1. The graphical technique that can be used before one applies the ANOVA is the
2. Boxplot
3. Bar Diagram
4. Histogram
5. Scatter Diagram

Correct answer: a

1. When shall we use the Tukey test in ANOVA
2. In case of multiple number of observations per cell
3. When the group means differ significantly
4. When the boxplots several outliers
5. Before performing ANOVA test

Correct Answer: b

1. The p-value of a one-way ANOVA is 0.04385. What shall be your conclusion at 5% level of significance.
2. The null hypothesis is rejected
3. The null hypothesis is accepted
4. Decision depends on the value of standard error
5. Need to perform two-way ANOVA to confirm

Correct Answer: (a)

1. Which of the following is not an assumption of ANOVA
2. The observations are independent.
3. The parent population from which the observations are taken is normal in nature.
4. Various treatment and environmental effects are additive in nature.
5. The correlations between the populations are zero.

Correct Answer: (d)

1. In Multivariate ANOVA the dependent variables are
2. Having the same units of measurement
3. Needs to be normally distributed
4. Correlated to each other
5. Are linearly related to each other

Correct Answer: (b)

1. The MANOVA is used when there are
2. Several number of dependent variables
3. One dependent variable and several independent variables
4. Data is classified in two dimensions
5. When each of the experimental material are at different levels

Correct Answer: (a)

1. The non-parametric counterpart of one way ANOVA is
2. Friedman’s Test
3. Kruskal-Wallis rank sum test
4. Wilcoxon Rank Sum test
5. Spearman’s test

Correct Answer: (b)

1. If the sd of X is σ , then standard error of X ̅ (Xbar) (n=sample size >2) is  
   (a) σ  
   (b) σ/√(n+1)  
   (c) σ/√n  
   (d) σ/√(n-1)

Correct Answer: (c)

1. Which of the following is true?  
   (a) Parameter is a characteristic of Population and Statistic is a characteristic of Sample  
   (b) Parameter is a characteristic of Sample and Statistic is a characteristic of Population.  
   (c) Parameter is a characteristic of Population and Statistic is a characteristic of Population  
   (d) Parameter is a characteristic of Sample and Statistic is a characteristic of Population

Correct Answer: (a)

1. Which of the following CAN NOT be a Statistical Hypothesis, where m is the population mean and X ̅ is the sample mean?  
   (a) m = 0  
   (b) m > 0  
   (c) X ̅ < 0  
   (d) µ

Correct Answer: (c)

1. Which of the following is correct (Notations have their usual significance)?

(a) Pr( tn-1 > t a, n-1) = a, Pr( tn-1 > - t a, n-1) = 1- a  
(b) Pr( tn-1 > t a, n-1) = 1-a, Pr( tn-1 > - t a, n-1) = a  
(c) Pr( tn-1 < t a, n-1) = a, Pr( tn-1 < - t a, n-1) = a

(d) Pr( tn-1 < t a, n-1) = 1- a, Pr( tn-1 < - t a, n-1) =1- a  
 Correct Answer: (a)

12. Let p1 and p2 be the population proportion of successes of two groups and p1 = p2 =p (unknown). Suppose the number of successes for the first group is 30 out of 60 and the number of successes for the first group is 50 out of 80. If we consider estimate of p1 as 30/60 and estimate of p2 as 50/80 what is the pooled estimate of p?  
 (a) 0.5514286  
 (b) 0.5625  
 (c) 0.5714286

(d) 0.5825  
 Correct Answer: (c)

13. Which of the following is a test for Normality?  
 (a) Kruskal-Wallis Test  
 (b) Shapiro-Wilk test  
 (c) Wilcoxon Rank-Sum Test  
 (d) Mann-Whitney U-test

Correct Answer: (b)

14. Consider two independent random samples X1, X2, X3, X4 ,…Xm and Y1, Y2, Y3, Y4 ,…Yn. Which of the following is the Mann-Whitney U-statistic?

