DEPARTMENT OF ARTIFICIAL INTELLIGENCE
III B.TECH I SEM
STATISTICS WITH R PROGRAMMING
MID II QUESTION BANK

UNIT III

1. Explain set operations in R with proper examples.
2. Explain different functions in R to access the keyboard and monitor.
3. Demonstrate the process of reading and writing files in R with proper examples.

UNIT IV

1. Discuss about changing the range of axes. Explain about magnifying a portion of the curve.
2. a) Discuss about plot( ) and abline( ) functions with examples.
   b) Explain about two density estimates on the same graph.
3. a) Write R program to create pie chart for the following data
   Housing -------600
   Food --------300
   Clothes ---------150
   Entertainment---100
   Others ---------200
   b) Write an R program to plot the function f(x)= sin(x) in the interval (-3,3) in the steps of 0.1 the point character of the plot is to be a triangle joined with the lines.
4. a) The maximum temperature in Celsius in a week is given as (35,42,38,25,28,36,40). Draw the bar plot for the given data. Also use legend function to describe hists.
   b) Explain the differences between stacked bar plot and bar plot.
5. a) How to plot multiple curves in the same graph? Explain with an example?
   b) Plot the function g(t) = (t2+ 1)0.5 for t between 0 and 5.(using curve and plot function)
6. a) What is Box plot? Explain the importance of boxplot with examples?
   b) Write about scatter plots and histograms with examples? Explain its importance?
7. a) Discuss in detail about Poisson Distributions.
   b) If only 5 percent of kids can secure A grade in a paper, find the probability of at most 2 out of 10 kids getting A grade in that paper.
8. a) Fit a poisson distribution to the following data
   x 0,1,2,3,4,5
   f 3,9,12,27,4,1A
   b) Calculate the coefficient of correlation to the following data
   X 10 12 18 24 23 27
   Y 13 18 12 25 30 10
9. a) Write about Binomial Distribution.
   b) Fit a Binomial distribution to the following data
   x=0 1 2 3 4 5
   f=2 16 28 12 9 3
10. a) Explain about descriptive statistics? Write examples?
    b) Calculate the coefficient of correlation to the following data

x	10	12	18	24	23	27
y	13	18	12	25	30	10

UNIT V

1. What is linear regression? How can linear regression be implemented in R? Explain with an example. What are the advantages & disadvantages of linear regression?
2. What is logistic regression? How can logistic regression be implemented in R? Explain with an example. Compare linear regression with logistic regression.
3. What is Poisson regression? How can Poisson regression be implemented in R? Explain with an example.
4. What is multiple regression? How can multiple regression be implemented in R? Explain with an example
5. The sales of a company (in million dollars) for each year are shown in the table below.

x(year)	2005	2006	2007	2008	2009
y(Sales)	12	19	29	37	45

   a) Find the least square regression line y = ax+b
   b) Use the least square regression line as a model to estimate the sales of the company in 2012
6. Fit a polynomial of degree 2 to the following data

x	0	1	2
y	1	6	17

7. Heights of father and son are given as follows

Father (X)	150	152	155	157	160	161	164	165
Son (y)	154	156	158	159	160	162	161	164

   Fit a regression line predicting the height of the son given the height of father
8. Fit the curve of type $y = ae^{bx}$ to the following data

