ulm university universität ulm

Dr. Tim Brereton
Lisa Handl

Winter Term 2014/15

Stochastics III

Problem Sheet 6

Deadline: January 28, 2015 at noon, before the practical

Exercise 1 (6 points)

Consider the ANOVA test statistic $\sup_{\boldsymbol{a}\in\mathcal{A}} T_{\boldsymbol{a}}^2$, where

$$T_{\boldsymbol{a}} = \frac{\left(\sum_{i=1}^k a_i \overline{Y}_i\right)}{\sqrt{S_p^2 \sum_{i=1}^k a_i^2/n_i}}$$

and $\mathcal{A} = \{\boldsymbol{a} = (a_1, \ldots, a_k)^\top : \boldsymbol{a} \neq \boldsymbol{o},\ \sum_{i=1}^k a_i = 0\}$. Show that

$$\sup_{\boldsymbol{a}\in\mathcal{A}} T_{\boldsymbol{a}}^2 = \frac{\sum_{i=1}^k n_i \left(\overline{Y}_i - \overline{Y}\right)^2}{S_p^2}$$

using the Cauchy-Schwarz inequality.

Hint: Proceed in a similar way as in the proof of Lemma 2.3.

Exercise 2 (4 points)

A bakery in Ulm measures the sales volume of its stores in three different districts of the city. It has 3 stores in district D_1, 5 stores in district D_2 and 2 stores in district D_2, the sales volumes are given in the following table:

district D_1	district D_2	district D_3
703	770	819
788	816	784
715	797	
	774	
	867	

Write out the design matrix, the parameter vector and the vector of response variables for the corresponding model of one-factor analysis of variance in two different ways:

a) such that the design matrix has full rank.

b) using the reparametrization into general mean and effects.

Hint: You don't need to estimate the parameter vector, just write out what the parameters are and outline their relationship to the θ_i.

Exercise 3 (3 points)

In the model of one-factor analysis of variance we considered the following reparametrization of $\theta_1, \ldots, \theta_k$:

$$\theta_i = \mu + \alpha_i \quad \forall i = 1, \ldots, k \quad \text{with} \quad \sum_{i=1}^{k} n_i \alpha_i = 0.$$

Show that this representation is unique.

Exercise 4 (2 + 3 + 5 points)

Consider the example on page 56 of the lecture notes.

a) Show that the matrix $(X^\top X)^-$ defined in (36) is a generalized inverse of the matrix $X^\top X$ as given in (35).

b) Following the example, calculate the least squares estimator for the parameters of model b) in Exercise 2.

c) Consider the model of balanced analysis of variance with $r = 3$ and two factors with 2 levels each. Write out the design matrix (and the parameter vector) and calculate $X^\top X$.

Hint: You may use **R** as a calculator.

Exercise 5 (Revision) (1 + 2 + 2 + 2 points)

Suppose the number of cars per 100 inhabitants of a country depends on the per capita income (in € 10.000) and the gasoline price (in €). The file *cars.txt* on the course website contains the following data:

country	cars	income	gasoline
Belgium	30	29.4	136
Denmark	28	33.0	129
Germany	35	31.2	113
Finland	24	21.3	113
France	33	26.4	140
Greece	8	10.2	129
Ireland	20	11.4	92
Iceland	34	29.4	131
Canada	42	26.1	39
Austria	27	23.1	113

Moreover, we know that in Norway the per capita income is $x_{0,2} = 27.9$ and the gasoline price is $x_{0,3} = 138$. You may assume that the vector of error terms is $\mathrm{N}(0, \sigma^2 I)$-distributed. Use a confidence level of $\alpha = 0.05$ for b) - d).

a) Write out a linear model for this relationship and estimate its parameters.

b) Determine a confidence ellipsoid for the parameter vector.

c) Determine a confidence interval for the expected response variable $\varphi(1, x_{0,2}, x_{0,3}) = \beta_1 + x_{0,2}\beta_2 + x_{0,3}\beta_3$.

d) Determine a prediction interval for the corresponding response variable Y_0.