STAT151A Quiz 1 (Sep 17th)

Please write your full name and email address:

You have 30 minutes for this quiz.

There are three parts, (a), (b), and (c), each weighted equally..

(a)

Let β_1 and β_2 denote two unknowns. Write the following system of equations in the matrix form $\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{Y}$, 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.

$$\begin{aligned} \beta_1 + 3\beta_2 &= 4 \\ \beta_1 + 3\beta_2 &= 5 \\ \beta_1 + 2\beta_2 &= 7 \\ \beta_1 + 1\beta_2 &= 1 \\ \beta_1 + 9\beta_2 &= 9 \end{aligned}$$

How many solutions does this sytem of equation have?

(b)

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

I will 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.

(1) $\boldsymbol{X}^\intercal \boldsymbol{X}$
(2) $\boldsymbol{X}\beta$
(3) $\boldsymbol{Y}^\intercal \boldsymbol{Y}$
(4) $\boldsymbol{Y}\boldsymbol{Y}^\intercal$
(5) $\text{trace}\,(\boldsymbol{Y}\boldsymbol{Y}^\intercal)$
(6) $\boldsymbol{Y} - \boldsymbol{X}\beta$
(7) $(\boldsymbol{X}^\intercal \boldsymbol{X})^{-1}$
(8) $\boldsymbol{X}^\intercal (\boldsymbol{Y} - \boldsymbol{X}\beta)$
(9) $\boldsymbol{X}^\intercal \boldsymbol{Y} - \boldsymbol{X}^\intercal \boldsymbol{X}\beta$
(10) $(\boldsymbol{X}^\intercal \boldsymbol{X})^{-1}\boldsymbol{X}^\intercal \boldsymbol{Y} - \beta$

(c)

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?

- Which of these two uses is a prediction problem, and which is an inference problem?

- Which question do you expect to be easier to answer with this dataset?

Briefly justify your answers.