x	0	1	2	3
y	1.05	2.1	3.85	8.3

9. Fit a parabola to the following data

x	1	2	3	4	5
y	10	12	8	10	14

10. What is survival analysis? How is it implemented in R? Explain with an example.

MULTIPLE CHOICE QUESTIONS

UNIT III

1. Which set operation combines all unique elements from two or more sets into a single set?

   A. Union

   B. Intersection

   C. Difference

   D. Complement

2. In set theory, the intersection of two sets contains:

   A. Elements that are common to both sets.

   B. Elements that are unique to each set.

   C. All elements from both sets.

   D. Elements that are not in either set.

3. The complement of a set A with respect to a universal set U contains:

   A. All elements in set A.

   B. All elements not in set A.

   C. All elements common to sets A and U.

D. All elements unique to set U.

4. In most programming languages, which function is commonly used to display output to the console or monitor?

   A. read()

   B. print()

   C. write()

   D. input()

5. What is the standard way to read input from the keyboard in many programming languages?

   A. read()

   B. input()

   C. print()

   D. display()

6. Which function is used to clear the console or terminal screen to improve readability?

   A. clear()

   B. cls()

   C. refresh()

   D. erase()

7. In file I/O, what does "I/O" stand for?

   A. Internal Operations

   B. Input/Output

   C. Information Optimization

   D. Integrated Operations

8. Which of the following modes would you typically use to open a file for reading in Python?

   A. "w"

   B. "r"

   C. "a"

   D. "x”

9. In Python, what is the primary function used to read the contents of a file?

   A. readfile()

   B. readtext()

   C. read()

   D. readline()

10. When writing to a file in Python, which method is commonly used to add new data to an existing file?

   A. write()

   B. append()

   C. update()

   D. insert()

11. The symmetric difference between two sets A and B contains:

   A. Elements common to both sets.

   B. All elements from both sets.

   C. Elements unique to either set but not both.

   D. Elements not present in either set.

12. If set A is a subset of set B, what is the relationship between the cardinality of A and B?

   A. $|A| = |B|$

B. $|A| < |B|$

C. $|A| > |B|$

D. $|A| = |B| - 1$

13. In set theory, the power set of a set S contains:

A. All elements of set S.

B. All non-empty subsets of set S.

C. Only the empty set.

D. The complement of set S.

14. What is the purpose of standard input and standard output in I/O operations?

A. Standard input is used to display output, and standard output is used to accept input.

B. Standard input is a type of set, and standard output is a file mode.

C. Standard input is used for user input, and standard output is used for program output.

D. Standard input and standard output are the same thing and can be used interchangeably.

15. Which function is commonly used to print a formatted message to the console in Python?

A. printf()

B. writef()

C. fprint()

D. print()

16. In Python, what is the purpose of the input() function?

A. To display text on the console.

B. To read user input from the keyboard.

C. To write data to a file.

D. To clear the console screen.

17. In file I/O, what does the "write" mode ("w") do in Python?

A. Opens the file for reading.

B. Opens the file for writing, creating a new file if it doesn't exist.

C. Opens the file for writing, but raises an error if the file already exists.

D. Appends data to an existing file.

18. In file I/O, what does the "append" mode ("a") do in Python?

A. Opens the file for reading.

B. Opens the file for writing, creating a new file if it doesn't exist.

C. Opens the file for writing, but raises an error if the file already exists.

D. Opens the file for writing, adding data to an existing file.

19 When writing data to a file in Python, which method is commonly used to write a sequence of strings as lines?

A. writelines()

B. write()

C. append()

D. print()

20. In Python, what is the purpose of the close() method in file I/O?

A. To delete the file.

B. To open the file for reading.

C. To close the file after reading or writing.

D. To move the file to a different directory.

ANSWERS

1. A. Union
2. A. Elements that are common to both sets.
3. B. All elements not in set A.
4. B. print()
5. B. input()
6. B. cls()
7. B. Input/Output
8. B. "r"
9. C. read()
10. B. append()
11. C. Elements unique to either set but not both.
12. B. |A| < |B|
13. B. All non-empty subsets of set S.
14. C. Standard input is used for user input, and standard output is used for program output.
15. D. print()
16. B. To read user input from the keyboard.
17. B. Opens the file for writing, creating a new file if it doesn't exist.
18. D. Opens the file for writing, adding data to an existing file.
19. A. writelines()
20. C. To close the file after reading or writing.

UNIT IV

Graphics - Creating Graphs, The Workhorse of R Base Graphics, the plot() Function –Customizing Graphs, Saving Graphs to Files.

Probability Distributions - Normal Distribution- Binomial Distribution- Poisson Distributions, Basic Statistics, Correlation and Covariance.

1. Which R function is often used for creating basic plots in R?

   A. barplot()

   B. hist()

   C. plot()

   D. scatterplot()

2. What is the primary purpose of customizing graphs in R?
   A. To make them more colorful
   B. To improve data visualization and interpretation
   C. To create 3D graphs
   D. To increase the file size of the graphs

3. Which R function is commonly used for saving graphs to files?
   A. save()
   B. export()
   C. write.graph()
   D. pdf(), png(), jpeg(), etc.

4. In R, which function is used to add labels to the x-axis and y-axis in a plot?
   A. axis()
   B. label()
   C. xlabel() and ylabel()
   D. annotate()

5. In R, what is the purpose of the legend() function when customizing graphs?
   A. To create a pie chart
   B. To add a title to the graph
   C. To add a legend describing the elements in the plot
   D. To remove labels from the plot

6. The normal distribution is also known as the:
   A. Uniform distribution
   B. Gaussian distribution
   C. Exponential distribution

D. Logarithmic distribution

7. In the binomial distribution, the random variable represents the number of:

   A. Continuous outcomes

   B. Trials until success

   C. Successes in a fixed number of trials

   D. Failures in a continuous process

8. Which of the following probability distributions is used to model the number of events occurring in a fixed interval of time or space?

   A. Normal distribution

   B. Binomial distribution

   C. Poisson distribution

   D. Exponential distribution

9. In a binomial distribution, if the probability of success (p) is 0.3 and the number of trials (n) is 5, what is the mean (expected value) of the distribution?

