Week 1

1. Question: Choose the wrong statement

A. Statistics helps in summarizing
B. Statistics helps in classifying numerical facts
C. Statistics studies individuals
D. Statistical laws are not exact

Answer: C

Explanation: PPT 1.a Slide Number: 9

2. Question: Which among these is not generally used as R based GUI for Statistics

A. R Studio
B. RKward
C. R Commander
D. Rattle

Answer: A

Explanation: PPT 1.b Slide Number: 10

3. Question :The place where we write commands to instruct R

A. Workspace
B. Script
C. Console
D. Package

Answer: C

Explanation : PPT 1.b Slide Number: 12

4. Question :Which among these can not be imported in RKward ?

A. XLSX
B. SPSS
C. Stata
D. OCX

Answer: D

Explanation : PPT 1.c Slide Number: 4

5. Question :Which among these data types are not supported in RKward ?

A. Logical
B. Factor
C. Cardinal
D. String

Answer: C

Explanation : PPT 1.c Slide Number: 13

6. Question :Which among these data types are not statistical data types ?

A. Ratio
B. Factor
C. Ordinal
D. Nominal

Answer: B

Explanation : PPT 1.c Slide Number: 14

7. Question :Where do we find attributes of variables in Statistical Software ?

A. Data View
B. Variable View
C. Help Menu
D. Descriptive Statistics

Answer: B

Explanation : PPT 1.c Slide Number: 13

8. Question :Which of these represents frequency distribution ?

A. Histogram
B. Box plot
C. Scatterplot
D. Descriptive Statistics

Answer: A

Explanation : PPT 1.e Slide Number: 3

9. Question :Which of these helps us to see descriptive statistics ?

A. Histogram
B. Box plot
C. Scatterplot
D. Ogive

Answer: B

Explanation : PPT 1.e Slide Number: 4

10. Question :Which among these can not be used in Statistical Analysis ?

A. Excel
B. SPSS
C. R
D. MS-Word

Answer: D

Explanation : MS word edits text.

Week 2

1. Question: Which is not an alternative name for the Galton Board?

A. Quincunx
B. Bean machine
C. Pascal Triangle
D. Poisson Table

Answer: D

Explanation: PPT g Slide Number: 6

2. Question: Which one is the central tendency?

A. Geometric Mean
B. Range
C. Variance
D. Correlation

Answer: A

Explanation: PPT of Slide Number: 5

3. Question: Which is the least sensitive measure of central tendency?

A. Median
B. Mean
C. Variance
D. Mode

Answer: A

Explanation: PPT of Slide Number: 10

4. Question: Which among these is correct in context of Negatively Skewed Data?

A. Mean=Mode=Median
B. Mean>Mode>Median
C. Median>Mode>Mean
D. Median>Mode<Mean

Answer: C

Explanation: PPT of Slide Number: 15

5. Question: Which among these represent the median?

A. Boundary of second and third quartile
B. Boundary of second and first quartile
C. Boundary of fourth and third quartile
D. None of the above

Answer: A

Explanation: PPT d Slide Number: 7

6. Question: Which among these second moments mean?

A. Skewness
B. Variance
C. Kurtosis
D. Mean

Answer: B

Explanation: PPT d Slide Number: 11

7. Question: Which among these is unitless?

A. Variance
B. Mean
C. Coefficient of Variance
D. range

Answer: C

Explanation: PPT d Slide Number: 22

8. Question: Which among these is not a measure of variability?

A. Range
B. Variance
C. Scatterplot
D. Interquartile Range

Answer: C

Explanation: PPT d Slide Number: 4

9. Question: Which among these can be used for curve fitting?

A. Scatter-Plot
B. Median
C. Mean
D. Correlation

Answer: A

Explanation: Scatter plot gives initial view

10. Question: Which of these can be used as a measure of model fit?

A. Scatter-Plot
B. Mean Square Error
C. Mean
D. Correlation

Answer: B

Week 3

1. Question: pnorm(-z) =?

A. 1 - pnorm(z)
B. - pnorm(z)
C. pnorm(z)
D. pnorm(1-z)

Answer: C

Explanation: Explanation : PPT g Slide Number : 28

2. Question: What is the variance in a Bernoulli Trial

A. np
B. npq
C. p/n
D. npn

Answer: A

Explanation: PPT g Slide Number: 11

3. Question: If in Gaussian function probability of success < 0.5, then the normal plot is

A. Left Skewed
B. Right Skewed
C. Symmetric
D. Insufficient data

Answer: B

Explanation: PPT g Slide Number: 11

4. Question: Leptokurtic deviation in normal graph means

A. Peakedness of frequency distribution
B. Flater than normal graph
C. More data away from mean
D. as same as normal graph