(a) Number of times a Y precedes an X in the combined ordered arrangement of the above two independent random samples

(b) Sum of the ranks of the X’s in the combined ordered arrangement of the above two independent random samples

(c) Product of the ranks of the X’s in the combined ordered arrangement of the above two independent random samples

(d) Sample Median of X’s – Sample Median of Y’s in the combined ordered arrangement of the above two independent random samples

Correct Answer: (a)

15. Arithmetic mean of a set of 20 observations is 500. If a value of 35 is added , the new arithmetic will

(a) remain the same

(b) increase

© decrease

(d)mean will not be affected

Correct Answer: (c)

16. Mode is used for

(a) Ordinal scale measure

(b) Ratio Scale measure

© Nominal Scale measure

(d)Interval Scale measure

Correct Answer: (c)

17. In a negatively skewed distribution

(a) median is less than mean

(b) mean is greater than median and mode

© mode is greater than mean but less than median

(d)median is between mean and mode.

Correct Answer: (d)

18. Mean ± sd of a set of normally distributed data covers

(a) 30% of the observations

(b) At least 68% of the observations

© 68% of the observations

(d)95% of the observations

Correct Answer: (b)

19. In a leptokurtic distribution

(a) the peak is narrow and high

(b) the peak is flat

© standard deviation is high

(d)none of these

Correct Answer: (a)

20. Coefficient of variation is

(a) an absolute measure of variation

(b) a relative measure of variation

© not a measure of variation

(d)none of these

Correct Answer: (b)

**Multiple Choice Questions for 09.08.2020**

**1. In practice, Line of best fit or regression line is found when \_\_\_\_\_\_\_\_\_\_\_\_\_  
a) Sum of residuals (∑(Y – h(X))) is minimum  
b) Sum of the absolute value of residuals (∑|Y-h(X)|) is maximum  
c) Sum of the square of residuals ( ∑ (Y-h(X))2) is minimum  
d) Sum of the square of residuals ( ∑ (Y-h(X))2) is maximum  
Answer: c  
2. Which of the following metrics can be used for evaluating regression models?**

**i) R Squared**

**ii) Adjusted R Squared**

**iii) F Statistics**

**iv) RMSE / MSE / MAE**

**a) ii and iv  
b) i and ii  
c) ii, iii and iv  
d) i, ii, iii and iv  
Answer: d  
3. How many coefficients do you need to estimate in a simple linear regression model (One independent variable)?  
a) 1  
b) 2  
c) 3  
d) 4  
Answer: b  
4. In a simple linear regression model (One independent variable), If we change the input variable by 1 unit. How much output variable will change?  
a) by 1  
b) no change  
c) by intercept  
d) by its slope**

**Answer: d  
5. In the mathematical Equation of Linear Regression Y = β1 + β2X + ϵ, (β1, β2) refers to \_\_\_\_\_\_\_\_\_\_  
a) (X-intercept, Slope)  
b) (Slope, X-Intercept)  
c) (Y-Intercept, Slope)  
d) (slope, Y-Intercept)**

**Answer: c**

**6. Which of the one is true about Heteroskedasticity?**

**A. Linear Regression with varying error terms**

**B. Linear Regression with constant error terms**

**C. Linear Regression with zero error terms**

**D. None of these**

**Solution: A**

**7. Which of the following indicates a fairly strong relationship between X and Y?**

**A. Correlation coefficient = 0.9**

**B. The p-value for the null hypothesis Beta coefficient =0 is 0.0001**

**C. The t-statistic for the null hypothesis Beta coefficient=0 is 30**

**D. None of these**

**Solution: A**

**8. Which of the following assumptions do we make while deriving linear regression parameters?**

1. **The true relationship between dependent y and predictor x is linear**
2. **The model errors are statistically independent**
3. **The errors are normally distributed with a 0 mean and constant standard deviation**
4. **The predictor x is non-stochastic and is measured error-free**

