

Python – Final Exam – Sample 1

Date: 09/04/2022

Marks: 60

Total Question: 5

All questions are mandatory

Q1.) (10 Marks)

Find Mode for,

a.) $X \sim N(\mu = 0, \sigma = 2.5)$; $X = -5, -4.9, \dots, 5$

b.) $X \sim \text{Exp}(\lambda = 0.55)$; $X = 0, 0.1, 0.2, \dots, 10$

c.) Use a while loop to generate 2 random sample of size 50 from a bag of 0-100 until the Correlation Coefficient between them is < 0 . Also, print the nth draw at which such a criteria is satisfied.

Q2.) (5 Marks)

A system takes in the marks of the students of out of 100.

However, due to a bug the system adds its own garbage values of any random size starting with 999. For eg, 999, 233, 23231, 3221 ...

Fix the above issue. You can take the marks from user / initialise them.

Q3.) (5 Marks)

Calculate mean for the below data while ignoring all the negative values.

$X = (12, 2, -81, 83, 94, -1, -332, -21, 33, 100, 56, 51, -9)$

(You are expected to write a program from scratch and NOT use functions like sum(), mean(), etc...)

Q4.) (5 Marks)

Import the data "Orange.csv".

Get the summary statistics for each of type of tree and Comment on your observations.

Q5.) (35 Marks)

You are given a Data set "Spam Email.csv".

Predict the target variable "Spam".

Every row represents one email and every column represents the word used in the email and every cell presents the frequency of that word used in that email.

Build the best possible model based on the metric: Accuracy, F-Beta Score.

Perform all required Data Cleaning / Feature Selection / Diagnostics steps.
