Part A: Descriptive Statistics & Data Concepts (18 Questions)

1. In a research study, data is collected from 5000 students across India. If the goal is to make conclusions about "all students in India," then:
   ☐a) 5000 students represent the population
   ☒ b) 5000 students represent a sample
   ☐ c) All students in India represent a sample
   ☐ d) No sampling is done

2. Which statement correctly differentiates population and sample?
   ☐ a) Population is always larger than sample
   ☐ b) Sample contains all possible outcomes
   ☒ c) Sample is a subset of population used for analysis
   ☐ d) Population is chosen from a sample

3. If a dataset has mean = 40, median = 30, and mode = 20, then the distribution is:
   ☐ a) Symmetric
   ☒ b) Positively skewed
   ☐ c) Negatively skewed
   ☐ d) Normal

4. A dataset contains exam marks of students. If one student's mark is wrongly entered as 900 instead of 90, this value is called:
   ☐ a) Noise
   ☒ b) Outlier
   ☐ c) Skewness
   ☐ d) Kurtosis

5. In data analysis, "noise" refers to:
   ☐ a) Extreme values
   ☒ b) Random error or irrelevant variations in data
   ☐ c) Correlated values
   ☐ d) Hidden patterns

6. Which of the following is most affected by outliers?
   ☒ a) Mean
   ☐ b) Median
   ☐ c) Mode
   ☐ d) Interquartile Range

7. The measure that indicates "peakedness" or "flatness" of a distribution is:
   ☐ a) Skewness
   ☒ b) Kurtosis
   ☐ c) Variance
   ☐ d) Standard Deviation

8. If a distribution has high kurtosis, it means:
   ☒ a) Heavy tails and more outliers
   ☐ b) Flat distribution with fewer outliers

☐ c) Symmetric bell-shape
☐ d) Zero variance

9. Which is an example of continuous random variable?
☐ a) Number of cars in a parking lot
☒ b) Temperature in a city
☐ c) Number of emails received
☐ d) Defective items in a batch

10. A scalar quantity can be represented as:
☒ a) A single number
☐ b) A column of numbers
☐ c) A 2D array
☐ d) A multidimensional cube

11. A vector is different from a scalar because:
☐ a) Vector has only magnitude
☒ b) Vector has magnitude and direction
☐ c) Vector is always positive
☐ d) Vector has no physical meaning

12. Tensor can be defined as:
☐ a) A single number
☐ b) A 1D vector only
☒ c) A generalization of scalars, vectors, and matrices to higher dimensions
☐ d) Only a 2×2 matrix

13. Which visualization is best to check skewness and outliers in data?
☒ a) Pie chart
☐ b) Histogram
☒ c) Box plot
☐ d) Scatter plot

14. Standardization (z-score scaling) is preferred when:
☐ a) Features have same units
☐ b) Data is categorical
☒ c) Features have very different scales and we want mean = 0, variance = 1
☐ d) Data has missing values

15. Normalization (min-max scaling) transforms data to:
☒ a) [0, 1] or [−1, 1] range
☐ b) Mean = 0, SD = 1
☐ c) Logarithmic scale
☐ d) Polynomial form

16. A dataset follows power-law distribution if:
☒ a) Large values are equally frequent as small values
☐ b) Few large values occur rarely while many small values occur frequently

☐ c) Distribution is symmetric around mean
☐ d) Variance = 0

17. Correlation between two variables measures:
☐ a) Difference between them
☒ b) Strength and direction of linear relationship
☐ c) Causation
☐ d) Variance of both variables

18. Which correlation value indicates the strongest linear relationship?
☒ a) −0.85
☐ b) +0.70
☐ c) 0.00
☐ d) +0.45

Part B: Probability & Distributions (12 Questions)

19. A fair coin is tossed 3 times. The probability of getting exactly 2 heads is:
☐ a) 1/8
☒ b) 3/8
☐ c) 1/2
☐ d) 5/8

20. The probability of an impossible event is:
☐ a) 0
☒ b) 1
☐ c) −1
☐ d) Undefined

21. In Poisson distribution, mean (λ) = variance. If λ = 4, then standard deviation = ?
☒ a) 4
☐ b) 2
☐ c) 8
☐ d) 16

22. Poisson distribution is suitable for:
☐ a) Continuous measurements
☒ b) Rare discrete events over fixed time/space
☐ c) Correlated variables
☐ d) Normal data only

23. In exponential distribution with mean = 5, the rate parameter (λ) is:
☐ a) 5
☐ b) 1/5
☒ c) 10
☐ d) 0.5

24. Which of the following is NOT a property of probability distribution?
    - ☐ a) All probabilities ≥ 0
    - ☐ b) Total probability = 1
    - ☒ c) Probabilities can be > 1
    - ☐ d) Each outcome has defined probability

25. If two events A and B are independent, then $P(A \cap B) = ?$
    - ☐ a) P(A) + P(B)
    - ☒ b) P(A) × P(B)
    - ☐ c) P(A)/P(B)
    - ☐ d) None

