



Unit 3: MCQ

**Correlations**

Q.1: What does a correlation coefficient measure?

- a. Causation
- b. Strength and direction of a linear relationship
- c. Standard deviation
- d. Mean

Answer: b. Strength and direction of a linear relationship

Q.2: What is the range of values for a correlation coefficient?

- a. -1 to 0
- b. 0 to 1
- c. -1 to 1
- d. 0 to infinity

Answer: c. -1 to 1

Q.3: If the correlation coefficient is -0.75, what does it indicate about the relationship between two variables?

- a. Strong positive correlation
- b. No correlation
- c. Moderate negative correlation
- d. Strong negative correlation

Answer: d. Strong negative correlation

Q.4: Which of the following correlation coefficients represents the strongest relationship between two variables?

- a. 0.2
- b. -0.8
- c. 0
- d. 0.7

Answer: b. -0.8

Q.5: In a scatter plot, if the points form a straight line that slopes upwards from left to right, what type of correlation is likely present?

- a. Positive correlation
- b. Negative correlation
- c. No correlation
- d. Perfect correlation

Answer: a. Positive correlation

Q.6: When interpreting a correlation coefficient, which value indicates the weakest relationship between variables?

- a. 0.9



Unit 3: MCQ

- b. -0.5
- c. 0.1
- d. -0.8

Answer: c. 0.1

Q.7: What does it mean if the correlation coefficient is exactly 0?

- a. No relationship between variables
- b. Perfect positive correlation
- c. Perfect negative correlation
- d. Weak positive correlation

Answer: a. No relationship between variables

Q.8: Which statistical test is commonly used to assess the significance of a correlation coefficient?

- a. T-test
- b. ANOVA
- c. Chi-square test
- d. Pearson correlation test

Answer: d. Pearson correlation test

Q.9: If the correlation coefficient is close to -1, what does it suggest about the relationship between variables?

- a. Strong positive correlation
- b. Weak negative correlation
- c. Perfect positive correlation
- d. Strong negative correlation

Answer: d. Strong negative correlation

Q.10: What is the primary limitation of using correlation to establish causation?

- a. Correlation does not imply causation
- b. Correlation coefficients are difficult to calculate
- c. Correlation is only applicable to linear relationships
- d. Correlation cannot be used in social sciences

Answer: A

### **Karl Pearson And The Ranking Method**

Q.11: Who is known as the "Father of Statistics"?

- A) Karl Pearson
- B) William Gosset
- C) Ronald A. Fisher
- D) Francis Galton



Unit 3: MCQ

Answer: A) Karl Pearson

Q.12: Karl Pearson is best known for his contributions to which branch of mathematics?

- A) Algebra
- B) Calculus
- C) Statistics
- D) Geometry

Answer: C) Statistics

Q.13: In statistics, what is Pearson's correlation coefficient used to measure?

- A) Association between two variables
- B) Central tendency
- C) Dispersion
- D) Probability

Answer: A) Association between two variables

Q.14: Karl Pearson developed the concept of the "p-value." What does the p-value represent?

- A) Probability of observing the data if the null hypothesis is true
- B) Probability of rejecting the null hypothesis when it is true
- C) Probability of accepting the alternative hypothesis
- D) Probability of a type II error

Answer: A) Probability of observing the data if the null hypothesis is true

Q.15: Which of the following is a statistical method developed by Karl Pearson for comparing the means of two groups?

- A) ANOVA
- B) t-test
- C) Chi-square test
- D) Regression analysis

Answer: B) t-test

Q.16: Karl Pearson's rank correlation coefficient is used to measure the degree of association between two variables when the data are:

- A) Nominal
- B) Ordinal
- C) Interval
- D) Ratio

Answer: B) Ordinal

Q.17: What is the range of Pearson's rank correlation coefficient?

- A) -1 to 0
- B) 0 to 1



Unit 3: MCQ

- C) -1 to 1
- D) 1 to  $\infty$

Answer: C) -1 to 1

Q.18: In the context of Karl Pearson's correlation coefficient, what does a positive value indicate?

- A) Strong negative correlation
- B) Weak positive correlation
- C) Strong positive correlation
- D) No correlation

Answer: C) Strong positive correlation

Q.19: Karl Pearson introduced the concept of the chi-square test for:

- A) Correlation
- B) Goodness of fit
- C) Regression analysis
- D) Analysis of variance

Answer: B) Goodness of fit

Q.20: What is the primary assumption when using Pearson's rank correlation coefficient?