**A. 1,2 and 3.**

**B. 1,3 and 4.**

**C. 1 and 3.**

**D. All of above.**

**Solution: D**

**9. To test linear relationship of y(dependent) and x(independent) continuous variables, which of the following plot best suited?**

**A. Scatter plot**

**B. Barchart**

**C. Histograms**

**D. None of these**

**Solution: A**

**10. Generally, which of the following method(s) is used for predicting continuous dependent variable?**

1. **Linear Regression**
2. **Logistic Regression**

**A. 1 and 2**

**B. only 1**

**C. only 2**

**D. None of these.**

**Solution: B**

**11. A correlation between age and health of a person found to be -1.09.  On the basis of this you would tell the doctors that:**

**A. The age is good predictor of health**

**B. The age is poor predictor of health**

**C. None of these**

**Solution: C**

**12. Suppose you are training a linear regression model. Now consider these points.**

1. **Overfitting is more likely if we have less data**
2. **Overfitting is more likely when the hypothesis space is small**

**Which of the above statement(s) are correct?**

**A. Both are False**

**B. 1 is False and 2 is True**

**C. 1 is True and 2 is False**

**D. Both are True**

**Solution: C**

**13. Which of the following statement(s) can be true post adding a variable in a linear regression model?**

1. **R-Squared and Adjusted R-squared both increase**
2. **R-Squared increases and Adjusted R-squared decreases**
3. **R-Squared decreases and Adjusted R-squared decreases**
4. **R-Squared decreases and Adjusted R-squared increases**

**A. 1 and 2**

**B. 1 and 3**

**C. 2 and 4**

**D. None of the above**

**Answer: A**

**14. If two variables are correlated, is it necessary that they have a linear relationship?**

**A. Yes**

**B. No**

**Solution: B**

**15. Suppose the distribution of salaries in a company X has median $35,000, and 25th and 75th percentiles are $21,000 and $53,000 respectively.**

**Would a person with Salary $1 be considered an Outlier?**

**A. Yes**

**B. No**

**C. More information is required**

**D. None of these.**

**Solution: C**

**16. How does number of observations influence overfitting? Choose the correct answer(s).**

**Note: Rest all parameters are same**

1. **In case of fewer observations, it is easy to overfit the data.**
2. **In case of fewer observations, it is hard to overfit the data.**
3. **In case of more observations, it is easy to overfit the data.**
4. **In case of more observations, it is hard to overfit the data.**

**A. 1 and 4**

**B. 2 and 3**

**C. 1 and 3**

**D. None of theses**

**Solution: A**

**17. If B­­­yx>1 then Bxy**

**(a) Greater than**

**(b) smaller than 1**

**(c) Both values are same**

**(d)equal to zero**

**Correct Answer (b)**

**18. The range of a partial correlation coefficient is:**

**(a) 0 to 1**

**(b) -1 to +1**

**(c) -1 to 0**

**(d) None of these**

**Correct Answer (b)**

**19. If the two lines of regression are perpendicular to each other , the correlation coefficient r is**

**(a) +1**

**(b)-1**

**(c) Zero**

**(d) Nothing to be said**

**Correct Answer (c)**

**20. Which of the following option is true regarding “Regression” and “Correlation” ?**

**Note: y is dependent variable and x is independent variable.**

**A. The relationship is symmetric between x and y in both.**

**B. The relationship is not symmetric between x and y in both.**

**C. The relationship is not symmetric between x and y in case of correlation but in case of regression it is symmetric.**

**D. The relationship is symmetric between x and y in case of correlation but in case of regression it is not symmetric.**

