



*Yogoda Satsanga
Mahavidyalaya*

Self-assessment Question Bank

NAME OF THE DEPARTMENT: COMMERCE

NAME OF THE FACULTY: Mrinal Gaurav

PROGRAMME: M.Com.

SEMESTER: I

COURSE: Business Statistics

COURSE CODE: CCCOM104

Very short answer type questions:

True or False:

1. Skewness is a measure of the symmetry of a probability distribution about its mean.
2. Bayes' theorem provides a method for updating probabilities based on new evidence or information.
3. The Central Limit Theorem states that the sampling distribution of the sample mean approaches a normal distribution regardless of the shape of the population distribution.
4. Hypothesis testing involves making a decision about a population parameter based on sample data and certain assumptions.
5. The F-test is used to test the equality of variances between two or more groups in an analysis of variance (ANOVA).
6. In multiple regression analysis, the dependent variable is predicted based on one or more independent variables.
7. The mode is always equal to the median in a perfectly symmetrical distribution.
8. The Poisson distribution is commonly used to model the number of successes in a fixed interval of time or space.
9. A confidence interval provides a range of values within which the population parameter is estimated to lie with a certain level of confidence.
10. A Type I error occurs when the null hypothesis is incorrectly rejected when it is actually true.
11. The Chi-square test for independence assesses the association between two categorical variables by comparing observed and expected frequencies.
12. In multiple linear regression analysis, multicollinearity occurs when independent variables are highly correlated with each other.
13. Of course, here are additional true or false statements based on the syllabus topics:
14. The hypergeometric distribution is used to model situations where sampling is done without replacement from a finite population.
15. Standard deviation is a measure of how spread out the values in a data set are around the mean.
16. An alpha level of 0.05 indicates a 95% confidence level in hypothesis testing.
17. The correlation coefficient can range from -1 to 1, where -1 indicates a perfect negative linear relationship, 1 indicates a perfect positive linear

relationship, and 0 indicates no linear relationship.

18. In ANOVA, the F-statistic is calculated as the ratio of between-group variance to within-group variance.

19. In regression analysis, the coefficient of determination (R-squared) measures the proportion of the variance in the dependent variable that is predictable from the independent variables.

20. The binomial distribution is used to model the number of successes in a fixed number of independent Bernoulli trials with the same probability of success.

Multiple Choice Questions:

1. Which of the following measures of central tendency is most affected by extreme outliers?

- a) Mean
- b) Median
- c) Mode
- d) Standard deviation

2. Bayes' theorem is used to:

- a) Calculate probabilities in experiments with a fixed number of trials.
- b) Update probabilities based on new evidence or information.
- c) Determine the spread of a probability distribution.
- d) Estimate the mean of a population from a sample.

3. The Central Limit Theorem states that, for a large sample size, the sampling distribution of the sample mean will:

- a) Approach a uniform distribution.
- b) Approach a normal distribution.
- c) Become skewed to the right.
- d) Be identical to the population distribution.

4. In hypothesis testing, a Type I error occurs when:

- a) The null hypothesis is incorrectly rejected when it is actually true.
- b) The null hypothesis is correctly rejected when it is actually false.
- c) The alternative hypothesis is incorrectly accepted when it is actually false.
- d) The alternative hypothesis is correctly accepted when it is actually true.

5. Which statistical test is used to determine if there is a significant association between two categorical variables?

- a) F-test
- b) t-test
- c) Chi-square test
- d) ANOVA

6. In regression analysis, the coefficient of determination (R-squared) measures:

- a) The strength and direction of the linear relationship between two variables.
- b) The proportion of the variance in the dependent variable that is predictable from the independent variables.

- c) The spread of the data points around the regression line.
- d) The likelihood of observing a particular set of data.

7. Which probability distribution is commonly used to model the number of occurrences of an event within a fixed interval of time or space?

- a) Normal distribution
- b) Binomial distribution
- c) Hypergeometric distribution
- d) Poisson distribution

8. What is the purpose of calculating a confidence interval?

- a) To determine the probability of making a Type I error.
- b) To provide a range of values within which the population parameter is estimated to lie.
- c) To test the null hypothesis.
- d) To compare means of different groups.

9. The F-test is used to test:

- a) The equality of variances between two or more groups.
- b) The significance of correlation coefficients.
- c) The independence of categorical variables.
- d) The normality of data distribution.