- A) Linearity
- B) Normality
- C) Independence
- D) Homoscedasticity

Answer: A) Linearity

**Regression Expressions:**

Q.21: What is the primary purpose of regression analysis?

- a) Classification
- b) Prediction
- c) Clustering
- d) Association

Answer: b) Prediction

Q.22: Which of the following is true about the slope (coefficient) in a linear regression equation?

- a) It represents the intercept of the line.
- b) It indicates the direction of the relationship between variables.
- c) It measures the spread of data points.
- d) It represents the variance in the dependent variable.

Answer: b) It indicates the direction of the relationship between variables.



Unit 3: MCQ

Q.23: In a simple linear regression equation, what does the y-intercept represent?

- a) The coefficient of the independent variable.
- b) The predicted value of the dependent variable when the independent variable is zero.
- c) The correlation between the variables.
- d) The standard error of the regression.

Answer: b) The predicted value of the dependent variable when the independent variable is zero.

Q.24: What is the purpose of residual analysis in regression?

- a) To determine the correlation between variables.
- b) To assess the fit of the regression model by examining the residuals.
- c) To calculate the slope of the regression line.
- d) To identify outliers in the dataset.

Answer: b) To assess the fit of the regression model by examining the residuals.

Q.25: What is multicollinearity in the context of regression?

- a) The correlation between the dependent and independent variables.
- b) The presence of more than one dependent variable.
- c) High correlation among independent variables.
- d) The spread of residuals in the regression model.

Answer: c) High correlation among independent variables.

Q.26: Which statistical test is used to assess the overall fit of a regression model?

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

Answer: b) F-test

Q.27: In logistic regression, what type of variable is the dependent variable?

- a) Continuous
- b) Categorical
- c) Ordinal
- d) Nominal

Answer: b) Categorical

Q.28: What does the coefficient of determination (R-squared) indicate?

- a) The strength of the relationship between variables.
- b) The slope of the regression line.
- c) The proportion of variance in the dependent variable explained by the independent variable(s).
- d) The p-value of the regression model.



Unit 3: MCQ

Answer: c) The proportion of variance in the dependent variable explained by the independent variable(s).

Q.29: What is the purpose of feature scaling in regression analysis?

- a) To standardize the units of measurement for variables.
- b) To increase the complexity of the regression model.
- c) To remove outliers from the dataset.
- d) To improve the interpretability of regression coefficients.

Answer: a) To standardize the units of measurement for variables.

Q.30: Which of the following statements is true about the residuals in regression?

- a) Residuals should always be normally distributed.
- b) Residuals represent the predicted values of the dependent variable.
- c) Residuals are the differences between observed and predicted values.
- d) Residuals are used to calculate the intercept of the regression line.

Answer: c) Residuals are the differences between observed and predicted values.

**Regression :**

Q.31: What is the primary goal of regression analysis?

- a) Classifying data
- b) Describing relationships between variables
- c) Hypothesis testing
- d) Finding outliers

Answer: b) Describing relationships between variables

Q.32: In simple linear regression, how many variables are involved?

- a) One independent variable and one dependent variable
- b) Two independent variables and one dependent variable
- c) One independent variable and two dependent variables
- d) Two independent variables and two dependent variables

Answer: a) One independent variable and one dependent variable

Q.33: What does the slope of the regression line represent?

- a) Intercept with the y-axis
- b) The strength of the relationship between variables
- c) The average change in the dependent variable for a one-unit change in the independent variable
- d) The residual error

Answer: c) The average change in the dependent variable for a one-unit change in the independent variable



Unit 3: MCQ

Q.34: Which of the following is a measure of goodness-of-fit in regression?

- a) R-squared
- b) Standard deviation
- c) Range
- d) Variance

Answer: a) R-squared

Q.35: What does a p-value in regression analysis indicate?

- a) The strength of the relationship
- b) The probability of observing the data if the null hypothesis is true
- c) The slope of the regression line
- d) The standard error of the residuals

Answer: b) The probability of observing the data if the null hypothesis is true

Q.36: What is multicollinearity in the context of regression?

- a) When the residuals are correlated
- b) When there is a strong linear relationship between two independent variables
- c) When there is a perfect fit between the regression line and the data
- d) When independent variables are highly correlated

Answer: d) When independent variables are highly correlated

Q.37: What is the purpose of the residual plot in regression analysis?

- a) To check for normality of residuals
- b) To check for homoscedasticity
- c) To identify influential outliers
- d) All of the above

Answer: d) All of the above

Q.38: What is the difference between correlation and regression?

