



Semester 1 Assessment, 2023

School of Mathematics and Statistics

## MAST30025 Linear Statistical Models Assignment 3

Submission deadline: **Friday May 26, 5pm**

This assignment consists of 3 pages (including this page) with 5 questions and 46 total marks

### Instructions to Students

#### *Writing*

- This assignment is worth 7% of your total mark.
- You may choose to either typeset your assignment in  $\text{\LaTeX}$ , or handwrite and scan it to produce an electronic version.
- You may use R for this assignment, including the `lm` function unless otherwise specified. If you do, include your R commands and output.
- Write your answers on A4 paper. Page 1 should only have your student number, the subject code and the subject name. Write on one side of each sheet only. Each question should be on a new page. The question number must be written at the top of each page.

#### *Scanning and Submitting*

- Put the pages in question order and all the same way up. Use a scanning app to scan all pages to PDF. Scan directly from above. Crop pages to A4.
- Submit your scanned assignment as a single PDF file and carefully review the submission in Gradescope. Scan again and resubmit if necessary.

**Question 1 (7 marks)**

Let  $A$  be an  $n \times p$  matrix with  $n \geq p$ .

- (a) Show that  $r(A^c A) = r(A)$ .
- (b) Show that  $I - A(A^T A)^c A^T$  is idempotent.
- (c) Show that  $r(I - A(A^T A)^c A^T) = n - r(A)$ .

**Question 2 (11 marks)**

We study the amount of rotting of a potato exposed to a variety of levels of oxygen, and a variety of temperatures. A small experiment is conducted and the following data obtained:

Temperature	Oxygen level		
	1	2	3
10	13	10	15
	11	4	2
	3	7	7
16	26	15	20
	19	22	24
	24	18	8

For this question, you may not use the `lm` or `ginv` functions in R.

- (a) Plot an interaction plot of the data. Does there appear to be interaction?
- (b) Fit an additive model, outputting your design matrix. Estimate the common variance.
- (c) Calculate a 95% confidence interval for the difference between the temperature effects.
- (d) Test the hypothesis that oxygen level has no effect on rotting at the 5% significance level.
- (e) Suppose we are interested in the effect of oxygen level only, but know that temperature affects the results, so we include it in our model. What type of design would this study be?

**Question 3 (5 marks)**

Consider a linear model with only categorical predictors, written in matrix form as  $\mathbf{y} = X_1 \boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1$ . Now suppose we add some continuous predictors, resulting in an expanded model  $\mathbf{y} = X \boldsymbol{\beta} + \boldsymbol{\varepsilon}$ .

Now consider a quantity  $\mathbf{t}^T \boldsymbol{\beta}$ , where  $\mathbf{t}^T = [\mathbf{t}_1^T | \mathbf{t}_2^T]$  is partitioned according to the categorical and continuous predictors. Show that if  $\mathbf{t}_1^T \boldsymbol{\beta}_1$  is estimable in the first model, then  $\mathbf{t}^T \boldsymbol{\beta}$  is estimable in the second model.

If you write  $X = [X_1 | X_2]$ , you may assume that  $r(X) = r(X_1) + r(X_2)$ .

*Hint: Use Theorems 6.9 and 6.3. For any vectors  $\mathbf{z}_1$  and  $\mathbf{z}_2$ , you can write*

$$\left[ \begin{array}{cc|c} X_1^T X_1 & X_1^T X_2 & X_1^T X_1 \mathbf{z}_1 \\ X_2^T X_1 & X_2^T X_2 & X_2^T X_2 \mathbf{z}_2 \end{array} \right] = \left[ \begin{array}{cc} X_1^T & \mathbf{0} \\ \mathbf{0} & X_2^T \end{array} \right] \left[ \begin{array}{cc|c} X_1 & X_2 & X_1 \mathbf{z}_1 \\ X_1 & X_2 & X_2 \mathbf{z}_2 \end{array} \right].$$

**Question 4 (11 marks)**

An experiment is conducted to investigate the distance travelled by three different toy cars on a smooth surface. Each car starts from a 16-inch long ramp tilted at different angles (in degrees), and the travel distance (in meters) is recorded. The data is given in the file `toycars.csv`, available on the LMS.

- (a) Plot the data, using different colours and/or symbols for each car. What do you observe?
- (b) Test for the presence of interaction between the angle and the type of toy car.
- (c) Use backward elimination to select relevant variables for the data.
- (d) For model from part (c), test the hypothesis that the type 1 toy car travels 0.05 meters more on average than the type 3 toy car, at the 5% significance level.
- (e) In the full model with interaction, test the hypothesis that the distances travelled by the type 2 and type 3 toy cars are the same, at the 5% significance level.

**Question 5 (12 marks)**

Consider the following model with both categorical and continuous predictors:

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

Here, we consider  $\mu$  to be a nuisance parameter.

**For this question, you may not use the `lm` function in R.**

- (a) State the reason why  $\mu$  can be regarded as a nuisance parameter.
- (b) Write down the matrices  $X_1$  and  $X_2$ .
- (c) Derive the reduced design matrix  $X_{2|1}$ .
- (d) Suppose  $(x_{11}, x_{12}, x_{13}, x_{21}, x_{22}, x_{23}) = (2, 4, 8, 7, 6, 4)$  and  $(y_{11}, y_{12}, y_{13}, y_{21}, y_{22}, y_{23}) = (4, 2, 10, 8, 8, 12)$ . Find a solution of the reduced normal equations.
- (e) Find the corresponding solution for  $\mu$  for your solution to part (d). Show directly that your solutions for  $(\mu, \tau_i, \beta)$  solve the full normal equations.

**End of Assignment — Total Available Marks = 46**