MAST30025: Linear Statistical Models

Assignment 3, 2020

Due: 5pm Friday, June 5 (week 11)

- This assignment is worth 7% of your total mark.
- You may use R for this assignment, including the `lm` function unless specified. If you do, include your R commands and output.
- Your assignment must be submitted on Canvas LMS as a single PDF document only (no other formats allowed). You may choose to either typeset your assignment or handwrite and scan it to produce an electronic version. The LMS will not accept late submissions. It is your responsibility to ensure that your assignments are submitted correctly and on time, and problems with online submissions are not a valid excuse for submitting a late or incorrect version of an assignment.
- Your assignment must clearly show your name and student ID number, your tutor's name and the time and day of your tutorial class. Your assignment must be submitted in the correct format and the correct orientation. Your answers must be clearly numbered and in the same order as the assignment questions.

1. Consider the matrix $A(A^TA)^cA^T$.
   (a) Show that this matrix is unique (invariant to the choice of conditional inverse).
   (b) Show that the rank of this matrix is $r(A)$.

Memory

Similar to the Tutorial question, first tutorial !

2. 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:

	Oxygen level		
Temperature	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.
   (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?

Seriously this question!, Complete Block Design, this will be later covered in Experimental Design

Read the question carefully and figure out the order, They tell you which is your tao and beta parameters respectively. Also ensure you compute your design matrix properly

3. Consider the two-factor model with interaction

$$y_{ij} = \mu + \tau_i + \beta_j + \xi_{ij}.$$

Suppose that there are a and b levels of the factors respectively. Now consider the set of equations

$$\xi_{ij} - \xi_{1j} - \xi_{i1} + \xi_{11} = 0, \quad i = 2, \ldots, a,\ j = 2, \ldots, b.$$

(a) Show that the equations are not redundant.

(b) Show that these equations are equivalent to the hypothesis of no interaction.

(c) Thereby calculate the rank of the hypothesis of no interaction.

(d) Show that the hypothesis is testable, provided there exists at least one sample from each combination of factor levels.

Annotate and clarify this later!

4. Maple trees have winged seeds called samara. An experiment is conducted to investigate the effect of shape on the speed of descent. Samara were collected from three trees, and their "disk loading" (a quantity based on size and weight, which was used to quantify shape) and descent velocity are calculated. The data is given in the file `heli.csv`, available on the LMS.

(a) Plot the data, using different colours and/or symbols for each tree. What do you observe?

(b) Test for the presence of interaction between disk loading and tree.

(c) Use backward elimination from the model with interaction to select variables for the data.

(d) Add lines corresponding to model from part (c), and the full model with interaction, to the plot from question (a).

(e) In the full model with interaction, test the hypothesis that a samara from tree 2 with a disk loading of 0.2 has an average descent velocity of 1.

5. An apple orchard has 32 trees set aside for an experiment which aims to examine the effect of mulching on tree growth. There are four mulching treatments: 1. Control (no mulch); 2. Wood chips; 3. Garden compost; 4. Clippings from a local council collection. The trees are in a 4×8 rectangle, labeled as shown in the diagram below. The experimenter has the resources to maintain 16 plots, each consisting of 2 adjacent trees. All trees in the same plot must have the same treatment.

beta1

beta2

Treatment 1 (Control)	1	2	3	4	5	6	7	8
Treatment 2 (Wood Chips)	9	10	11	12	13	14	15	16
Treatment 3 (Garden compost)	17	18	19	20	21	22	23	24
Treatment 4 (Clippings)	25	26	27	28	29	30	31	32

In this question, if you need randomisation, use R and reproduce your R commands and output.

(a) Construct an appropriate experimental design. Draw the allocated treatments and write down the matrices in the corresponding linear model.

(b) Now suppose the ground slopes down from the left to the right of the diagram. Repeat question (a).

by memory

We will cover this a lot in Experimental Design