- a) Correlation measures the strength and direction of a linear relationship, while regression predicts one variable based on another.
- b) Regression measures the strength and direction of a linear relationship, while correlation predicts one variable based on another.
- c) Correlation and regression are identical concepts.
- d) Correlation is used for categorical variables, and regression is used for continuous variables.

Answer: a) Correlation measures the strength and direction of a linear relationship, while regression predicts one variable based on another.

Q.39: What does a negative coefficient in a regression equation indicate?

- a) A negative relationship between variables
- b) A positive relationship between variables
- c) No relationship between variables



Unit 3: MCQ

d) The intercept with the y-axis

Answer: a) A negative relationship between variables

Q.40: Which type of regression is suitable when there are multiple independent variables?

- a) Simple linear regression
- b) Multiple linear regression
- c) Polynomial regression
- d) Logistic regression

Answer: b) Multiple linear regression

Q.41. What is the primary purpose of regression analysis?

- a. Describing relationships between variables
- b. Testing hypotheses
- c. Making predictions
- d. All of the above

Answer: d. All of the above

Q.42. In a simple linear regression, how many variables are involved?

- a. One
- b. Two
- c. Three
- d. Four

Answer: b. Two

Q.43. Which of the following is the regression equation for a simple linear regression model?

- a.  $Y=a+bX$
- b.  $Y=aX+b$
- c.  $Y=a \cdot X+b$
- d.  $Y=a \div X+b$

Answer: a.)  $Y=a+bX$

Q.44. What does the term "slope" (b) represent in a regression equation?

- a. Intercept
- b. The change in the dependent variable for a one-unit change in the independent variable
- c. Coefficient
- d. Residual

Answer: b. The change in the dependent variable for a one-unit change in the independent variable





Unit 3: MCQ

- Q.45. What is the purpose of the least squares method in regression analysis?
- Minimizing the sum of squared residuals
  - Maximizing the sum of squared residuals
  - Minimizing the sum of absolute residuals
  - Ignoring residuals

Answer: a. Minimizing the sum of squared residuals

- Q.46. In multiple regression, how many independent variables can be included in the model?
- One
  - Two
  - Three or more
  - There is no limit

Answer: c. Three or more

- Q.47. What is the coefficient of determination (R-squared) used for in regression analysis?
- Measuring the strength of the relationship between variables
  - Assessing the goodness-of-fit of the regression model
  - Predicting future values
  - Both a and b

Answer: d. Both a and b

- Q.48. What does a correlation coefficient of -1 indicate in a simple linear regression?
- Perfect positive correlation
  - Perfect negative correlation
  - No correlation
  - Weak correlation

Answer: b. Perfect negative correlation

- Q.49. In a scatter plot, if points cluster closely around a line with a positive slope, what can be said about the correlation?
- Positive
  - Negative
  - Zero
  - Cannot be determined

Answer: a. Positive

- Q.50. Which of the following is true about the residual plot in regression analysis?
- It should show a clear pattern.
  - It should be a straight line.
  - It should be random with no clear pattern.
  - It is not necessary to examine the residual plot.



Unit 3: MCQ

Answer: c. It should be random with no clear pattern.

**Interpolation And Extrapolation Mcqs:**

Q.51. What is interpolation?

- a. Predicting values within the given data range
- b. Predicting values outside the given data range
- c. Both a and b
- d. None of the above

Answer: a

Q.52. Which method involves estimating values between known data points?

- a. Extrapolation
- b. Interpolation
- c. Regression
- d. Prediction

Answer: b

Q.53. What is extrapolation?

- a. Estimating values within the given data range
- b. Estimating values outside the given data range
- c. Both a and b
- d. None of the above

Answer: b

Q.54. Which interpolation method uses a polynomial to estimate values between data points?

- a. Linear interpolation
- b. Cubic interpolation
- c. Exponential interpolation
- d. Power interpolation

Answer: b

Q.55. In linear interpolation, how are values estimated between two data points?

- a. Using a straight line
- b. Using a curve
- c. Using a parabola
- d. Using an exponential function

Answer: a

Q.56. Which of the following is a disadvantage of linear interpolation?

