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 8

Solutions are due on Monday, December 9th, 2013, 3:15pm.
Every completely and correctly solved exercise gives 4 points.

Exercises

29. (a) Prove Theorem 1.37. Hint: Use Theorem 3.29.

    (b) Under the assumptions of Theorem 3.33, derive the restricted least squares estimator

    $$\hat{\beta}_{H_0} = \hat{\beta} - (X^t X)^{-1} K^t [K (X^t X)^{-1} K^t]^{-1} (K\hat{\beta} - d)$$

    in a direct manner. You can use the following result from quadratic optimization theory without proof.

    Result 29.1 *Let $A \in \mathbb{R}^{r \times p}$ and $b \in \mathbb{R}^r$ fixed and define $\mathcal{M} = \{z \in \mathbb{R}^p : Az = b\}$. Moreover, let $f : \mathbb{R}^p \to \mathbb{R}$, given by $f(z) = z^t Q z/2 - c^t z$ for a symmetric, positive semi-definite $(p \times p)$-matrix Q and a vector $c \in \mathbb{R}^p$. Then, the unique minimum of f over the search space $\mathcal{M}$ is characterized by solving the system of linear equations*

    $$Qz - A^t y = c \qquad (1)$$
    $$Az = b \qquad (2)$$

    for (y, z). The component z of this solution minimizes f over $\mathcal{M}$.

30. **Prognostic intervals in multiple linear regression models.**

    (a) Prove Theorem 3.36.3).
    Hint: Notice that $\mu_0 = \mathbb{E}[Y_0 | \vec{X}_0 = \vec{x}_0] = \vec{x}_0 \beta$ and derive the distribution of $\hat{\mu}_0 = \vec{x}_0 \hat{\beta}$.

    (b) Prove Theorem 3.36.4).
    Hint: Derive the distribution of the prognostic error $\hat{\varepsilon}_0 = Y_0 - \vec{x}_0 \hat{\beta}$.

31. **Programming exercise: Keuls (1952).**
    Make yourself familiar with the publication by M. Keuls (1952, Euphytica 1, 112-122, file: `keuls1952.pdf`) which you can download from the lecturer's homepage. If you should encounter problems in downloading the file or if you do not have access to the internet, you may alternatively get the file via USB stick during the lecturer's consulting hour.

    (a) Model the experiment with an `ANOVA1` model. Ignore the potential influence of the field block on the gross yield per head of cabbage and focus only on the effect of the cabbage varieties.

    (b) Reproduce the left half of Figure 1 (page 113 in the publication). What does "Basic number" mean here? Create a dataset in statistics software which comprises the `ANOVA1` design matrix and the response vector corresponding to the data in the publication.

(c) Perform a decomposition of spread as in Theorem 3.39. Test the global hypothesis that varieties have no influence at all on the gross yield per head of cabbage.
Hint: Use Theorem 3.40 (without having to prove it).

32. **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. For $n \to \infty$, Fisher's F-distribution with r and $n - p$ degrees of freedom converges weakly to the distribution of Q/r, where Q is chi-squared distributed with r degrees of freedom.
    2. One-sided linear hypotheses cannot be tested with an F-test, unless $r = 1$.
    3. Introducing an intercept in an `ANOVA1` model leads to a singular design matrix.
    4. `ANOVA1` models are special cases of multiple linear regression models.