IJ - I - J)} \sim F_{1,(IJ-I-J)}.$$

(d) Interpret your result.

36. **Multiple Select.** Which of the following statements are true and which are false?
Please give reasons for your respective decisions (one short sentence each is sufficient).

1. All three coding methods discussed in Remark 3.43 lead to different estimated responses $(\hat{y}_{ij})_{\substack{1\le i\le k,\\ 1\le j\le n_i}}$
2. If the test for interaction effects in an `ANOVA2` model rejects the null hypothesis, then the estimated joint effects of both factors are always larger than the sums of the individual effect estimates.
3. The sum of the respective degrees of freedom of all sums of squares resulting from decomposition of spread in an `ANOVA2` model equals the total sample size.
4. The grand average in an `ANOVA2` model cannot be represented as a weighted average of all group means corresponding to the factor level combinations if interactions are present in the model.