- a. Simple to implement



Unit 3: MCQ

- b. May produce inaccurate results for non-linear data
- c. Works well for all types of datasets
- d. None of the above

Answer: b

- Q.57. What is the purpose of extrapolation in data analysis?
- a. To estimate values within the given data range
  - b. To predict values outside the given data range
  - c. Both a and b
  - d. None of the above

Answer: b

- Q.58. Which interpolation method uses neighboring data points to estimate values?
- a. Cubic interpolation
  - b. Nearest neighbor interpolation
  - c. Linear interpolation
  - d. Exponential interpolation

Answer: b

- Q.59. When should caution be taken while extrapolating data?
- a. Always
  - b. When the data is linear
  - c. When the data is non-linear
  - d. Never

Answer: a

- Q.60. Which method is suitable for estimating values beyond the given dataset based on a curve fit?
- a. Extrapolation
  - b. Interpolation
  - c. Linear regression
  - d. Nearest neighbor interpolation

Answer: c

- Q.61. What is the probability distribution used to model the number of successes in a fixed number of independent Bernoulli trials?
- A. Normal Distribution
  - B. Poisson Distribution
  - C. Binomial Distribution
  - D. Exponential Distribution

Answer: C. Binomial Distribution



Unit 3: MCQ

Q.62. In a binomial distribution, if 'n' is the number of trials and 'p' is the probability of success, what is the mean ( $\mu$ ) of the distribution?

- A.  $\mu = n$
- B.  $\mu = np$
- C.  $\mu = p$
- D.  $\mu = n/p$

Answer: B.  $\mu = np$

Q.63. In a binomial distribution, what is the variance ( $\sigma^2$ ) of the distribution?

- A.  $\sigma^2 = n$
- B.  $\sigma^2 = np$
- C.  $\sigma^2 = p$
- D.  $\sigma^2 = np(1-p)$

Answer: D.  $\sigma^2 = np(1-p)$

Q.64. The condition for a binomial distribution is that each trial must be:

- A. Independent
- B. Dependent
- C. Mutually exclusive
- D. Both A and C

Answer: A. Independent

Q.65. What is the probability of exactly 'k' successes in 'n' trials in a binomial distribution, denoted as  $P(X = k)$ ?

- A.  $\binom{n}{k} p^k (1 - p)^{n - k}$
- B.  $n! / (k! \cdot (n - k)!)$
- C.  $p^k (1 - p)^{n - k}$
- D.  $\frac{1}{2}$

Answer: A

Q.66. The mean and variance of a binomial distribution become approximately equal to each other when:

- A. n is small
- B. p is small
- C. np is large
- D. Both A and B



Unit 3: MCQ

Answer: C. np is large

Q.67. What is the probability of at least 'r' successes in 'n' trials in a binomial distribution, denoted as  $P(X \geq r)$ ?

- A.  $\sum_{k=r}^n \binom{n}{k} p^k (1-p)^{n-k}$
- B.  $\sum_{k=0}^r \binom{n}{k} p^k (1-p)^{n-k}$
- C.  $\binom{n}{r} p^r (1-p)^{n-r}$
- D.  $1 - \binom{n}{r} p^r (1-p)^{n-r}$

Answer : A

Q.68. The continuity correction is often applied when approximating a binomial distribution using a:

- A. Poisson Distribution
- B. Exponential Distribution
- C. Normal Distribution
- D. Hypergeometric Distribution

Answer: C. Normal Distribution

Q.69. Which of the following statements about the binomial distribution is false?

- A. It is symmetric when  $p = 0.5$
- B. The mean is equal to the median
- C. It is always positively skewed
- D. The variance increases as p gets closer to 0 or 1

Answer: C. It is always positively skewed

Q.70. In the binomial expansion formula, what does  $\binom{n}{k}$  represent?

- A. Number of trials
- B. Number of successes
- C. Number of ways to choose 'k' successes from 'n' trials
- D. Probability of success

Answer: C. Number of ways to choose 'k' successes from 'n' trials

Q.71: What is Lagrange interpolation used for in statistics?

- a. Estimating population parameters
- b. Approximating missing data points
- c. Testing hypotheses
- d. Calculating probabilities

Answer: b. Approximating missing data points



Unit 3: MCQ

Q.72: In Lagrange interpolation, what is the degree of the interpolating polynomial?

- a. Equal to the number of data points
- b. One less than the number of data points
- c. Two less than the number of data points
- d. It varies depending on the data set

Answer: b. One less than the number of data points

Q.73: What is the Lagrange basis polynomial for a set of data points  $\{x_0, x_1, x_2\}$ ?

- a.  $L_0(x) = (x - x_1)(x - x_2)$
- b.  $L_1(x) = (x - x_0)(x - x_2)$
- c.  $L_2(x) = (x - x_0)(x - x_1)$
- d. All of the above

Answer: d. All of the above

Q.74: In Lagrange interpolation, what does the weight of each term in the polynomial depend on?

