

1. Define and differentiate between qualitative and quantitative data with an example of each. Explain why a boxplot is more suitable than a histogram for comparing the spread of two different datasets. (5 Marks)
2. A dataset of 10 values has a sample mean of 22 and a sample standard deviation of 5. It is discovered that one recorded value, 30 was incorrect and should have been 20. Determine the new mean and the new standard deviation. (5 Marks)
3. A dataset has a mean of 50, a median of 45, and a standard deviation of 10. Interpret the skewness of the distribution. If the calculated kurtosis value is less than 3, what does this imply about the shape of the distribution compared to a normal curve. (5 Marks)
4. The coefficient of variation (CV) for Dataset A is 15% and for Dataset B is 25%. Both have the same mean. What does the CV tell you about the relative variability? Explain a practical scenario in business where understanding both the standard deviation and the CV is crucial for decision-making. (5 Marks)
5. State the axioms of probability. From a standard deck of 52 cards, what is the probability of drawing either a heart or a queen? (5 Marks)
6. The number of system failures in a data center per week is modeled by a Poisson random variable X with mean $\lambda = 2.5$. The center also tracks Y , the number of weeks in a 4-week period with at least one failure. (5 Marks)
 - (a) Find $P(X \geq 1)$ for a single week
 - (b) Identify the probability distribution that Y follows and state its parameters.
 - (c) Calculate the probability that there is at least one failure in exactly 3 out of the 4 weeks.
7. Using Bayes' theorem, solve : In a factory, Machine A produces 60% of items and Machine B produces 40%. The defect rate for A is 2% and for B is 4%. If a randomly selected item is found to be defective, what is the probability it was produced by Machine B? (5 Marks)

8. The lifetime of a component follows an exponential distribution with a mean of 1000 hours. What is the probability that the component lasts more than 1200 hours? Justify why the exponential distribution is often used to model time-to-failure events? (5 Marks)
9. Differentiate between a population and a sample. Explain one key advantage of stratified random sampling over simple random sampling (SRSWOR). (5 Marks)
10. Consider two estimators for the population variance (σ^2): $\hat{\sigma}_1^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$, $\hat{\sigma}_2^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$. (5 Marks)
11. Describe the concept of the sampling distribution of the sample mean. State the Central Limit Theorem (CLT) and explain its crucial importance in statistical inference, even when the population distribution is not normal. (5 Marks)
12. A survey estimates the proportion of satisfied customers as 0.70 from a sample of 100. Construct an appropriate 95% confidence interval for the true population proportion. Interpret this interval. What would happen to the width of this interval if the sample size were increased to 400? (5 Marks)
13. (a) Define null and alternate hypotheses.
- (b) A survey claims that 9 out of 10 doctors recommend aspirin for their patients with headaches. To test this claim, a random sample of 100 doctors is obtained. Of these 100 doctors, 82 indicate that they recommend aspirin. Is this claim accurate? Use $\alpha = 0.05$.
 $\left(Z_{\frac{\alpha}{2}} = 1.96 \right)$ (5 Marks)
14. (a) Write the formula for 't' statistic for testing one mean.
- (b) An automobile tyre manufacturer claims that the average life of a particular grade of tyre is 20,000 km. A random sample of 16 tyres is having mean 22,000 km with a standard deviation of 5000 km. Validate the claim of the manufacturer at 5%. (for 15 d.f, $t_{\alpha} = 1.75$) (5 Marks)

15. A survey of 100 customers gives the following data (5 Marks)

Gender	Type of brand	
	A	B
Men	30	20
Women	20	30

Test whether preference of brand is independent of gender. ($\alpha = 0.05$, given table value = 3.84)

16. (a) What are the assumptions of ANOVA.
 (b) Sketch the ANOVA table for a one-way classified data. (5 Marks)
17. (a) Define scatter diagram. What is the purpose of scatter plot?
 (b) What do you mean by Correlation and Regression analysis? (5 Marks)
18. The following table shows the sales and Advertisement expenditure of a firm.

	Sales (x)	Advertisement expenditure (y) (in Crores)
Mean	40	6
S.D.	10	1.5

Coefficient of correlation $r = 0.9$. Write the regression equation on x and y . Estimate the likely sales for a proposed Advertisement expenditure of Rs.10 crores. (5 Marks)

19. (a) What is R-square (coefficient of determination)
 (b) Assume that the coefficient of correlation between rainfall and per acre yield of rice is 0.8. Find out the coefficient of determination and comment on its value. (5 Marks)
20. (a) Differentiate between simple and multiple regression.
 (b) Define Multicollinearity (5 Marks)

21. (a) Define sign test.
(b) Define Spearman's rank correlation coefficient and when is it used. (5 Marks)
22. Define Wilcoxon signed rank test and explain its importance. (5 Marks)
23. Explain the Mann-Whitney U test and state its assumptions. (5 Marks)
24. Write short note on Kruskal-Wallis test. (5 Marks)
25. Explain the three basic principles of Design of Experiments. Illustrate each with an example from forestry or ecological studies. (5 Marks)
26. Describe the structure, layout, advantages and limitations of a Completely Randomized Design. Mention one suitable environmental application. (5 Marks)
27. Distinguish between Randomized Block Design and Latin square Design with respect to experimental layout, sources of variation controlled and application situations. (5 Marks)
28. Explain the role of Analysis of Variance (ANOVA) in experimental design. Describe how ANOVA helps in testing treatment effects in ecological or agricultural experiments. (5 Marks)
29. Explain the concept of Multivariate Data Analysis. Why is it important in environmental and ecological data studies? (5 Marks)
30. Explain the basic idea of Principal Component Analysis (PCA). How does PCA help in reducing dimensionality of environmental data? (5 Marks)
31. Describe Cluster Analysis. Distinguish between Hierarchical clustering and K-means clustering with suitable environmental applications. (5 Marks)
32. Explain Shannon Diversity Index and Simpson's Diversity Index. State their importance in biodiversity and wildlife studies. (5 Marks)

33. Explain the advantages of Holt-Winters forecasting method over simple exponential smoothing. (5 Marks)
34. State the additive and multiplicative models of time series and explain which is suitable in the following situations :
- (a) A forestry technology company's user growth may exhibit a seasonal pattern where late-summer adoption rates increase by 15% relative to the underlying trend. Assume that the trend grows exponentially.
 - (b) If a wholesale business experiences a consistent quarterly sales increase of Rs.10,00,000 during the summer each year irrespective of overall sales growth. (5 Marks)
35. Define the Autocorrelation Function (ACF) in the context of time series analysis. Describe the underlying characteristics and behaviour of a time series when its ACF plot exhibits the following properties : (5 Marks)
- (a) consistently positive values,
 - (b) periodic peaks,
 - (c) rapid decay towards zero, and
 - (d) near-zero values for all lags $k > 0$.
36. What is an ARIMA model? Explain the significance of the three core components of an ARIMA (p, d, q) model. (5 Marks)
37. Provide the primary use of the following R functions : (5 Marks)
- (a) rep ()
 - (b) seq ()
 - (c) data.frame ()
 - (d) rank ()
 - (e) factor ()

38. Explain various approaches for importing data available in Excel and text formats in R. (5 Marks)
39. How can you enter a missing observation in R? How do you handle missing observations in a data set in R during statistical data analysis? (5 Marks)
40. Explain Spatially Explicit Capture-Recapture (SECR) method for estimating animal population. Suggest a package in R as a primary tool for SECR. (5 Marks)
-