26. Central Limit Theorem is important because:
    - ☐ a) Population is always normal
    - ☒ b) Sample mean distribution tends to normal for large n
    - ☐ c) Standard deviation always decreases with sample size
    - ☐ d) Variance becomes zero

27. If a distribution is symmetric and bell-shaped, it is:
    - ☒ a) Normal distribution
    - ☐ b) Poisson distribution
    - ☐ c) Exponential distribution
    - ☐ d) Power-law distribution

28. In probability, a random variable is:
    - ☐ a) A fixed number
    - ☒ b) A function assigning numbers to outcomes of an experiment
    - ☐ c) Always continuous
    - ☐ d) Always discrete

29. If two dice are rolled, the sample space has:
    - ☐ a) 6
    - ☐ b) 12
    - ☐ c) 18
    - ☒ d) 36 outcomes

30. Which probability distribution is used for modeling "time between arrivals"?
    - ☐ a) Poisson
    - ☐ b) Normal
    - ☒ c) Exponential
    - ☐ d) Uniform

Part C: Inferential Statistics & Hypothesis Testing (20 Questions)

31. The null hypothesis (H_o) generally states that:
    - ☐ a) A difference exists
    - ☒ b) No difference exists

☐ c) Data is always skewed
☐ d) Sample size is large

32. Type-I error occurs when:
☒ a) Rejecting a true null hypothesis
☐ b) Accepting a true null hypothesis
☐ c) Rejecting a false null hypothesis
☐ d) None

33. Type-II error occurs when:
☐ a) Rejecting a true null hypothesis
☒ b) Accepting a false null hypothesis
☐ c) Rejecting a false null hypothesis
☐ d) None

34. The probability of Type-I error is denoted by:
☒ a) β
☒ b) α
☐ c) μ
☐ d) σ

35. The power of a statistical test is defined as:
☒ a) $1 - \alpha$
☐ b) $1 - \beta$
☐ c) $\alpha + \beta$
☐ d) β/α

36. A p-value less than significance level ($\alpha = 0.05$) means:
☐ a) Fail to reject H_o
☒ b) Reject H_o
☐ c) Increase sample size
☐ d) Accept alternative only if $\alpha < 0.01$

37. A 95% confidence interval means:
☐ a) 95% of population lies in interval
☒ b) 95% probability that parameter lies in interval
☒ c) 95% of such intervals constructed from samples will contain true parameter
☐ d) Both b and c

38. Larger sample size leads to:
☐ a) Larger standard error
☒ b) Smaller standard error
☐ c) No effect
☐ d) More bias

39. Z-test is generally used when:
☐ a) Sample size is small and σ unknown
☒ b) Sample size is large and σ known

☐ c) Comparing categorical variables
☐ d) Variance is unequal

40. T-test is used when:
☐ a) Population variance is known
☒ b) Sample size is large
☒ c) Population variance is unknown and sample is small
☐ d) Data is categorical

41. The chi-square test is most appropriate for:
☐ a) Comparing means of two groups
☒ b) Testing independence between categorical variables
☐ c) Testing slope of regression line
☐ d) Analyzing correlation

42. The F-test is generally used to compare:
☐ a) Two sample means
☒ b) More than two means (ANOVA) or variances
☐ c) Two proportions
☐ d) Skewness

43. Correlation ≠ Causation because:
☐ a) High correlation always means randomness
☒ b) A third factor may influence both variables
☐ c) Correlation is always zero
☐ d) It measures only causality

44. If correlation coefficient r = 0, it means:
☐ a) No relationship at all
☐ b) No linear relationship
☒ c) Variables are independent
☐ d) Variables are strongly dependent

45. The sampling distribution of the mean refers to:
☐ a) Distribution of population
☒ b) Distribution of all possible sample means
☐ c) Normal distribution always
☐ d) Distribution of sample variance

46. Which test would you use to compare the average salary of male and female employees?
☐ a) Chi-square test
☒ b) T-test for independent samples
☐ c) Z-test
☐ d) F-test

47. If standard deviation of population is unknown, which distribution is used for hypothesis test?
☐ a) Z-distribution
☒ b) T-distribution

☐ c) F-distribution
☐ d) Chi-square distribution

48. The critical region in hypothesis testing refers to:
☒ a) Values where null hypothesis is rejected
☐ b) Values where null hypothesis is accepted
☐ c) Always $\alpha = 0.05$
☐ d) Confidence interval

49. Which one is TRUE about confidence level and significance level?
☐ a) Confidence level + α = 100%
☒ b) Confidence level − $\alpha = 1$
☐ c) Both are equal
☐ d) Both are independent

50. When the p-value is 0.85, at $\alpha = 0.05$, the correct decision is:
☐ a) Reject H_o
☒ b) Fail to reject H_o
☐ c) Accept H_o without doubt
☐ d) Increase sample size