Linear Models - exam 12th of November 2008

1. A general linear model has a matrix presentation $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. The assumptions made on the model are, however, not clear based on that presentation. State <u>all</u> assumptions made on the model and explain shortly the interpretation of the model. Explain also what is ment by the concepts of the (least squares) residual and the fitted value. Calculate $\mathsf{Cov}(\hat{\boldsymbol{\varepsilon}})$ and $\mathsf{Cov}(\hat{\boldsymbol{\varepsilon}}, \hat{\mathbf{Y}})$ when $\hat{\boldsymbol{\varepsilon}}$ is the (random) residual vector computed from the sample and $\hat{\mathbf{Y}}$ is the (random) vector of fitted values.

2. It is assumed that in problem 1 the matrix $\mathbf{X}$ is orthogonal. Estimate the parameter vector $\boldsymbol{\beta}$ by applying the estimation theory of linear models and verify the probability distribution of the estimator. Derive also the $100(1-\alpha)\%$ confidence bound for the parameter β_j (= the jth component of $\boldsymbol{\beta}$) and explain how the confidence bound will change if the kth column is removed from the matrix $\mathbf{X}$ ($k \neq j$).

3. Assume that n independent observations $Y_1, ..., Y_n$ ($n > 2$) are sampled from a normal distribution (Gaussian distribution) with unknown variance σ^2. The first $n-1$ observations are known to originate from the same normal distribution, but it is suspected that the last observation originates from a normal distribution with a different mean parameter than the previous $n-1$ observations. Formulate the situation using a linear model and derive a test for the hypothesis that the nth observation originates from the same normal distribution as the first $n-1$ observations.

4. Analysis of variance (one-way): The arrangements of the analysis, the statistical model used, the hypothesis most commonly tested, and how it can be tested.

Recall that

- Randomvector $\mathbf{X} \sim \mathsf{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ has density function

$$f(x) = (2\pi)^{-k/2} \det(\boldsymbol{\Sigma})^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})\right\},$$

  where $\det(\boldsymbol{\Sigma})$ is the determinant of the covariance matrix $\boldsymbol{\Sigma}$ and the one-dimentional case is obtained setting $k = 1$.

- $F_{k,m}$-distribution is defined by a random variable $m\chi_k^2/k\chi_m^2$, where $\chi_k^2 \perp\!\!\!\perp \chi_m^2$. Further, $\mathsf{E}(\chi_k^2) = k$ and $\mathsf{Var}(\chi_k^2) = 2k$.

- t_k-distribution is defined by a random variable $Z/\sqrt{\frac{1}{k}\chi_k^2}$, where $Z \sim \mathsf{N}(0,1)$ and $Z \perp\!\!\!\perp \chi_k^2$.

- If $\mathbf{X} \sim \mathsf{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $(\mathbf{X}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X}-\boldsymbol{\mu}) \sim \chi_k^2$.

- If $\mathbf{X} \sim \mathsf{N}_k(\boldsymbol{\mu}, \sigma^2\mathbf{I}_k)$ and matrix $\mathbf{P}$ ($k \times k$) is orthogonal projection with rank r, then $(\mathbf{X}-\boldsymbol{\mu})' \mathbf{P} (\mathbf{X}-\boldsymbol{\mu})/\sigma^2 \sim \chi_r^2$.