KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DEPARTMENT OF MATHEMATICS

MATH 405: Learning from Data
Term 231, Major Exam I
Saturday October 14, 2023, 07:30 PM

Name: _____ ID #: _____

Question No	Full Marks	Marks Obtained
1	**08**	
2	**03**	
3	**05**	
4	**07**	
5	**05**	
6	**05**	
Total	**33**	

Instructions:

1. Mobiles are not allowed in the exam. If you have your **mobile** with you, **turn it off** and put it **on the table/floor** so that it is visible to the proctor.

2. Show all the calculation steps. There are points for the steps so if you miss them, you lose points. For multiple choice type questions, showing calculation steps is not required.

3. Report **at least 4 decimal points** of your numerical answers.

Q1: (8 pts.) Suppose that a discrete random variable X follow the following distribution:

$$f(x; n, p) = {}^nC_x p^x (1-p)^{n-x} \text{ for } x = 0,1,2,\ldots,n.$$

Here, $n \geq 0$ and $0 < p \leq 1$ are the parameters of distribution. Mathematically, derive an expression for the variance of X.

Q1 continues…

Q2: (3 pts.) Briefly explain the major differences between the Pearson's and Spearman's correlation coefficients.

Q3: (5 pts.) Let X be a random variable following a continuous probability distribution given by

$f(x; \alpha, \beta) = \alpha\beta \exp[\alpha + \beta x - \alpha e^{\beta x}]$ and $F(x; \alpha, \beta) = 1 - \exp[-\alpha(e^{\beta x} - 1)]$ for $x \geq 0$.

Here, $\alpha > 0$ is the shape parameter of distribution and $\beta > 0$ is the scale parameter.

Write down a detailed procedure for generating random numbers from this given distribution with specific values of α and β.

Note: In the second part of the exam, you will be asked to use RStudio and draw these random numbers.

Name: _____ ID #: _____

Q4: (7 pts.) Let X be a random variable following a continuous probability distribution given by

$$f(x; \alpha, \beta) = \alpha\beta \exp[\alpha + \beta x - \alpha e^{\beta x}] \text{ and } F(x; \alpha, \beta) = 1 - \exp[-\alpha(e^{\beta x} - 1)] \text{ for } x \geq 0.$$

Setting $\alpha = \beta = 1$, generate 10^6 random numbers from this distribution and compute the following:

The mean is _____

The median is _____

The mode is _____

Also write your code for the function of Mode you developed.

Comment on the shape of the distribution:

Q5: (5 pts.) Download the dataset from Blackboard and write Code # _____

Data on the jet turbine engine are available on four variables with $n = 32$. If the variables x1 and x4 do not follow a bivariate normal distribution, test the hypothesis that there is negative linear association between x1 and x4.

H0: _____

H1: _____

p-value of the test: _____

Decision and conclusion:

Q6: (2+2+1 = 5 pts.) Generate 10^6 random numbers from Student's T distribution with 10 degrees of freedom using the command `rt(n = 10^6, df = 10)`.

    (i)   Compute the proportion of the values that are within mean ± 2 SD. Write all the details of your solution.

    (ii)   Between what two symmetric values, 99% of the data lie? Write all the details of your solution.

    (iii)   Compute the interquartile range. Write just the final answer.

Good luck