Answer: A

Explanation: PPT g Slide Number: 23

5. Question: What approaches as it heads toward infinity but never actually intersects with the curve.

A. Asymptote
B. Hyperbola
C. Parabola
D. Straight Line

Answer: A

Explanation: In the video

6. Question: The shape of cumulative density function of normal function is

A. Not Defined
B. Sigmoidal
C. Parabola
D. Straight Line

Answer: A

Explanation: PPT Slide Number: 16

7. **Question: Z table gives**

A. Probability Density Function
B. Cumulative Density Function
C. Circle
D. Straight Line

Answer: B

Explanation: PPT Slide Number: 19

8. **Question: In the the normal x**

A. 50
B. 68
C. 95
D. 99.7

Answer: B

Explanation: PPT Slide Number: 26

9. Question: In the Pascal triangle of 6 levels, if 100 balls are passed at level 1. How many balls will arrive at the fourth slot from right?

A. 25
B. 75
C. 50
D. 50

Answer: A

Explanation: 3 successes out of 6 trials with p = 0.5

10. Question: pnorm (1,0,1)= ?

A. 0.5
B. 0.16
C. 0.84
D. 0.7

Answer: C

Explanation: The function pnorm(1, 0, 1) calculates the cumulative distribution function (CDF) of a standard normal distribution at the value 1

Week 4

1. Question: Accept null hypothesis which is false

A. Type 2 error
B. Type 1 error
C. Type 3 error
D. not able to define

Answer: A

Explanation: PPT h Slide Number: 4

2. Question: x+y+z=3, x+y-z=4, What is the degree of freedom?

A. insufficient info
B. 2
C. 0
D. 1

Answer: D

Explanation: PPT i slide number: 10

3. Question: Find the odd one out

A. Unexplained Variance
B. Random Variance
C. Residual Variance
D. Model Variance

Answer: D

Explanation: PPT i slide number: 9

4. Question: X could be calculated by taking the difference between each sample mean and the overall mean, squaring these differences, adding them up, and then dividing by the number of samples. X =?

A. Random Variance
B. Confidence Interval
C. Standard Error
D. Estimate

Answer: C

Explanation: PPT i slide number: 2

5. Question: Confidence Interval should be

A. Low
B. High
C. There is no constraint
D. depends on researcher

Answer: D

Explanation: PPT i slide number: 7

6. Question: **if b represents the estimated coefficient, and SE represents the standard error of the coefficient estimate. then**

A. t =b/SE
B. t = b * SE
C. b =t/SE
D. SE = t*b

Answer A

Explanation: PPT i slide number: 16

7. Question: Effectiveness of drug needs to be tested on patients. Hence pre-test and post-test parameters were taken for comparison. Which test is appropriate

A. One sample t-test
B. Paired Sample test
C. Two sample t- test
D. none of these

Answer: B

Explanation: PPT i slide number: 25

8. Question: What does Analysis of Variance (ANOVA) calculate?

A. z-score
B. t-score
C. chi-square
D. F-score

Answer: D

Explanation: PPT J slide number: 11

9. Question: To use a t-test, dependent variable must have

A. Nominal
B. Ordinal
C. Ratio
D. d: none of these

Answer: C

Explanation: Mean is possible for ratio data

10. Question: The degree of freedom for Mean Square error explained is

A. N-1
B. k-1
C. N-k
D. zero

Answer: B

Explanation: PPT J slide number: 18

Week 5

1. Question: For what type of data is Spearman correlation appropriate

A. Nominal
B. Ordinal
C. Ratio
D. All of the above

Answer: C

Explanation: PPT K Slide Number: 4

2. Question: Which among these serves the same purpose as correlation?

A. Covariance
B. Regression coefficient
C. Variance
D. Z-score

Answer: A

Explanation: PPT K Slide Number: 8

3. Question: Which of the following is true for the coefficient of correlation?

A. The coefficient of correlation is not dependent on the change of scale.
B. The coefficient of correlation is not dependent on the change of origin
C. The coefficient of correlation is not dependent on both the change of scale and change of origin.
D. None of the above

Answer: C

Explanation: It's a unitless and measurement of association

4. Question: What is the null hypothesis for correlation in inferential statistics?

A. There is no association between variables.
B. There is an association between independent and dependent variables.
C. There is no hypothesis in correlation
D. It depends on convenience of the researcher

Answer: A

Explanation: In correlation analysis, the null hypothesis typically states that there is no association or correlation between the two variables being tested. In other words, the population correlation coefficient
is equal to zero, indicating no linear relationship.

