Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

Contact person:
Håvard Rue 73593533/92600021

Exam in TMA4267 Linear Statistical Models

Tuesday June 2, 2009
Time: 09:00–13:00

Permitted assisting material: **None**

You may answer in English or Norwegian.
Du kan besvare enten på engelsk eller norsk.

Notation: $\boldsymbol{y} \sim \mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$, denotes that $\boldsymbol{y} = (y_1, y_2, \ldots, y_p)^T$ is normal distributed with mean $\boldsymbol{\mu}$ and covariance matrix Σ. $\boldsymbol{I}$ is the diagonal matrix.

Problem 1 Let

$$y \sim \mathcal{N}_2\left(\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}9 & 2\\2 & 9\end{pmatrix}\right)$$

a) What is the distribution of $z_1 = y_1 + y_2$?

What is the distribution of $z_2 = z_1 - 2y_2$?

What is $\text{Cov}(z_1, z_2)$?

Problem 2 Assume the data $\boldsymbol{y}$ follows the model

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}_n(\boldsymbol{0}, \sigma^2\boldsymbol{I})$, but you estimate the following model using ordinary least squares,

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1^* + \boldsymbol{\epsilon}^*.$$

Here, $\boldsymbol{X}_1 \neq \boldsymbol{X}_2$, and both $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ have full rank.

a) Show that $\widehat{\boldsymbol{\beta}}_1^*$ has the following properties

$$\text{E}\left(\widehat{\boldsymbol{\beta}}_1^*\right) = \boldsymbol{\beta}_1 + \left(\boldsymbol{X}_1^T\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1^T\boldsymbol{X}_2\boldsymbol{\beta}_2$$

and

$$\text{Cov}\left(\widehat{\boldsymbol{\beta}}_1^*\right) = \sigma^2\left(\boldsymbol{X}_1^T\boldsymbol{X}_1\right)^{-1}.$$

b) Under what condition (of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$) is $\widehat{\boldsymbol{\beta}}_1^*$ unbiased?

What is the geometrical interpretation of this condition?

Problem 3 The Cook's distance is defined in the book as

$$D_i = \frac{\left(\widehat{\boldsymbol{y}}_{(i)} - \widehat{\boldsymbol{y}}\right)^T\left(\widehat{\boldsymbol{y}}_{(i)} - \widehat{\boldsymbol{y}}\right)}{(k+1)s^2}$$

a) Explain each term in the Cook's distance.

Explain the use of the Cook's distance.

Problem 4 Let

$$y_1, y_2, \ldots, y_n \overset{\text{iid}}{\sim} \mathcal{N}_1(\mu, \sigma^2).$$

Define

$$\boldsymbol{j} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{J} = \boldsymbol{j}\boldsymbol{j}^T = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ 1 & 1 & \ldots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \ldots & 1 \end{pmatrix}$$

and $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

a) Show that

$$\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} (y_i - \overline{y})^2 + n\overline{y}^2$$

can be expressed as

$$\boldsymbol{y}^T \boldsymbol{I} \boldsymbol{y} = \boldsymbol{y}^T \left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J} \right) \boldsymbol{y} + \boldsymbol{y}^T \left(\frac{1}{n}\boldsymbol{J} \right) \boldsymbol{y}.$$

b) Show that

1. $\boldsymbol{I} - \frac{1}{n}\boldsymbol{J}$ and $\frac{1}{n}\boldsymbol{J}$ are idempotent.
2. $\left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J}\right)\left(\frac{1}{n}\boldsymbol{J}\right) = \boldsymbol{0}$

Problem 4 continuous on the next page...

Here is a transcript of two main results from the book:

Theorem 5.5. Let $\boldsymbol{y}$ be distributed as $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, let $\boldsymbol{A}$ be a symmetric matrix of constants of rank r, and let $\lambda = \frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{A}\boldsymbol{\mu}$. Then $\boldsymbol{y}^T\boldsymbol{A}\boldsymbol{y}$ is $\chi^2(r, \lambda)$, if and only if $\boldsymbol{A}\boldsymbol{\Sigma}$ is idempotent.

Theorem 5.6a. Suppose $\boldsymbol{B}$ is a $k \times p$ matrix of constants, $\boldsymbol{A}$ is a $p \times p$ symmetric matrix of constants, and $\boldsymbol{y}$ is distributed as $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then $\boldsymbol{B}\boldsymbol{y}$ and $\boldsymbol{y}^T\boldsymbol{A}\boldsymbol{y}$ are independent if and only if $\boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{A} = \boldsymbol{0}$.

c) Use Theorem 5.5 and Theorem 5.6a to show that for $\mu = 0$, then

1.
$$\frac{\sum_{i=1}^n (y_i - \overline{y})^2}{\sigma^2} \sim \chi^2_{n-1}$$

2.
$$n\frac{\overline{y}^2}{\sigma^2} \sim \chi^2_1$$

3.
$$\frac{\sum_{i=1}^n (y_i - \overline{y})^2}{\sigma^2} \quad \text{and} \quad n\frac{\overline{y}^2}{\sigma^2} \quad \text{are independent.}$$

d) Use the results in **c)** to construct a F-test (Fisher-test) for

$$H_0 : \mu = 0, \qquad H_1 : \mu \neq 0$$

with significance level α.

e) Show that the test in **d)** is equivalent to a (classical) student-t test for testing the same hypothesis.