- a. The data point's x-coordinate
- b. The data point's y-coordinate
- c. The degree of the polynomial
- d. The number of data points

Answer: a. The data point's x-coordinate

Q75: What is the main limitation of Lagrange interpolation?

- a. It is computationally intensive
- b. It is sensitive to the distribution of data points
- c. It cannot handle non-linear relationships
- d. It requires a fixed number of data points

Answer: d. It requires a fixed number of data points

Q.76: Lagrange interpolation is most suitable for interpolating data when:

- a. Data points are equally spaced
- b. Data points are randomly distributed
- c. Data points are few and widely spaced
- d. Data points are closely packed

Answer: a. Data points are equally spaced

Q.77: What is the purpose of Lagrange multipliers in optimization problems?



Unit 3: MCQ

- a. To find the maximum value of a function
- b. To handle inequality constraints
- c. To minimize the number of data points
- d. To calculate the mean of a dataset

Answer: b. To handle inequality constraints

Q.78: In Lagrange interpolation, what happens if two data points have the same x-coordinate?

- a. The interpolation is not possible
- b. The Lagrange polynomial becomes undefined
- c. The Lagrange basis polynomials become identical
- d. The interpolation is less accurate

Answer: c. The Lagrange basis polynomials become identical

Q.79: Which of the following methods is an alternative to Lagrange interpolation?

- a. Newton's divided difference
- b. Least squares regression
- c. Principal component analysis
- d. Chi-square test

Answer: a. Newton's divided difference

Q.80: What is the primary advantage of using Lagrange interpolation over linear interpolation?

- a. Higher accuracy
- b. Simplicity of computation
- c. Faster convergence
- d. Ability to handle non-linear relationships

Answer: a. Higher accuracy

**MCQs on Newton's methods**

Multiple-Choice Questions:

Q.81: What is Newton's method primarily used for in statistics?

- a. Hypothesis testing
- b. Estimation of parameters
- c. Data visualization
- d. Random sampling

Answer: b. Estimation of parameters

Q.82: Which of the following is a key assumption of Newton's method?



Unit 3: MCQ

- a. Normal distribution of data
- b. Linearity of the model
- c. Convergence of the initial guess
- d. Independence of observations

Answer: c. Convergence of the initial guess

Q.83: What type of optimization problem does Newton's method address in statistics?

- a. Univariate optimization
- b. Multivariate optimization
- c. Nonlinear optimization
- d. Convex optimization

Answer: b. Multivariate optimization

Q.84: In Newton's method, what is the role of the Hessian matrix?

- a. Measures the curvature of the objective function
- b. Determines the step size in each iteration
- c. Represents the initial guess
- d. Checks the normality assumption

Answer: a. Measures the curvature of the objective function

Q.85: Which term is often used to describe the rate of convergence in Newton's method?

- a. Learning rate
- b. Momentum
- c. Tolerance
- d. Quadratic convergence

Answer: d. Quadratic convergence

Q.86: What happens if the Hessian matrix is not positive definite in Newton's method?

- a. The method converges faster
- b. The method may not converge or give incorrect results
- c. The method becomes computationally less expensive
- d. The method is not affected

Answer: b. The method may not converge or give incorrect results

Q.87: Newton's method is sensitive to the choice of:

- a. Sample size
- b. Random seed
- c. Initial guess
- d. Significance level

Answer: c. Initial guess





Unit 3: MCQ

Q.88: Which of the following statements is true about Newton's method for optimization?

- a. It is always guaranteed to find the global minimum
- b. It is more suitable for linear models
- c. It requires the objective function to be convex
- d. It is not applicable to multivariate problems

Answer: c. It requires the objective function to be convex

Q.89: In logistic regression, what does Newton's method optimize?

- a. Mean squared error
- b. Maximum likelihood
- c. Residual sum of squares
- d. R-squared

Answer: b. Maximum likelihood

Q.90: Which iteration method is commonly used to improve the stability of Newton's method?

- a. Gradient descent
- b. Random search
- c. Levenberg-Marquardt
- d. Markov chain Monte Carlo (MCMC)

Answer: c. Levenberg-Marquardt

Q.91: What is the main advantage of Newton's method over gradient descent?

- a. Simplicity of implementation
- b. Robustness to non-convex functions
- c. Faster convergence in many cases
- d. Independence from initial guess

Answer: c. Faster convergence in many cases

Q.92: In the context of Newton's method, what does a negative eigenvalue of the Hessian matrix indicate?

- a. Convergence
- b. Divergence
- c. Local maximum
- d. Local minimum

Answer: b. Divergence