5. Question: Line x= 1,2,3,1,4,5, y= 2x+1, z=3y+2. What can be said the correlation coefficients a= rxy and b=rxz, then

A. a=b
B. a<b
C. a>b
D. insufficient data.

Answer: A

Explanation a=b=1, as both are perfect straight line

6. Question: If x= 1,2,3,1,4,5, y= 2x+1, z=3x+2. What can be said the regression coefficients byx and bzx, then

A. bzx = byx
B. bzx > byx
C. bzx < byx
D. bxz = bxy

Answe : B

Explanation bzx = 3, byx=2

7. Question: Which statement accurately describes heteroscedasticity in regression analysis?

A. Heteroscedasticity arises when the residuals in a regression model exhibit a systematic pattern or structure, violating the assumption of constant variance.
B. Heteroscedasticity implies that the variance of the residuals remains consistent across all levels of the independent variables
C. Heteroscedasticity leads to unbiased and efficient estimates of the regression coefficients.
D. Heteroscedasticity only affects the precision of the estimates and has no impact on the validity of the regression model.

Answer: A

Explanation: This violates one of the key assumptions of linear regression, which assumes that the variance of the error terms (residuals) is constant. Heteroscedasticity can lead to biased and inefficient estimates of the regression coefficients, affecting the validity of the statistical inference drawn from the regression analysis.

8. Question: What does the adjusted R-squared measure in regression analysis?

A. The adjusted R-squared measures the proportion of variance in the dependent variable that is explained by the independent variables in the model.
B. The adjusted R-squared measures the proportion of variance in the independent variables that is explained by the dependent variable.
C. The adjusted R-squared accounts for the number of independent variables in the model and penalizes excessive complexity, providing a more accurate assessment of the model's goodness of fit.
D. The adjusted R-squared is identical to the regular R-squared and does not provide any additional information about the regression model

Answer: C

Explanation: Adjusted R-squared adjusts the regular R-squared (which measures the proportion of variance explained by the independent variables) by penalizing the inclusion of additional independent variables. This adjustment helps to prevent overfitting and provides a more reliable measure of how well the model fits the data.

9. Question: What does multiple regression analysis involve?

A. Multiple regression analysis involves predicting a single dependent variable based on two or more independent variables.
B. Multiple regression analysis involves predicting multiple dependent variables based on a single independent variable.
C. Multiple regression analysis involves predicting multiple independent variables based on a single dependent variable.
D. Multiple regression analysis involves comparing the means of two or more groups.

Answer: A

Explanation: Multiple regression analysis allows us to predict a single dependent variable based on the values of two or more independent variables. It helps in understanding the relationship between the dependent variable and multiple predictors simultaneously.

10. Question: What does a Q-Q plot (Quantile-Quantile plot) assess regarding data distribution?

A. A Q-Q plot assesses the linearity of the relationship between two variables.
B. A Q-Q plot assesses the normality of the distribution of a single variable by comparing its quantiles to those of a theoretical normal distribution.
C. A Q-Q plot assesses the homogeneity of variance across groups in an analysis of variance (ANOVA).
D. A Q-Q plot assesses the multicollinearity among independent variables in a regression analysis.

Answer: B

Explanation: A Q-Q plot compares the quantiles of a dataset to the quantiles of a theoretical distribution, usually a normal distribution. This comparison helps to visually assess whether the data are approximately normally distributed.

Week 6

1. Question: What is the chi-square test used for?

A. Testing the difference between two population means
B. Testing the difference between two population proportions
C. Testing the association between two categorical variables
D. Testing the correlation between two continuous variables

Answer: C

Explanation: Testing the association between two categorical variables

2. Question: In a chi-square test, the null hypothesis states

A. There is a significant difference between groups
B. There is no significant relationship between groups
C. The groups are perfectly correlated
D. The groups have identical means

Answer: B

Explanation: In a chi-square test, the null hypothesis typically states that there is no significant association or relationship between the categorical variables being tested. The chi-square test checks if the observed frequencies differ significantly from expected frequencies under the assumption that the null hypothesis is true.