10. Which measure of dispersion is not affected by extreme values in a dataset?

- a) Range
- b) Variance
- c) Standard deviation
- d) Interquartile range

11. What is the formula for calculating skewness?

- a) $(\text{Mean} - \text{Median}) / \text{Standard Deviation}$
- b) $3 * (\text{Mean} - \text{Median}) / \text{Standard Deviation}$
- c) $(\text{Mean} - \text{Mode}) / \text{Standard Deviation}$
- d) $3 * (\text{Mean} - \text{Mode}) / \text{Standard Deviation}$

12. If the probability of an event is 0.25, what is the odds of the event occurring?

- a) 0.75
- b) 0.25
- c) 4
- d) $1/4$

13. A normal distribution has a mean of 50 and a standard deviation of 10. What is the z-score corresponding to a value of 60?

- a) 0.5
- b) 1

c) 2

d) 10

14. What is the expected value of a binomial distribution with 10 trials and a success probability of 0.3?

a) 3

b) 5

c) 7

d) 10

15. In a random sample of 100 items, the sample mean is 25, and the sample standard deviation is 5. What is the standard error of the sample mean?

a) 0.5

b) 0.05

c) 0.5

d) 0.05

16. Which of the following statements is true about the p-value in hypothesis testing?

a) A smaller p-value indicates stronger evidence against the null hypothesis.

b) A larger p-value indicates stronger evidence against the null hypothesis.

c) The p-value represents the probability of making a Type I error.

d) The p-value is always greater than the significance level.

17. Which test is appropriate for comparing the means of three or more groups?

- a) Paired t-test
- b) Chi-square test
- c) ANOVA
- d) Mann-Whitney U test

18. If a dataset has a variance of 16 and a standard deviation of 4, how many data points are there?

- a) 16
- b) 8
- c) 4
- d) 64

19. What is the rank correlation coefficient used for?

- a) Measuring the strength and direction of the relationship between two continuous variables.
- b) Measuring the strength and direction of the relationship between two categorical variables.
- c) Testing the independence of two continuous variables.
- d) Testing the equality of variances between two groups.

20. In regression analysis, which of the following statements about residuals is true?

- a) Residuals represent the observed values of the dependent variable.
- b) A positive residual indicates that the actual value is lower than the predicted value.
- c) Residuals should ideally follow a normal distribution.

d) Residuals are calculated as the difference between observed and predicted values.

Very Short Answer Type questions:

1. What is the measure of central tendency that is not influenced by extreme values?
2. What does Bayes' theorem allow us to do with probabilities?
3. What does the Central Limit Theorem state about the sampling distribution of the sample mean?
4. What type of error occurs when the null hypothesis is incorrectly rejected?
5. What statistical test is used to compare the means of two or more groups?
6. What does the coefficient of determination measure in regression analysis?
7. What distribution is commonly used to model the number of occurrences of an event within a fixed interval?
8. What does a confidence interval provide?
9. What does the F-test assess in analysis of variance (ANOVA)?
10. What is the purpose of calculating the standard error of the sample mean?
11. What measure of dispersion describes the spread of data around the mean?
12. What is the probability of an event that cannot occur?
13. What term is used to describe the deviation of a data point from the mean in units of standard deviation?
14. What type of probability distribution is commonly used to model the number of successes in a fixed number of independent trials?
15. What statistical concept is used to estimate population parameters based on sample data?
16. What is the significance level typically denoted as in hypothesis testing?

17. What statistical test is used to determine if there is a significant association between two categorical variables?

18. What is the range of possible values for the correlation coefficient?

19. What is the name of the statistical test used to determine if variances are equal between two or more groups?

20. What does multicollinearity refer to in regression analysis?

Short Answer Type Questions:

1. Explain the concept of skewness and its implications in statistical analysis.

2. Describe the steps involved in hypothesis testing, including the formulation of null and alternative hypotheses, selection of significance level, and interpretation of results.

3. Discuss the importance of the Central Limit Theorem in inferential statistics and its practical implications in real-world data analysis.

4. Explain the difference between point estimation and interval estimation, providing examples of each.

5. Describe the key assumptions underlying the analysis of variance (ANOVA) and explain why violating these assumptions can affect the validity of the results.

6. Discuss the relationship between correlation and causation, providing examples to illustrate the distinction between the two concepts.

7. Explain how multicollinearity can impact the results of multiple regression analysis and discuss strategies for detecting and dealing with multicollinearity.

8. Describe the process of constructing a confidence interval for a population parameter, including the interpretation of the interval and its relationship to the sampling distribution.

9. Discuss the role of the p-value in hypothesis testing and explain how it is used to make decisions about the null hypothesis.

