Coding Quiz 3

April 13, 2023

Instructions:

- **BS students**: Drop only your .R file here: https://www.dropbox.com/request/AiDC7YaCebyMdwWvXJje.

  MSc students: Drop only your .R file here: https://www.dropbox.com/request/OZbZC9osbgV7WSBLFSz4.

  NO other files should be submitted through this link.
- Accepted format of the file: 111111.R (where '111111' is your roll number).

  Use your full name (e.g., Subhajit Dutta) and IITK email address (e.g., duttas@iitk.ac.in) while submitting your .R file.

  If you submit multiple files, only the .R file will be considered.

  Moreover, **grading will be based solely on this .R file**.
- Output should be printed strictly in the order of the questions given below.
- Total marks: 20
- Time: 30 minutes (6:30pm to 7pm).

  Needless to say, R codes dropped **after 7pm will NOT be graded**.
- **Set the seed to be** 1 **at the beginning of your R code**.
- Only text written below in Red should be printed when the R code is executed.
- Please avoid spamming by NOT uploading incorrect and/or multiple files.

Read the set of instructions given above again before moving to the next page.

Question 1: Generate a random sample of size 50 from the following density function:

$$f_X(x) = \frac{1}{\sqrt{72\pi}}(e^{-\frac{x^2}{32}}(1 + e^{-\frac{3x^2}{32}})),$$

where $x \in \mathbb{R}$. Further, generate Y as follows:

$$Y = 7.5 + 0.68X + \epsilon,$$

where $\epsilon \sim N(0, 1)$.

a. Run regression for the above generated values of Y and X. Now, plot the fitted and the population regression curves in different colours.
Your code should only display the plot.

b. Let β_0 and β_1 be the true intercept and slope terms in the above regression equation. Plot $\frac{\beta_0}{\hat{\beta}_0}$ and $\frac{\beta_1}{\hat{\beta}_1}$ for increasing sample sizes $n = 100, 200, 500, 1000, 5000$. Here, $\hat{\beta}_0$ and $\hat{\beta}_1$ are the corresponding estimates. Are they consistent?
In addition to displaying the plot, your code also should print "Consistent", or "Inconsistent" as the case may be.

Question 2: Read data directly from the following link:

https://www.dropbox.com/s/geke5ykega8lytr/Q3_data.csv?dl=1.

a. Fit a logistic regression model for the above data taking Y as the response variable. Which of the following predictors are statistically significant at level $\alpha = 0.05$?
Your code should display the statistically significant predictor variables only.

b. Compute the proportion of correct predictions made by the model for the data (assign 1 as the predicted value if the predicted probability > 0.5, and 0 otherwise).
Your code should only print the proportion of correct predicted values.

Output marks: $(3 + 3) + (2 + 2) = 10$

R code marks: $6 + 4 = 10$