3. Question: When is the chi-square test considered appropriate?

A. When sample sizes are small
B. When data is continuous
C. When data is normally distributed
D. When data is categorical

Answer: D

4. Question: Which among this is non-parametric test

A. One-Way ANOVA
B. Friedman test
C. Regression
D. Correlation

Answer: B

Explanation: PPT L Slide: 8

5. Question: Choose the correct definition for Posterior Probability

A. refers to the probability of an event occurring after considering new evidence or information.
B. refers to the initial probability of an event occurring based on logic or prior knowledge, without considering any additional evidence.
C. refers to the initial probability of an event occurring based on logic or prior knowledge, with considering any additional evidence.
D. none of these

Answer: A

Explanation PPT L Slide 5

6. Question: If P(B|A) = P(B), then

A. P (A intersection B) = P(A).P(B)
B. P (A intersection B) = P(A)
C. P (A intersection B) = .P(B)
D. P (A intersection B) = P (A union B)

Answer: A

Explanation: PPT L slide 2

7. Question: which one is an incorrect command in R?

A. glm(y ~x)
B. lm(y ~x)
C. rlm(y ~x)
D. regr(y ~x)

Answer: A

Explanation: PPT N slide 8, rlm(y ~x) is part of the MASS package and is used to fit robust linear models.

8. Question: y = 1/ (1 + exp (k - x)), if x is independent variable, k is constant, then y represents

A. Logistic function
B. Exponential function
C. Logarithmic function
D. Sigmoidal function

Answer: D

Explanation: PPT M slide 2

9. Question: If probability of winning is 4/7, The odds ratio in favor is

A. 4:3
B. 3:7
C. 3:4
D. 4:7

Answer: A

Explanation: PPT N slide 3

10. Question: if the dependent variable is binary, then which technique can be applied

A. chi square and Logistic regression
B. logistic and linear regression
C. linear regression and ANOVA
D. ANOVA and logistic regression

Answer: A

Explanation PPT N slide: 2

Week 7

1. Question: What is a time series?

A. A collection of data points indexed in time order
B. A statistical method for analyzing trends in data
C. A type of regression analysis
D. A machine learning algorithm

Answer: A

Explanation: A time series is a collection of data points indexed in time order, typically at regular intervals.

2. Question: What are the components of a time series?

A. Trend, season, and cycle
B. Mean, median, and mode
C. Variance, standard deviation, and range
D. Mean, trend, and seasonality

Answer: A

Explanation: The components of a time series are trend, season, and cycle.

3. Question: What is stationarity in the context of time series analysis?

A. The property where the mean, variance, and auto covariance of the series do not change over time
B. The process of converting non-stationary data into stationary data
C. The measure of how well a model predicts future observations
D. The property where data points are evenly spaced in time

Answer: A

Explanation: Stationarity refers to the property where the mean, variance, and auto covariance of the series do not change over time.

4. Question: What are some common forecasting methods used in time series analysis?

A. Moving average, exponential smoothing, ARIMA
B. Linear regression, logistic regression, decision trees
C. ANOVA, MANOVA, ANCOVA
D. Clustering, principal component analysis, factor analysis

Answer: A

Explanation: Common forecasting methods used in time series analysis include moving average, exponential smoothing, and ARIMA (Autoregressive Integrated Moving Average).

5. Question: What does the statistical expectation of a random variable represent?

A. The most likely value of the variable
B. The median of the variable
C. The mean of the variable
D. The range of the variable

Answer: C

Explanation: The statistical expectation of a random variable represents the mean of the variable.

6. Question: In which scenario would the Poisson distribution be most appropriate?

A. Modeling the number of heads obtained when flipping a fair coin multiple times.
B. Estimating the average height of students in a classroom.
C. Predicting the number of customers arriving at a store within a given hour.
D. Analyzing the distribution of scores on a standardized test.

Answer: C

Explanation: The Poisson distribution is commonly used to model the number of events occurring within a fixed interval of time or space, such as the number of customers arriving at a store within a given hour.

7. Question: What is the main difference between the hypogeometric and normal distributions?

A. The hypogeometric distribution is discrete, while the normal distribution is continuous.
B. The hypogeometric distribution has a single peak, while the normal distribution is symmetric.
C. The hypogeometric distribution is used for modeling time series data, while the normal distribution is used for modeling cross-sectional data.
D. There is no difference; they are both continuous distributions.