**Solution: D**

**MCQ for 16.08.2020**

**1) True-False: Is Logistic regression a supervised machine learning algorithm?**

**A) TRUE  
B) FALSE**

**Solution: A**

**2) True-False: Is it possible to design a logistic regression algorithm using a Neural Network Algorithm?**

**A) TRUE  
B) FALSE**

**Solution: A**

**3) Which of the following methods do we use to best fit the data in Logistic Regression?**

**A) Least Square Error  
B) Maximum Likelihood  
C) Jaccard distance  
D) Both A and B**

**Solution: B**

**4) Which of the following evaluation metrics can not be applied in case of logistic regression output to compare with target?**

**A) AUC-ROC  
B) Accuracy  
C) Logloss  
D) Mean-Squared-Error**

**Solution: D**

**5 One of the very good methods to analyze the performance of Logistic Regression is AIC, which is similar to R-Squared in Linear Regression. Which of the following is true about AIC?**

**A) We prefer a model with minimum AIC value  
B) We prefer a model with maximum AIC value  
C) Both but depend on the situation  
D) None of these**

**Solution: A**

**Context: 6-7**

**Consider a following model for logistic regression: P (y =1|x, w)= g(w0 + w1x)  
where g(z) is the logistic function.**

**In the above equation the P (y =1|x; w) , viewed as a function of x, that we can get by changing the parameters w.**

**6) What would be the range of p in such case?**

**A) (0, inf)  
B) (-inf, 0 )  
C) (0, 1)  
D) (-inf, inf)**

**Solution: C**

**7) In above question what do you think which function would make p between (0,1)?  
  
A) logistic function  
B) Log likelihood function  
C) Mixture of both  
D) None of them**

**Solution: A**

**8) Suppose you have been given a fair coin and you want to find out the odds of getting heads. Which of the following option is true for such a case?**

**A) odds will be 0  
B) odds will be 0.5  
C) odds will be 1  
D) None of these  
Solution :C**

**9) The logit function(given as l(x)) is the log of odds function. What could be the range of logit function in the domain x=[0,1]?**

**A) (– ∞ , ∞)  
B) (0,1)  
C) (0, ∞)  
D) (- ∞, 0)**

**Solution: A**

**10) Which of the following option is true?**

**A) Linear Regression errors values has to be normally distributed but in case of Logistic Regression it is not the case  
B) Logistic Regression errors values has to be normally distributed but in case of Linear Regression it is not the case  
C) Both Linear Regression and Logistic Regression error values have to be normally distributed  
D) Both Linear Regression and Logistic Regression error values have not to be normally distributed**

**Solution:A**

**11. What would you use Box’s test for?**

**Answer choices**

1. **To test for multivariate normality.**
2. **To test for independence of residuals**
3. **To test for homogeneity of variance**
4. **To test the assumption of homogeneity of covariance matrices.**

**Correct Answer : D**

**12. An examination of differences across groups lies at the heart of the basic concept of \_\_\_\_\_.**

1. **regression analysis**
2. **discriminant analysis**
3. **conjoint analysis**
4. **factor analysis**

**Correct Answer (b)**

**13. The unstandardized coefficients are multiplied by the values of the variables. These products are summed and added to the constant term to obtain the \_\_\_\_\_.**

1. **discriminant function coefficients**
2. **classification matrix**
3. **discriminant scores**
4. **canonical correlation**

**Correct Answer : ©**

**14. If you are using the leave-out option of SPSS, you are at the \_\_\_\_\_ step of discriminant analysis.**

1. **estimate the discriminant coefficients**
2. **determine the significance of the discriminant function**
3. **interpret the results**
4. **assess validity of discriminant analysis**

**Correct answer : (d)**

**15. Discriminant analysis can be conducted via all of the following except:**

1. **SPSS**
2. **EXCEL**
3. **MINITAB**
4. **SAS**

**Correct Answer : (b)**

**16. A \_\_\_\_\_ is a tool for assessing discriminant analysis results that plots the group membership of each case on a graph.**

1. **characteristic profile**
2. **scattergram**
3. **territorial map**
4. **pie chart**