   A. 0.3

   B. 1.5

   C. 3

   D. 5

10. The Poisson distribution is commonly used to model:

    A. The number of successes in a fixed number of trials.

    B. Continuous random variables.

    C. The number of events occurring in a fixed interval of time or space.

    D. The difference between two normal distributions.

11. In statistics, the term "mean" refers to:

A. The middle value of a dataset

B. The sum of all data values

C. The most frequent value in a dataset

D. The average of a set of data values

12. What is the range of the correlation coefficient (r) in statistics?

A. -1 to 0

B. 0 to 1

C. -∞ to ∞

D. 0 to ∞

13. Covariance is a measure of:

A. How two variables are related in absolute terms.

B. The strength and direction of the linear relationship between two variables.

C. How two variables are related in relative terms.

D. The ratio of two variances.

14. Which statistic measures the spread or dispersion of a dataset?

A. Median

B. Variance

C. Correlation coefficient

D. Mean

15. What is the purpose of the standard deviation in statistics?

A. To measure the spread of a dataset

B. To calculate the median

C. To estimate the correlation coefficient

D. To calculate the sum of squared differences

Answers

1. C. plot()
2. B. To improve data visualization and interpretation
3. D. pdf(), png(), jpeg(), etc.
4. A. axis()
5. C. To add a legend describing the elements in the plot
6. B. Gaussian distribution
7. C. Successes in a fixed number of trials
8. C. Poisson distribution
9. B. 1.5
10. C. The number of events occurring in a fixed interval of time or space
11. D. The average of a set of data values
12. B. 0 to 1
13. A. How two variables are related in absolute terms.
14. B. Variance
15. A. To measure the spread of a dataset

UNIT V

Linear Models - Simple Linear Regression, Multiple Regression Generalized Linear Models, Logistic Regression, Poisson Regression- other Generalized Linear Models-Survival Analysis.

1. In simple linear regression, what is the dependent variable?

   A. X

   B. Y

   C. Intercept

   D. Slope

2. What does the slope (b) in a simple linear regression model represent?

   A. The predicted Y value when X is 0.

   B. The change in Y for a one-unit change in X.

   C. The Y-intercept.

D. The correlation coefficient between X and Y.

3. What is the purpose of the least squares method in simple linear regression?
   A. To maximize the R-squared value.
   B. To minimize the sum of squared residuals.
   C. To find the best-fitting line.
   D. To maximize the intercept (a) value.
4. In multiple regression, how many independent variables are used to predict the dependent variable?
   A. One
   B. Two
   C. Three or more
   D. It depends on the dataset.

5. The coefficient of determination (R-squared) in multiple regression measures:
   A. The proportion of variance explained by the dependent variable.
   B. The proportion of variance explained by all independent variables combined.
   C. The proportion of variance explained by each independent variable.
   D. The correlation between dependent and independent variables.

6. What is the purpose of the F-statistic in multiple regression?
   A. To measure the strength of the relationship between independent variables.
   B. To test the overall significance of the regression model.
   C. To calculate the standard error of the estimate.
   D. To determine the slope coefficients of the independent variables.

7. In logistic regression, what type of dependent variable is commonly used?
   A. Continuous
   B. Categorical

C. Ordinal

D. Binary

8. Logistic regression is used for:

   A. Predicting continuous outcomes.

   B. Predicting probabilities of binary outcomes.

   C. Analyzing correlations between variables.

   D. Descriptive statistics.

9. In logistic regression, what is the logistic function used for?

   A. Estimating the mean of the dependent variable.

   B. Transforming the independent variables.

   C. Predicting the probability of a specific outcome.

   D. Calculating the residual sum of squares.

10. Poisson regression is commonly used for modeling data that:

    A. Follows a normal distribution.

    B. Involves binary outcomes.

    C. Involves count data, such as the number of events in a fixed interval.

    D. Is strictly linear.

11. Generalized Linear Models (GLMs) extend linear regression by:

    A. Allowing for multiple dependent variables.

    B. Modeling the relationship between dependent and independent variables.

    C. Adapting to various types of dependent variable distributions.

    D. Restricting the types of data that can be analyzed.

12. Which type of distribution is commonly used in logistic regression?

A. Normal distribution

B. Binomial distribution

C. Poisson distribution

D. Exponential distribution

13. Survival analysis is primarily used for studying:

    A. Time-to-event data, such as time until failure.

    B. Categorical data.

    C. Continuous data.

    D. Linear relationships between variables.

14. In survival analysis, the Kaplan-Meier estimator is used to:

    A. Estimate the survival probability over time.

    B. Estimate the mean survival time.

    C. Calculate hazard ratios.

    D. Determine the standard error of the estimate.

15. What is the primary output of a survival analysis?

    A. R-squared value

    B. Hazard ratio

    C. Kaplan-Meier survival curve

    D. p-value

Answers

1. B. Y
2. B. The change in Y for a one-unit change in X.
3. B. To minimize the sum of squared residuals.
4. C. Three or more
5. B. The proportion of variance explained by all independent variables combined.
6. B. To test the overall significance of the regression model.

7. B. Categorical
8. B. Predicting probabilities of binary outcomes.
9. C. Predicting the probability of a specific outcome.
10. C. Involves count data, such as the number of events in a fixed interval.
11. C. Adapting to various types of dependent variable distributions.
12. B. Binomial distribution
13. A. Time-to-event data, such as time until failure.
14. A. Estimate the survival probability over time.
15. C. Kaplan-Meier survival curve