Answer: A

Explanation: The main difference between the hypo geometric and normal distributions is that the hypo geometric distribution is discrete, while the normal distribution is continuous.

8. Question: What type of data is commonly modeled using a negative binomial distribution?

A. Continuous data
B. Categorical data
C. Count data with overdispersion
D. Time series data

Answer: C

Explanation: A negative binomial distribution is commonly used to model count data with overdispersion, where the variance exceeds the mean.

9. Question: In time series analysis, what does the term "effect size" typically refer to?

A. The size of the time series dataset
B. The magnitude of the relationship between variables
C. The length of the time period covered by the dataset
D. The complexity of the forecasting model

Answer: B

Explanation: "effect size" typically refers to the magnitude of the relationship between variables.

10. Question: What is the term for the repetition of patterns at fixed intervals in a time series?

A. Trend
B. Seasonality
C. Cycle
D. Random noise

Answer: B

Explanation: Seasonality refers to the repetition of patterns at fixed intervals in a time series, such as daily, weekly, or yearly patterns.

Week 8

1. Question: What is RStudio?

A. A programming language
B. An integrated development environment (IDE) for R
C. A statistical analysis software
D. A database management system

Answer: B

Explanation: R Studio is an integrated development environment (IDE) designed specifically for working with the R programming language.

2. Question: How can you install an R package named "ggplot2"?

A. install. Packages("ggplot2")
B. library("ggplot2")
C. require("ggplot2")
D. source("ggplot2")

Answer: A

Explanation: To install an R package named "ggplot2," you can use the install. packages () function in R.

3. What is the purpose of documentation in R?

A. To provide tutorials on using R
B. To document the functions and datasets available in packages
C. To list programming languages similar to R
D. To describe the history of R development

Answer: B

Explanation: Documentation in R is used to describe the functions and datasets available in packages, along with their usage and parameters.

4. What are the primary differences between RStudio and RKWard?

A. RStudio is an IDE for R, while RKWard is a standalone statistical analysis software.
B. RStudio provides a graphical user interface (GUI), while RKWard is a command-line interface (CLI).
C. RStudio is designed for beginners, while RKWard is designed for advanced users.
D. RStudio is free, while RKWard requires a subscription.

Answer: A

Explanation: RStudio is an integrated development environment (IDE) for R, whereas RKWard is a standalone statistical analysis software.

5. Question: What is the function of the R Console?

A. To execute R scripts
B. To write documentation
C. To visualize data
D. To access R packages

Answer: A

Explanation: The R Console is used to execute R scripts, commands, and expressions.

6. Question: What does the R Environment represent?

A. The physical location where R is installed
B. The memory space where R objects are stored and manipulated during an R session
C. The online community of R users
D. The folder structure of R packages

Answer: B

Explanation: The R Environment represents the memory space where R objects are stored and manipulated during an R session.

7. Question: Which of the following are the four panels in RStudio?

A. Script Editor, Console, History, Environment/Files
B. Data, Models, Visualizations, Console
C. Code, Output, Plots, Files
D. Editor, Console, Terminal, Files

Answer: A

Explanation: The four panels in RStudio are Script Editor, Console, History, and Environment/Files.

8. Question: What is R Markdown?

A. A programming language similar to R
B. A file format for saving R scripts
C. A markup language that allows integration of R code and text to produce dynamic documents
D. A graphical user interface (GUI) for R

Answer: C

Explanation: R Markdown is a markup language that allows integration of R code and text to produce dynamic documents, including reports, presentations, and websites.

9. Question: Which of the following RStudio panels is used for writing and editing R scripts?

A. Plots
B. Environment
C. Files
D. Source

Answer: D

Explanation: The Source panel in RStudio is used for writing and editing R scripts.

10. What are R packages?

A. Pre-installed datasets in R
B. Functions and datasets bundled together for specific purposes
C. Programming languages used within R
D. Text files containing R code

Answer: B

Explanation: R packages are collections of functions and datasets bundled together for specific purposes, such as data analysis, visualization, or statistical modeling.

11. Question: In RStudio, what does the Console panel primarily serve for ?

A. Displaying plots and visualizations
B. Editing R scripts
C. Executing R commands and displaying their output
D. Managing files and directories

Answer: C

Explanation: In RStudio, the Console panel primarily serves for executing R commands and displaying their output.