**Correct Answer: ©**

**17. Eigen value in Discriminant Analysis [E=Error, R –Regression]**

1. **ESS/TSS**
2. **RSS/TSS**
3. **RSS/ESS**
4. **ESS/RSS**

**Correct Answer is ©**

**18. Wilk’s Lambda in discriminant analysis [B=Between group W =within group]**

**(a) BSS/TSS**

**(b) WSS/TSS**

**© BSS/WSS**

**(d)none of these**

**Correct Answer : (b)**

**19. Statistical significance of the discriminant function**

**(a) is judged on the basis of structure matrix**

**(b) is judged on the basis of canonical discriminate function**

**© is judged on the basis of wilk’s lambda**

**(d)cannot be judged.**

**Correct Answer: (c)**

**20 The dependent variable in discriminant analysis is**

1. **Measured in ratio scale (b) measured in interval scale**

**© measured in ordinal scale (d) measured in nominal scale.**

**Correct Answer: (d)**

**MCQ for 22.08.2020**

1. In a 7x7 correlation matrix what is the sum of eigen values?
2. 7
3. b) 8
4. c) 3
5. d) none of these

Answer :(a)

2. if all eigen values are positive then the matrix is said to be

a) positive semi definite

b) positive definite

c) negative definite

d) Indefinite

Answer: (b)

3) In factor analysis with 19 variables what is the degree’s of freedom ?

(a) 172 (b) 171 (c) 190 (d) 210

Answer: (b)

**4. If 2 = 0.29 & =0.124 find the value of KMO ?**

**a)0.5**

**b)0.6**

**c) 0.65**

d) 0.7

Answer: (d)

**5. Varimax rotation should be used when:**

Answer choices

1. You believe that the underlying factors will be correlated.
2. You believe that the underlying factors are non-orthogonal.
3. You believe that the underlying factors are independent.
4. Kaiser’s criterion is met.

#### Answer: ©

**6. A Cronbach’s alpha value of .85 for a questionnaire means that:**

Answer choices

1. The questionnaire is valid.
2. The questionnaire has good reliability.
3. The questionnaire has too few items.
4. The questionnaire would produce different scores if used on the same people at two different points in time.

#### Answer: (b)

**7. Kaiser criterion for retaining factors is:**

Answer choices

1. Retain any factor with an eigenvalue greater than 1.
2. Retain any factor with an eigenvalue greater than 0.3.
3. Retain factors before the point of inflexion on a scree plot.
4. Retain factors with communalities greater than 0.7.

#### Answer: (a)

**8. Which of these is a form of oblique rotation?**

1. Equamax
2. Quartimax
3. Varimax
4. Promax

#### Answer: D

**9. What is a latent variable?**

1. It is a variable that cannot be measured directly.
2. It is another name for a factor.
3. Latent variables represent clusters of variables that correlate highly with each other.
4. All of these are correct.

#### Answer: D

  10. Which of the following is true?

(a) Shorter scales tend to have greater internal reliability but lower reliability over time.

(b) Shorter scales tend to have lower internal reliability but higher reliability over time.

© Longer scales tend to have higher internal reliability and lower reliability over time.

(d) Longer scales tend to have lower internal reliability and lower reliability over time.

Answer : (a)

11. Alternate-form reliability is also known as…

(a) Split-half reliability

(b) Convergent reliability

© Test-retest reliability.

(d) Parallel forms

Answer : (d)

12. Using a properly calibrated set of bathroom scales to measure mood would be:

(a) Neither reliable nor valid.

(b) Reliable but not valid

© Reliable and valid.

(d) Valid but not reliable.

Answer: (b)

13. Why might test-retest reliability become less important?

(a) If alternate-form reliability cannot be calculated.

(b) If participants become familiar with a test and all perform better on the second occasion.

© If the concept being measured is not expected to be stable over time (e.g. mood).

(d) If convergent validity was not high.

