

Linear Statistical Models

Week-2: Graded Assignment

Subjective Assignment: (Manual-grading)

Max. Marks: 30

NOTE: All the plots should be properly labelled.

1.
 - (a) Generate 1000 discrete random numbers within the range $[1,5]$, and store it in a vector named 'x'. [1 Mark]
 - (b) Generate 1000 random numbers from the normal distribution with mean μ and variance σ^2 , and store it in a new vector 'y'. [1 Mark]
Note: You can choose any value of μ and σ .
 - (c) Using 'ggplot' in R, plot the scatter plot between the variables 'y' and 'x'. [1 Mark]
 - (d) Using 'ggplot' in R, plot the scatter plot between x and y . [1 Mark]
 - (e) Based on the visualization of the above plotted scatter plot, categorize the pairs (x, y) into 5 appropriate categories, in the above plotted scatter plot (by representing them in different colors). Comment on the visualization of categories in scatter plot. [3 Marks]
2. Explore the site: Click to access.
Pick any section of your choice which you find interesting and elaborate on it. Also, try it out on any publicly available datasets. [8 Marks]
3. Use the in-built data set `cars` in R-software of the library `datasets` to answer the following questions.
 - (a) Give a brief description of the dataset and read the data set as data frame in R. [1 Mark]
 - (b) Using 'ggplot' in R, plot the scatter plot between the variables 'dist' and 'speed'. On visualizing the scatter-plot, comment on the relationship between 'dist' and 'speed'. [2 Marks]
 - (c) Create and plot 200 random linear models for the 'cars' data set of cars ('dist' on 'speed'). Also, find the mean squared error for each of the fitted models.[3 Marks]
 - (d) From the above fitted 200 models, select the best one and write down its equation. Also, mention the reason why you think the selected model is best among the fitted 200 models. [3 Marks]
 - (e) Use the inbuilt function `lm()` in R, to find the equation of best fitted linear model ('dist' on 'speed') for the 'cars' dataset. [1 Mark]

- (f) Using ‘[ggplot](#)’ in [R](#), plot the linear models obtained in part (*d*) and (*e*) in the same scatter plot (plotted in part (*b*)). [2 Marks]
- (g) For each of the fitted models (part (d) and (e)), plot the ‘residuals’ vs ‘speed’ using ‘[ggplot](#)’ in [R](#). Comment on the obtained plots. [3 Marks]