Second mandatory assignment in STK3100/STK4100-f13

Posted: Monday October 21st.

Deadline: Thursday November 7th, 14:30 pm.

The answers shall be delivered in the box un the hallway on the 7th floor in Niels Henrik Abels house. Use the standard front page which can be downloaded from www.mn.uio.no/math/studier/admin/obligatorisk-innlevering/obligforside.pdf and indicate whether you attend STK3100 or STK4100.

This is the second out of two mandatory assignments in STK31000/STK4100-f13. It consists of two problems. Both handwritten reports and answers using a word processing system are acceptable. Where you use R (or some other statistical package) the relevant part of the output must be enclosed or pasted into the report. It is OK if you cooperate, but each student must deliver a separate and individually formulated report. Also in case of co-operation it should be indicated in the report whom the others are. There is more information in "Regelverk for obligatoriske oppgaver" which can be obtained from the course web page.

Problem 1

Consider a one-way layout of the form

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij},\ i = 1, \dots, I,\ j = 1, \dots, J \tag{1}$$

where $E(\varepsilon_{ij}) = 0$ and $Var(\varepsilon_{ij}) = \sigma^2$ and $\varepsilon_{ij},\ i = 1, \dots, I,\ j = 1, \dots, J$ are uncorrelated and the constraint $\sum_{i=1}^{I} \alpha_i = 0$ is satisfied. For analyzing such model it is known, see e.g. Devore and Berk, that the following quantities are important

$$SSTr = J \sum_{i=1}^{I} (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2 \text{ and } SSE = \sum_{i=1}^{I} \sum_{j=1}^{J} (Y_{ij} - \bar{Y}_{i\cdot})^2$$

where $\bar{Y}_{i\cdot} = \frac{\sum_{j=1}^{J} Y_{ij}}{J}$ and $\bar{Y}_{\cdot\cdot} = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} Y_{ij}}{IJ}$. They are also called the sum of squares between and within groups respectively.

a) Assuming the fixed effects model (1) what are the expected values of SSTr and SSE? What is a reasonable test statistic for the null hypothesis $\alpha_1 = \cdots = \alpha_I = 0$? What is the distribution of this statistic under the null hypothesis if we in addition assume that $\varepsilon_{ij} \sim N(0, \sigma^2),\ i = 1, \dots, I,\ j = 1, \dots, J$?

The random effects or variance component model corresponding to (1) is

$$Y_{ij} = \mu + A_i + \varepsilon_{ij},\ i = 1, \dots, I,\ j = 1, \dots, J \tag{2}$$

where where $E(A_i) = E(\varepsilon_{ij}) = 0$, $Var(A_i) = \sigma_A^2$, $Var(\varepsilon_{ij}) = \sigma^2$ and $A_i,\ \varepsilon_{ij},\ i = 1, \ldots, I,\ j = 1, \ldots, J$ are uncorrelated.

b) Find the expectations of SSTr and SSE under model (2).

c) Use the result in part b) to show that

$$\hat{\sigma}^2 = \frac{SSE}{I(J-1)}$$
$$\hat{\sigma}_A^2 = \frac{SSTr}{J(I-1)} - \frac{\hat{\sigma}^2}{J}$$

are unbiased estimators of σ^2 and σ_A^2 respectively.

From now on we assume that also $A_i \sim N(0, \sigma_A^2)$ and $\varepsilon_{ij} \sim N(0, \sigma^2),\ i = 1, \ldots, I,\ j = 1, \ldots, J$.

d) Formulate model (2) as a linear mixed effects model as it is defined in Zuur et al.

e) Explain why for model (2) $SSE/\sigma^2 \sim \chi^2_{I(J-1)}$.

f) Explain why for model (2) $SSTr/(J\sigma_A^2 + \sigma^2) \sim \chi^2_{(I-1)}$.

g) Why are $SSTr$ and SSE independent if model (2) is assumed?

h) Use this to suggest a test for $H_0 : \sigma_A^2 = 0$ versus the alternative $H_A : \sigma_A^2 > 0$ and find an expression for the power function.

i) Plot the power function for the case where $I = 3,\ J = 5,\ \sigma^2 = 1$ and the level is 5%.

Problem 2

The data set `pigs.txt` on the course web page contains weekly observations of the weights of 16 pigs for a period of 9 weeks.

In this problem we will analyze the development of the weights of the 16 pigs.

a) Compute the mean of the observed measurements for each week and plot the means as a function of time. Plot also, in the same figure, the observed response profiles for each of the 16 pigs. Describe the main features of the plot.

Consider a model of the form

$$Y_i = \mathbf{1}_9\beta_1 + \mathbf{t}_9\beta_2 + \mathbf{e}_i, i = 1, \ldots, 16$$

where the error terms $\mathbf{e}_i, i = 1, \ldots, 16$ are assumed to be independent, multivariate normally distributed with expectation $\mathbf{0}$ and a 9×9 covariance matrix Σ. The vectors $\mathbf{1}_9$ and $\mathbf{t}_9$ are defined as $\mathbf{1}_9 = (1, 1, 1, 1, 1, 1, 1, 1, 1)'$ and $\mathbf{t}_9 = (1, 2, 3, 4, 5, 6, 7, 8, 9)'$.

b) Fit a model of the form described above to the data and find the estimate of the covariance matrix Σ. Consider both the case where the covariance matrix Σ is unrestricted and the case where the nine variances, i.e. the elements on the main diagonal of Σ, are assumed to be equal.

In the rest of the problem we consider a linear mixed effects model of the form

$$Y_i = \mathbf{1}_9\beta_1 + \mathbf{t}_9\beta_2 + Z_i\mathbf{b}_i + \epsilon_i, i = 1, \ldots, 16$$

where $b_i, i = 1, \ldots, 16$ and $\epsilon_i, i = 1, \ldots, 16$ are assumed to be independent, multivariate normally distributed random vectors with expectation 0 and an unstructured $q \times q$ covariance matrix G and $\sigma^2 I_9$ respectively. The matrices Z_i are $9 \times q$ matrices.

c) Fit a random intercept and slope model, i.e. let $q = 2$ and $Z_i = (\mathbf{1}_9, \mathbf{t}_9), i = 1, \ldots, 16$ and fit the model to the measurements of the 16 pigs. What are the estimates of the fixed effects and the estimates of the variance and covariance of the random effects?

d) Now, consider a random intercept model, i.e. let $q = 1$ and $Z_i = \mathbf{1}_9, i = 1, \ldots, 16$ and fit the model to the measurements of the 16 pigs.

e) Discuss how the models in part d) and c) can be compared using a likelihood ratio test.

f) Compute the estimated (marginal) covariance matrices of Y_i using the estimates og the variances and covariances in the models from parts c) and d) and compare them to the estimates of Σ you found in part b). What is your impression?