

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 $\mathbf{X}\boldsymbol{\beta} = \mathbf{Y}$, where $\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$ is the 2-vector of unknowns, \mathbf{X} is a 5×2 matrix, and \mathbf{Y} is a 5-vector.

$$\beta_1 + 3\beta_2 = 4$$

$$\beta_1 + 3\beta_2 = 5$$

$$\beta_1 + 2\beta_2 = 7$$

$$\beta_1 + 1\beta_2 = 1$$

$$\beta_1 + 9\beta_2 = 9$$

How many solutions does this system of equations have?

(b)

For this problem, I will use the following definitions.

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

I will take $N > P > 1$. You may assume that \mathbf{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) $\mathbf{X}^\top \mathbf{X}$
- (2) $\mathbf{X} \beta$
- (3) $\mathbf{Y}^\top \mathbf{Y}$
- (4) $\mathbf{Y} \mathbf{Y}^\top$
- (5) $\text{trace}(\mathbf{Y} \mathbf{Y}^\top)$
- (6) $\mathbf{Y} - \mathbf{X} \beta$
- (7) $(\mathbf{X}^\top \mathbf{X})^{-1}$
- (8) $\mathbf{X}^\top (\mathbf{Y} - \mathbf{X} \beta)$
- (9) $\mathbf{X}^\top \mathbf{Y} - \mathbf{X}^\top \mathbf{X} \beta$
- (10) $(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} - \beta$

(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 prerequisite 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 semester 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.