

620-371 Linear Models

Semester 1 Exam — June 9, 2009

DEPARTMENT OF MATHEMATICS AND STATISTICS
THE UNIVERSITY OF MELBOURNE

Exam duration: 3 hours
Reading time: 15 minutes
This exam has 6 pages – including this page.

Authorised materials:

Calculators are permitted, but you may be requested to erase the memory of programmable calculators.

Any printed or handwritten material is permitted, including textbooks.

Computers are NOT permitted.

Instructions to invigilators:

The exam paper may be taken out of the examination room.

Instructions to students:

There are 6 questions. All questions should be attempted.

The approximate number of marks for each question is indicated.

The total number of marks available is 100.

This paper may be reproduced and lodged with the Baillieu Library.

1. [12 marks] Consider the matrices

$$A = \frac{1}{5} \begin{bmatrix} 3 & 4 \\ -4 & 3 \end{bmatrix}, B = \frac{1}{10} \begin{bmatrix} 1 & -3 \\ -3 & 9 \end{bmatrix}.$$

- (a) Show that A is orthogonal.
 (b) Find the trace of B .
 (c) Show that B is idempotent.
 (d) Using your answer from part 1b, find the rank of B .
 (e) Suppose that X and Y are two symmetric matrices with a common diagonalizing matrix, i.e. there exists an orthogonal matrix P so that both $P^T X P$ and $P^T Y P$ are diagonal. Show from first principles that $XY = YX$. You may use the property that $D_1 D_2 = D_2 D_1$ if D_1 and D_2 are diagonal matrices.
 (f) Find $\frac{\partial}{\partial \mathbf{y}}(\mathbf{y}^T A \mathbf{y})$, where $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$.

2. [14 marks] Let \mathbf{y} be a normal random vector with mean $\boldsymbol{\mu} = \begin{bmatrix} 0 \\ -4 \\ 1 \end{bmatrix}$ and variance I . Let

$$A = \frac{1}{3} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}.$$

- (a) Find $E[\mathbf{y}^T A \mathbf{y}]$.
 (b) Find the distribution of $\mathbf{y}^T A \mathbf{y}$.
 (c) Let \mathbf{z} be a normal random vector that is independent of \mathbf{y} , with mean $\begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix}$ and variance I . Find the distribution of $\mathbf{y}^T \mathbf{y} + \mathbf{z}^T \mathbf{z}$.
 (d) Under what condition on B is $\mathbf{y}^T A \mathbf{y}$ independent from $B \mathbf{y}$?

3. [24 marks] We wish to fit a linear model to explain the amount of diesel fuel a truck uses, given the number of hours that it runs. A study is performed and the following data collected:

Amount of fuel (litres)	Time motor runs (hours)
5	0.6
10	2
15	2.1
20	2
22	2.4

We fit the linear model $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where \mathbf{y} is the vector of response values, X is the design matrix, $\boldsymbol{\varepsilon}$ is the vector of errors, and

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

is the parameter vector, with β_0 an intercept term and β_1 the parameter associated with the running time.

- What are the common assumptions used to fit a linear model?
- Calculate the least squares estimators of the parameters.
- Calculate the fitted value and residual for the first sample.
- Show that in a general linear model, $SS_{Res} = SS_{Total} - \mathbf{y}^T X \mathbf{b}$, where SS_{Res} is the residual sum of squares, SS_{Total} is the total sum of squares, and \mathbf{b} is the least squares estimator of the parameter vector.
- What is the smallest possible variance of a linear estimator of β_0 ? You may express this in terms of σ^2 , the variance of the errors.
- According to the model, what is the average length of time a motor would take to consume 12 litres of fuel?
- Find a 95% confidence interval for β_1 . You may take $SS_{Res} = 62.86$ and the critical value for a t distribution with 3 degrees of freedom as $t_{0.025} = 3.182$.
- Explain the difference between a confidence interval and a prediction interval for a given set of predictor variables.
- Simultaneously taking 95% confidence intervals for all the parameters gives us a joint confidence region for the parameter vector. What is the confidence level of this region?

4. [14 marks] Consider the study in question 3.
- (a) Test for model adequacy, $H_0 : \boldsymbol{\beta} = \mathbf{0}$, at the 5% level. You may take the critical value of an F distribution with 2 and 3 degrees of freedom as $f_{0.05} = 9.55$.
 - (b) Find $R(\beta_0|\beta_1)$, the regression sum of squares of β_0 in the presence of β_1 .
 - (c) What is a difference between a partial test of $H_0 : \beta_0 = 0$ and a sequential test of $H_0 : \beta_0 = 0$, assuming that the sequential test tests the parameters in the order β_0, β_1 ?
 - (d) Express the hypothesis $5\beta_0 = \beta_1 = 1$ as a general linear hypothesis in matrix form.
 - (e) Give an advantage of mutually orthogonal design.

5. [24 marks] A study of two major coal seams in a region is conducted, to compare the sulfur content of coal drawn from each seam. The following data is collected (in terms of percentage of sulfur):

Seam	
1	2
1.51	1.69
1.92	0.64
1.08	0.9

The linear model that we use is

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, i = 1, 2, j = 1, 2, 3,$$

where τ_i is the effect on the sulfur content associated with seam i , and ε_{ij} is the error for the j th sample from seam i .

- (a) Write down the linear model in a matrix form.
- (b) Find two different conditional inverses for $X^T X$, where X is the design matrix.
- (c) Under what conditions does a matrix have 0, 1, or an infinite number of conditional inverses?
- (d) What is the best linear unbiased estimator for μ ?
- (e) Is $2\mu + \tau_1 + \tau_2$ estimable? Why or why not?
- (f) Estimate $\tau_1 - \tau_2$.
- (g) Calculate s^2 , the estimator for the variance of the errors.
- (h) Find a 95% confidence interval for $\tau_1 - \tau_2$. You may take the critical value for a t distribution with 4 degrees of freedom as $t_{0.025} = 2.78$.

6. [12 marks] We are interested in modelling the life span of paint in years, based on two different types of undercoating and three different types of paint formula. A study is conducted with one sample from each pair of factor levels and the following data is obtained:

		Formula (factor 1)		
		1	2	3
Undercoat	1	1.09	1.35	1.6
(factor 2)	2	1.16	1.38	2.18

- Express this data as a two-factor linear model without interaction, in matrix form.
- We wish to test the hypothesis that the formula used has no effect on the lifespan of the paint. How many degrees of freedom do we use in the F distribution that we are testing against?
- Express the hypothesis that neither formula nor undercoat has an effect on paint life in matrix form.
- Is this hypothesis testable? Justify your answer.
- Write down the design matrix if we were to express this as a two-factor linear model with interaction.
- Suppose the true means are:

		Formula (factor 1)		
		1	2	3
Undercoat	1	1	1.3	1.5
(factor 2)	2	1.1	1.4	2

Is there interaction between the factors? Why or why not?