STAT151A Quiz 1

Please write your full name and email address here:

Also, please put your intials on each page in case the pages get separated.

You have 30 minutes for this quiz.

There are three questions, (1), (2), and (3), each weighted equally..

There are extra pages at the end if you need more space for solutions.

Question 1

Let β_1 and β_2 denote two unknowns, and consider the following system of equations.

$$\beta_1 + 3\beta_2 = 4$$
$$\beta_1 + 3\beta_2 = 5$$
$$\beta_1 + 2\beta_2 = 7$$
$$\beta_1 + 1\beta_2 = 1$$
$$\beta_1 + 9\beta_2 = 9$$

(a) Write the system of equations in matrix form $\boldsymbol{X\beta} = \boldsymbol{Y}$, where $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$ is the 2–vector of unknowns, $\boldsymbol{X}$ is a 5×2 matrix, and $\boldsymbol{Y}$ is a 5–vector.

(b) How many solutions does this sytem of equations have? Briefly justify your answer.

Question 2

For this problem, I will use the following definitions.

- $\boldsymbol{X}$ denotes an $N \times P$ matrix
- $\boldsymbol{Y}$ denotes an N–vector (i.e. an $N \times 1$ matrix)
- $\boldsymbol{\beta}$ denotes a P–vector

Take $N > P > 1$. You may assume that $\boldsymbol{X}$ is full column rank.

For each expression, write the dimension of the result, or write "badly formed" if the expression is not a valid matrix expression.

(a) $\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X}$
(b) $\boldsymbol{X}\beta$
(c) $\boldsymbol{Y}^{\mathsf{T}}\boldsymbol{Y}$
(d) $\boldsymbol{Y}\boldsymbol{Y}^{\mathsf{T}}$
(e) $\mathrm{trace}\left(\boldsymbol{Y}\boldsymbol{Y}^{\mathsf{T}}\right)$
(f) $\boldsymbol{Y} - \boldsymbol{X}\beta$
(g) $(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X})^{-1}$
(h) $\boldsymbol{X}^{\mathsf{T}}\left(\boldsymbol{Y} - \boldsymbol{X}\beta\right)$
(i) $\boldsymbol{X}^{\mathsf{T}}\boldsymbol{Y} - \boldsymbol{X}^{\mathsf{T}}\boldsymbol{X}\beta$
(j) $(\boldsymbol{X}^{\mathsf{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{Y} - \beta$

Question 3

Suppose I have a dataset in which each row is a student, and the columns contain the following variables for a given year:

- Their final score in a linear models class
- The self-reported total number of hours spent studying per week
- Their grade in a prerequesite theoretical statistics course (like STAT135)
- Whether or not they are a foreign exchange student

Consider two ways I might consider using this dataset:

(Use 1) How much can a student expect to increase their linear models grade by spending more hours studying?

(Use 2) How can I identify students who may need extra assistance at the beginning of the next semseter of linear models?

(a) Which of these two uses is a prediction problem, and which is an inference problem?
(b) Which question do you expect to be easier to answer with this dataset?

Briefly justify your answers.

Extra space for answers (indicate clearly which problem you are working on)

Extra space for answers (indicate clearly which problem you are working on)