MAST30025: Linear Statistical Models

Assignment 1, 2019

Due: 5pm Friday March 29 (week 4)

- This assignment is worth 6% of your total mark. Remember to complete the plagiarism declaration form online before you submit this assignment.

- You may use R for this assignment, but for matrix calculations only (you may not use the `lm` function). If you do, include your R commands and output.

- Your assignment must be submitted to Turnitin on the LMS as a single PDF document only. You may choose to either typeset your assignment or handwrite and scan it to produce an electronic version. Turnitin will not accept late submissions.

- Turnitin gives you an option to preview your work prior to submission. Please check this preview carefully to ensure you are submitting the correct document. After a successful submission to Turnitin, you will see a submission ID. This confirmation will also be sent to your University email address. If you do not see a submission ID, you should assume that your assignment has not been submitted successfully. Either try to submit again or contact the tutor co-ordinator (Rheanna Mainzer) immediately to arrange an alternate means of submission. Issues with Turnitin are not a valid excuse for submitting a late assignment or an incorrect version of an assignment.

- **(2 marks)** 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. Suppose that A is a symmetric matrix with $A^k = A^{k+1}$ for some integer $k \geq 1$. Show that A is idempotent.

2. Let $A_1, A_2, \ldots, A_m$ be a set of symmetric $k \times k$ matrices. Suppose that there exists an orthogonal matrix P such that $P^T A_i P$ is diagonal for all i. Show that $A_i A_j = A_j A_i$ for every pair $i, j = 1, 2, \ldots, m$.

3. Show directly that for any random vector $\mathbf{y}$ and compatible matrix A, we have var $A\mathbf{y} = A(\text{var } \mathbf{y})A^T$.

4. Let $\mathbf{y}$ be a 3-dimensional multivariate normal random vector with mean and variance

$$\mu = \begin{bmatrix} 3 \\ 0 \\ -2 \end{bmatrix}, \quad V = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Let

$$A = \frac{1}{10} \begin{bmatrix} 4 & -2 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 10 \end{bmatrix}.$$

(a) Describe the distribution of $A\mathbf{y}$.

(b) Find $E[\mathbf{y}^T A \mathbf{y}]$.

(c) Describe the distribution of $\mathbf{y}^T A \mathbf{y}$.

(d) Find all linear combinations of $\mathbf{y}$ elements which are independent of $\mathbf{y}^T A \mathbf{y}$.

5. The table below shows prices in US cents per pound received by fishermen and vessel owners for various species of fish and shellfish in 1970 and 1980. (Taken from Moore & McCabe, Introduction to the Practice of Statistics, 1989.)

Type of fish	Price (1970)	Price (1980)
Cod	13.1	27.3
Flounder	15.3	42.4
Haddock	25.8	38.7
Menhaden	1.8	4.5
Ocean perch	4.9	23.0
Salmon, chinook	55.4	166.3
Salmon, coho	39.3	109.7
Tuna, albacore	26.7	80.1
Clams, soft-shelled	47.5	150.7
Clams, blue hard-shelled	6.6	20.3
Lobsters, american	94.7	189.7
Oysters, eastern	61.1	131.3
Sea scallops	135.6	404.2
Shrimp	47.6	149.0

We will model the 1980 price of fish, based on the 1970 price.

(a) The linear model is of the form $\mathbf{y} = X\beta + \varepsilon$. Write down the matrices and vectors involved in this equation.

(b) Find the least squares estimates of the parameters.

(c) Calculate the sample variance s^2.

(d) A fisherman sold ocean trout for 18 cents per pound in 1970. Predict the price for ocean trout in 1980.

(e) Calculate the standardised residual for sea scallops.

(f) Calculate the Cook's distance for sea scallops.

(g) Does sea scallops fit the linear model? Justify your argument.