STAT 650 Assignment-04

Due: October 3, 2022 5:59PM

Instructions :

- This assignment is based on materials coverved in Lectures 09, 10 and 11.
- We highly recommend that you write your solutions in **Jupyter Notebook** and convert them to a **PDF** file. However, you may write the solutions by hand, scan and upload it as **.pdf** file.
- The PDF file should be under 15MB in size. It must be uploaded as a single file and not separate files for separate pages. Do not take a photo of each page and then paste them into a document - this will make your file too big and the results will generally not be very readable anyway.
- Please make sure that the solutions are neat, legible and in order (even if you choose to solve them in different order).
- Include **STAT650--UIN** at the top of the first page.
- Name the file as **UIN_assign4.pdf** (For eg, if someone's UIN is 123456789, then the file should be named 123456789_assign4.pdf). Otherwise, your submission will not be graded.
- You should upload your file through Canvas. You can make multiple submissions within the deadline, but only the latest submission will be considered for grading.
- You may take 6 hours extra after the due time, but 10% of your marks will be deducted.
- It is strictly prohibited to share or distribute the content in this document.

The aim of this assignment is to get familiar with basic concepts in statistics.

Question 1

Suppose X is a random variable having the following probability mass function:

$$P(X = x) = \begin{cases} 0.2 & \text{if } x = 0 \\ 0.2 & \text{if } x = 1 \\ 0.2 & \text{if } x = 3 \\ 0.4 & \text{if } x = 5 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the cumulative density function for X.
(b) Find $E(X)$ and $Var(X)$.
Show the calculations clearly.
[$5 + 10 = 15$ marks]

Question 2

A large company has 100 potential new prescription drugs under clinical test. About 20% of all drugs that reach this stage are equivalently licensed for sale.

(a) What is the probability that at least 15 of the 100 drugs are equivalently licensed ?
(b) Calculate probability in part (a) using binormial-normal approximation
(Show the steps clearly and then you may use Python to calculate probabilities in part (a) and (b) in the final step.)
$[5 + 10 = 15$ marks$]$

Question 3

Suppose Y has a Poisson distribution with mean $\lambda = 9$.
(a) Find $P(7 \leq Y \leq 11)$.
(b) Calculate the probability in part (a) using Normal approximation to Poisson
(Show the steps clearly and then you may use Python to calculate probabilities in part (a) and (b) in the final step.)
$[5 + 10 = 15$ marks$]$

Question 4

Suppose X is a Normal distribution with mean $\mu = 3$ and variance $\sigma^2 = 9$.
(a) Find $P(1 < X < 6)$ in terms of the cdf of standard Normal distribution.
(b) Find $E((2 + X)^2)$ and $Var(4 + 3X)$. Show the steps clearly (No Python calculations for this part)
$[5 + 5 = 10$ marks$]$

Question 5

Suppose X has an exponential density with parameter $\lambda = 2$ (i.e. $f(x) = \lambda e^{-\lambda x}, x \geq 0$). Define $Y = 2X + 4$.

(a) Find the cumulative distribution of Y. Also find the density of Y.
(b) What is the mean and variance of Y?
(c) What is the 95th percentile of Y? [Use of Python is allowed and encouraged.]
(d) Use Python to find $P[2 \leq Y \leq 36]$.
$[5 \times 4 = 20$ marks$]$

Question 6

A book club classifies members as *heavy, medium*, or *light* purchasers, and seperate mailings are prepapred for each of these groups. Overall, 20% of the members are heavy perchasers, 30% medium, and 50% light. A member is not classified into a group untill 18 months after joining the club, but a test is made of the feasibility of using the first 3 months' purchases to classify members. The following percentages are obtained from existing records of individuals classified as heavy, medium, or light purchasers

First 3 months' Purchases	Heavy	Medium	Light
0	5	15	60
1	10	30	20
2	30	40	15
3+	55	15	5

If a member purchases no books in the first three months, what is the probability that the member is a light purchaser ?

[10 marks]

Question 7

(Use of Python is allowed and encouraged for the following questions)

(a) Suppose $X_1, X_2, \cdots, X_{100}$ is a random sample of size 100 from $N(\mu, \sigma = 3)$ distribution, μ is unknown. If the sample mean is $\bar{X} = 45$, find the 95% confidence interval of μ.

(b) Suppose the mean of the first 50 observations $(X_1, \cdots, X_{50})$ is 42, then on the basis of the last 50 observations $(X_{51}, \cdots, X_{100})$ find the 95% confidence interval of μ.

(c) Which confidence interval of μ do you think is better? Justify your answer.

[$5 + 5 + 5 = 15$ marks]