10. Explain the concept of statistical power in hypothesis testing and discuss factors that can influence the power of a statistical test.
11. If a dataset has the following values: 5, 7, 9, 11, 13, what is the mean?
12. A company produces 100 units of a product. If 5% of the units are defective, how many units are defective?
13. The ages of five individuals are 20, 25, 30, 35, and 40 years. What is the median age?
14. If the probability of an event occurring is 0.3, what are the odds in favor of the event occurring?
15. The heights of five students in inches are: 60, 62, 64, 66, and 68. What is the range of heights?
16. A die is rolled 50 times. What is the expected number of times a 3 will appear?
17. The mean score of a class of 20 students on a test is 75. If one student's score was incorrectly recorded as 85 instead of 55, what is the corrected mean score?
18. If a dataset has a standard deviation of 4 and a mean of 10, what is the variance?
19. A factory produces 5000 items per day. If the defect rate is 2%, how many defective items are produced per day?
20. A survey of 50 people found that 20 preferred option A, 15 preferred option B, and the remaining preferred option C. What percentage of people preferred option C?

Long Answer Type Questions:

1. Imagine you are conducting a study to analyze the sales performance of two different marketing strategies. Describe the steps you would take to design and implement the study, including the formulation of research questions, selection of samples, data collection methods, and statistical analysis techniques you would employ to compare the effectiveness of the two strategies.
2. Discuss the concept of sampling distribution and its importance in inferential statistics. Explain how the Central Limit Theorem relates to sampling distribution and its practical implications in estimating population parameters from sample data.

3. Consider a scenario where you are tasked with analyzing the impact of a new training program on employee productivity. Outline the steps you would take to design an experiment to evaluate the effectiveness of the training program, including the selection of experimental and control groups, measurement of productivity, and statistical methods for analyzing the results.
4. Imagine you are conducting a study to assess the relationship between customer satisfaction scores and sales revenue in a retail business. Describe the process of data collection, including the selection of variables, sampling methods, and potential challenges in collecting and analyzing the data. Discuss how you would use correlation and regression analysis to examine the relationship between customer satisfaction scores and sales revenue.
5. Discuss the role of hypothesis testing in statistical analysis. Explain the difference between null and alternative hypotheses and describe the steps involved in hypothesis testing, including setting the significance level, calculating test statistics, and making decisions based on p-values. Provide examples to illustrate the concept of hypothesis testing in real-world scenarios.
6. Consider a scenario where you are analyzing the performance of students in two different schools on a standardized test. Describe the statistical methods you would use to compare the mean scores of the two schools, including the selection of appropriate hypothesis tests, interpretation of results, and potential limitations of the analysis.
7. Imagine you are tasked with analyzing customer feedback data to identify factors that influence customer satisfaction in a retail business. Describe the process of data analysis, including data cleaning, exploratory data analysis, and statistical techniques such as regression analysis or factor analysis. Discuss how you would interpret the results and provide recommendations for improving customer satisfaction based on your analysis.
8. Discuss the concept of multicollinearity in regression analysis. Explain how multicollinearity can affect the estimation of regression coefficients and the interpretation of results. Describe techniques for detecting and dealing with multicollinearity in regression models, including variable selection methods and regularization techniques.
9. Consider a scenario where you are analyzing the relationship between employee engagement and job performance in a company. Describe the steps you would take to conduct a correlation analysis, including the selection of variables, calculation of correlation coefficients, and interpretation of results. Discuss potential limitations of correlation analysis and alternative statistical methods for examining the relationship between employee engagement and job performance.
10. Imagine you are conducting a study to assess the effectiveness of a new drug in treating a medical condition. Describe the process of designing a clinical trial, including the selection of participants, randomization procedures, measurement of outcomes, and statistical methods for analyzing the results. Discuss ethical considerations in conducting clinical trials and potential challenges in interpreting the findings.

11. A dataset consists of the following values: 10, 15, 20, 25, 30. Calculate the mean, median, and mode of the dataset.
12. A factory produces 5000 items per day. If the defect rate is 3%, how many defective items are produced per day?
13. The heights of students in a class are recorded as follows (in inches): 60, 62, 64, 66, 68. Calculate the range and standard deviation of the heights.
14. In a survey, 80 out of 200 respondents preferred option A. Calculate the percentage of respondents who preferred option A.
15. A random sample of 50 students took a test, and their scores were as follows: 70, 75, 80, 85, 90. Calculate the sample standard deviation of the test scores.
16. A company produces 1000 units of a product. If 5% of the units are defective, how many units are not defective?
17. The monthly salaries (in dollars) of a group of employees are: 3000, 3500, 4000, 4500, 5000. Calculate the mean and median monthly salary.
18. The ages (in years) of a group of people are: 25, 30, 35, 40, 45. Calculate the variance and standard deviation of the ages.
19. A bag contains 10 red marbles, 15 blue marbles, and 20 green marbles. If a marble is selected at random from the bag, what is the probability of selecting a blue marble?
20. The lengths (in inches) of fish caught by a group of fishermen are: 12, 14, 16, 18, 20. Calculate the mean and median length of the fish.