

PUSASQF202
Probability and Statistics-2

Time: 2 hours
Total Marks: 60 marks

Note:

- 1) The candidate has option to either attempt question 4A or question 4B. Rest all questions are mandatory.
- 2) Numbers to the right indicate full marks.
- 3) The candidates will be provided with the formula sheet and graph papers (if required) for the examination.
- 4) Use of approved scientific calculator is allowed.

Q1. Attempt All Questions

A)

5Marks

(i) If the daily number of motor claims registered with Suraksha General Insurance Company follows a Poisson distribution with mean = 28, find the probability that 200 or more claims will be registered in the next week.

(ii) At its smaller peer, Raksha General Insurance Company, motor claims follow a Poisson Distribution with mean of 0.5 claims per day. Calculate the Probability that more than 2 claims occur in any given week.

B)

5Marks

A game of dice offers you two options as reward:

- Option A: You roll a fair die once and you get a sum of money (in rupees) computed as the rolled number multiplied by 1000
- Option B: You roll a fair die 1000 times and get a sum of money (in rupees) equal to the sum of all the numbers rolled on the die

What is the expected reward and its variance under both options? Which one would you choose and why?

[Note: Please ignore the physical effort involved in rolling the die 1000 times and answer only on the basis of the statistical properties of the payoff.]

C)

5Marks

For a quantity following the standard normal distribution, 10 underlying samples are randomly chosen.

- (i) What is the probability that the sample mean is greater than 1?
- (ii) What is the probability that the sample Standard Deviation is greater than 1?
- (iii) What is the probability that both sample mean and sample variance are greater than 1?
- (iv) How would your answer to part (iii) change if the sample size was doubled?

Q2. Attempt All Questions

A)

5Marks

For a normally distributed sample of size 84 with mean = 20 and given the population variance = 21. Find the following confidence intervals of Population mean :

- (a) 90% Left – sided Confidence Interval
- (b) 99% Right – sided Confidence Interval
- (c) 95% Shortest length Confidence Interval

B)

5Marks

A university wishes to examine if the performance of its male and female students is significantly different. The marks of the 16 male students have a sample mean of 56 and a standard deviation of 6. The marks of the 25 female students have a sample mean of 62 and a sample standard deviation of 5.

- (i) State assumptions required to perform a two – sample t-test on the above data.
- (ii) Perform an appropriate test to determine whether the difference in performance is statistically significant.

C)

5Marks

In the Actuarial Soccer League (ASL), the number of goals in the 50 matches had the following distribution:

No. of goals	0	1	2	3	4	5
No. of matches with those many goals	4	15	17	7	5	2

- (i) Calculate the sample mean for the above data
- (ii) Conduct a suitable statistical test to ascertain whether the number of goals follow a Poisson distribution.
- (iii) An official who was part of the last year's ASL mentions that the Number of Goals followed a $Poi(2)$ Distribution for the last ASL. How would your answer to part(ii) change to test if the data given above follows a $Poi(2)$ Distribution?

Q3. Attempt All Questions

A)

5Marks

Consider the following statements and answer the subsequent questions:
Statement A: Pearson's correlation=100%; Statement B: Spearman's rank correlation=100%

- i. If A is true, does it imply that B must be true?
- ii. If A is false, does it imply that B must be false?
- iii. If B is true, does it imply that A must be true?
- iv. If B is false, does it imply that A must be false?
- v. How would the above answers change if statement B referred to Kendall's rank correlation (tau) instead of Spearman's rank correlation?

B)

5Marks

The logit function is defined as: $f(x) = \log\left(\frac{x}{1-x}\right)$. For which distribution is the logit function a natural choice as a link function (in the context of GLM) and why? What are some other possible functions which can act as a link function for that distribution?

C)

5Marks

For a linear regression of y over x with 5 data pairs, it was found that $s_{xx} = 10$, $s_{yy} = 44.126$, $s_{xy} = 20.828$. Construct an ANOVA table of the sum of squares of y and make comments interpreting the same as an F-test.

Q4. Attempt Q4 A or Q4 B

A)

15Marks

You are provided with a Tabletop Random Number Generator (TaRaNG) that returns a real number uniformly from 0 to n (n is an unknown positive real number).

- i. Given k sample observations $\{x_1, x_2, \dots, x_k\}$ from this TaRaNG, find two estimators for n , one using the method of moments (MoM) and one using maximum likelihood estimation (MLE).
- ii. Compute the bias for MoM estimator and comment on the bias for MLE estimator.
- iii. Comment on the applicability of the result on the asymptotic distribution of MLEs and Cramer-Rao Lower Bound.
- iv. Describe the issue with each of the two approaches and suggest an alternative estimator for n .
- v. What would your estimators be under both the approaches (MoM and MLE) if the range of TaRaNG was $-n$ to n (instead of 0 to n)?

B)

15Marks

For the paired data given below,

x	0	1	2	3	4
y	0.97	1.09	1.29	1.47	1.59

- i. Derive the estimates for the slope and the intercept parameters of a linear regression model on the data above from first principles (i.e. deriving coefficients by minimizing the sum of square of errors without the formulae).
- ii. Compute the total variation of Y and partition it into that which is explained by the model and that which is not. Hence determine the coefficient of determination (R^2).
- iii. Compute the residuals and comment on the model based on them.
- iv. It is later discovered that (10,3.39) and (20,11.50) also belong to the same dataset. How does your model perform in predicting them and why?