Assignment 0 (two pages)

Statistics 32950-24620 (Spring 2024)

Submit ASAP before 11:59 pm Thursday, March 23.

Instructions

- One of the main purposes of this assignment is for you to get familiar with Gradescope, which will be used for homework submission and grading for this course.
- Submission guide

  Start by typing your name, your course section (either STAT24620 or STAT32950), and Assignment 0.

  Name the file of your answers as `LastnameFirstnamePset0.pdf` (can be abbreviated; example: `PeterPanPset0.pdf`).

  Submit the file to `246Pset0` or `329Pset0` on Gradescope (in the left column menu on Canvas)

  Important: submission to the wrong section will cause problems and will be deleted.
  If you accidentally submitted to a wrong link, please email the instructor asap.

  Important: Mark (i.e., tag) the page numbers corresponding to each problem and its sub-questions.
- This assignment also prompts you to practice on typing math formula, using R, and producing data plot.
- You should start to review the background materials assumed for this course.

  Basic probability and statistics, as in chapters 1-6 and 8-10 in Rice (Mathematical Statistics and Data Analysis);
  Basic linear regression models, as in Rice's chapter 14;
  Basic vector matrix algebra, as in Johnson and Wichern's 2A.

Problem assignments

1. (*Typing math formula*)

   For this question, type out your answers, including the mathematical formula.
   (You may use LaTeX, Word and its Equation Editor, or any writing programs of your choice.)

   Consider a dataset consisting of n pairs of real numbers, (x_i, y_i), $i = 1, \cdots, n$.
   Suppose the data are n independent observations of a pair of random variables (X, Y).

   Type the formula of Pearson sample correlation coefficient of (X, Y) in terms of the x_i's and y_i's.

2. (*Producing R plot*)

   Produce data and 2-dimensional scatter plot with least square line by using the following R commands (shown at R prompt `>`). Replace `???` in the last command by an appropriate plot title.

```
> x= runif(30); y=x^3+rnorm(30)/3
> plot(x,y); abline(lm(y~x))
> title("???")
```

3. (*Matrix operation on rows and columns*)
Define three matrices (and their column vectors) as

$$A = [a_1\ a_2\ a_3] = \begin{bmatrix} 3 & -4 & -1 \\ -6 & 0 & 5 \\ 4 & 5 & 7 \end{bmatrix}, \quad B = [b_1\ b_2\ b_3] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -2 \end{bmatrix}, \quad E = [e_1\ e_2\ e_3] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

(a) What happens to A when A is multiplied by B from the left?
Note: You are asked to describe BA in terms of changes in the row or column vectors of matrix A.
What happens to A when A is multiplied by B from the right? (Now describe AB.)
What happens to A when A is multiplied by E from the left?
What happens to A when A is multiplied by E from the right?

(b) $v = \begin{bmatrix} 7 \\ 3 \\ 24 \end{bmatrix} = [7\ 3\ 24]^T$ is a column vector, often represented by its coordinates as $(7, 3, 24)$.
Write v as a linear combination of the a_i's.
Write v as a linear combination of the b_i's.
Write v as a linear combination of the e_i's.

4. (*Mathematical induction, sigma summation notation, derivation type of proofs*)

(a) Use mathematical induction to prove that $\sum_{k=1}^{n} k^3 = (\sum_{k=1}^{n} k)^2$.
(State and prove the two steps of the induction procedure clearly.)

(b) A and B are 4×4 matrices. Show that $AB = BA$ is not always true.

5. (*Function of random variables*) The random variable U is uniform on $[-1, 1]$. Define $X = U^2$.

(a) Derive the probability density function of X.

(b) Calculate the mean $\mathbb{E}(X)$.