

MAHATMA GANDHI UNIVERSITY, KOTTAYAM
MGU-UGP (HONOURS)
THIRD SEMESTER EXAMINATION
(2024 ADMISSION ONWARDS)
MG3DSEMATA07–PROBABILITY AND STATISTICS FOR DATA SCIENCE
AND MACHINE LEARNING

Duration: 2 hrs

Maximum Marks: 70

Students should attempt at least one question from each course outcome to enhance their overall outcome attainability.

Part A

Answer any five questions. Each question carries 2 marks

1. Two coins are flipped, yielding the sample space $S = \{(H, H), (H, T), (T, H), (T, T)\}$. Let E be the event that at least one head appears, and F be the event that at least one tail appears. Find the intersection event EF (or $E \cap F$). [U][1]
2. A fair six-sided die is rolled, where each outcome is equally likely. Calculate $P(\{2, 4, 6\})$, the probability of rolling an even number. [U][1]
3. State Bayes's formula. [K][1]
4. Suppose X is a continuous random variable with probability density function:

$$f(x) = \begin{cases} C(4x - 2x^2), & 0 < x < 2 \\ 0, & \text{otherwise} \end{cases}$$

Find the value of C . [U][2]

5. If X is a continuous random variable with cumulative distribution function $F(x)$, express $P(a \leq X \leq b)$ in terms of $F(a)$ and $F(b)$. [U][2]
6. If X is a continuous random variable with density function $f(x) = \frac{1}{100}e^{-x/100}$ for $x \geq 0$, calculate $P(X < 100)$. [A][2]
7. If a sample $x_1, x_2, \dots, x_n$ has a sample mean of $\bar{x} = 5$, find the sample mean of the modified observations $y_i = 3x_i + 4$ for $i = 1, 2, \dots, n$. [U][3]
8. Calculate the sample variance of the observations: 2, 4, 6, 8. [A][5]

Part B

Answer any **five** questions. Each question carries **6** marks.

9. Three machines A , B , and C produce 50%, 30%, and 20% of the total output in a factory, respectively. The defect rates for the three machines are 3%, 4%, and 5%, respectively. Given that a randomly chosen item from the output is found to be defective, find the probability that it was produced by machine A . [U][1]

10. The probability mass function of a discrete random variable X is given by

$$p(i) = c \frac{\lambda^i}{i!}, \quad \text{for } i = 0, 1, 2, \dots$$

where $\lambda > 0$. Find:

- (a) c [U][1]
- (b) $P(X = 0)$ [U][1]
- (c) $P(X > 2)$ [U][1]

11. Consider continuous random variables X and Y with joint probability density function:

$$f(x, y) = \begin{cases} 2, & 0 < x < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

Find the covariance σ_{XY} . [A][3]

12. (a) Define the sample mean and sample median for a set of n observations $x_1, x_2, \dots, x_n$. [K][3]
- (b) Explain why the median is considered a robust measure of central tendency compared to the mean when outliers are present in a dataset. [K][3]

13. Calculate the variances σ_X^2 , σ_Y^2 and correlation coefficient ρ_{XY} for

$$f(x, y) = \begin{cases} 8xy, & 0 < y < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

[U][5]

14. A sample of 36 measurements taken from a population with known standard deviation $\sigma = 0.3$ yields a sample mean of $\bar{x} = 2.6$.

- (a) Construct a 95% confidence interval for the population mean μ . [A][4]
- (b) Construct a 99% confidence interval for μ and briefly explain why it is wider than the 95% interval. [A][4]

15. Suppose a researcher wishes to estimate a population mean μ from a population with a known standard deviation $\sigma = 0.3$.

- (a) State the formula for the sample size n required to be $100(1 - \alpha)\%$ confident that the error in estimating μ will not exceed e . [K][4]
- (b) Determine the minimum sample size required to be 95% confident that the sample mean $\bar{x}$ differs from μ by less than 0.05. [K][4]

16. (a) Define the Simple Linear Regression model $Y = \beta_0 + \beta_1 x + \epsilon$ and state the standard assumptions on the random error term ϵ . [K][5]
- (b) Using the Method of Least Squares, derive the normal equations used to estimate the regression parameters β_0 and β_1 . [U][5]

Part C

Answer any three questions. Each question carries 10 marks.

17. (a) If A and B are two independent events defined on a sample space S , prove that their complements A' and B' are also independent. [U][1]
- (b) Distinguish clearly between pairwise independence and mutual independence for three events A , B , and C . Provide an example or scenario illustrating that pairwise independence does not guarantee mutual independence. [U][1]
18. (a) A person speaks the truth 4 out of 5 times. A die is tossed, and the person reports that the outcome is a "6". Find the actual probability that the die rolled a "6". [A][2]
- (b) Now suppose two independent witnesses, who each speak the truth with probability $\frac{4}{5}$, both report that the die rolled a "6". Formulate the events and calculate the updated conditional probability that the die actually landed on a "6". [A][2]
19. A continuous random variable X has probability density function:

$$f(x) = \begin{cases} 3x^2, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

- (a) Verify that $f(x)$ is a valid probability density function. [U][2]
- (b) Find the expected value $E(X)$. [U][2]
- (c) Calculate $E(X^2)$ and use it to find the variance $\text{Var}(X)$. [U][2]
20. The probability that a component produced by a machine is defective is 0.05. A sample of 10 components is selected at random.
- (a) Find the probability that exactly 2 components are defective. [U][2]
- (b) Find the probability that at most 1 component is defective. [U][2]
21. For a given data distribution, the mean is 50, the mode is 44, and the standard deviation is 12.
- (a) Calculate Karl Pearson's coefficient of skewness. [U][3]
- (b) Interpret the result to explain the nature and direction of asymmetry in the distribution. [An][3]
22. A random sample of 100 items has a sample mean of 53. The population is known to have a mean $\mu = 50$ and standard deviation $\sigma = 10$.
- (a) Test whether the sample mean differs significantly from the population mean at a 5% significance level ($Z_{\text{critical}} = 1.96$). [A][5]
- (b) Calculate the p -value for this test and state your conclusion. [A][5]