Answer: ©

14. If care has been taken to ensure that a measure is measuring all aspects of something fully, it has high…

(a) Criterion validity

(b) Internal validity

© Face validity

(d) Content validity

Answer: (d)

15. If a measure correlates highly with other established measures of the same thing, it is said to have high….

|  |  |
| --- | --- |
| (a) Construct validity. | |
| |  | | --- | | (b) Face validity. | |  | | |
| © Content Valdity |
| (d)Criterion validity.  Answer : (d) |

16.The total of all eigen values will equal:

1. 7
2. 50
3. 100
4. Impossible to tell without further information
5. None of these

Answer: (d)

17. **To determine which variables relate to which factors, a researcher would use:**

**(a) Factor loadings**

**(b) Communalities**

**© Eigen values**

**(d) Beta coefficients**

Answer :(a)

18.If a researcher wants to determine the amount of variance in the original variables that is associated with a factor, s/he would use:

(a) Factor Loading

(b) Eigen Value

© Beta Coefficient

(d) Communalities

Answer: (b)

19. In SPSS, orthogonal rotation in factor analysis is called:

|  |
| --- |
| (a) Oblimin |
| (b) Oblimax |
| © Oblique |
| (d) Varimax |
| (e)None of the above |

Answer: (d)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 20. Within group and between group correlations are called:  (a) Inter and Intra  (b) Intra and Inter  © Intra and Intra  (d) None of the above  Answer: (b)  MCQ for 23.08.2020 | | | | | |
|  | |  | |  | |
| 1. The number of brands or stimuli selected for use in MDS analysis and the specific brands included determines the nature of the resulting dimensions and configurations. 2. TRUE (b) FALSE   Answer : (a)   1. Suppose a researcher is interested in obtaining consumer perceptions of automobiles, the choice of the number and specific brands or stimuli to be included in the MDS analysis should be based on the statement of the marketing research problem and theory only. The researcher should not bias the research by also using his judgment to determine what should be included in the analysis. 2. TRUE (b) FALSE   Answer : (b)   1. In direct approaches to gathering perception data, respondents are asked to make similarity judgments on pairs of various brands or stimuli. 2. TRUE (b) FALSE   Answer : (a)   1. Derived approaches to collecting perception data require respondents to rate the brands or stimuli on the identified attributes using semantic differential scales or Likert scales. 2. TRUE (b) FALSE   Answer : (a)   1. The fit of an MDS solution is commonly assessed by the stress measure. Stress is a goodness-of-fit measure; higher values of stress indicate better fits. 2. TRUE (b) FALSE   Answer: (b)   1. Non -metric MDS procedures assume that the input data are ordinal, and they result in ordinal output. 2. TRUE (b) FALSE   Answer: (b)   1. The metric MDS procedures assume that input data are metric and the output is also metric. 2. TRUE (b) FALSE   Answer : (a)   1. The metric and non-metric MDS methods often produce vastly different results. 2. TRUE (b) FALSE   Answer : (b)   1. Spatial maps are computed in such a way that the fit increases as the number of dimensions decreases. 2. TRUE (b) FALSE   Answer : (b)   1. The objective in MDS is to obtain a spatial map that best fits the input data in the smallest number of dimensions. 2. TRUE (b) FALSE   Answer: (a)   1. Gaps in the spatial map may indicate potential opportunities for introducing new products. 2. TRUE (b) FALSE   Answer (a)   1. Each dimension determined in MDS represents only one attribute. 2. TRUE (b) FALSE   Answer: (b)   1. MDS solutions are subject to substantial random variability. 2. TRUE (b) FALSE   Answer: (a)   1. Stress values indicate the proportion of variance of the optimally scaled data that is not accounted for by the MDS model. 2. TRUE (b) FALSE   Answer: (b)   1. An assumption of MDS is that the similarity of stimulus A to B is the same as the similarity of stimulus B to A. 2. TRUE (b) FALSE   Answer: (a)   1. The text reports on a research project which used MDS to plot the perceived ethics of marketing research firms using a broad-based moral equity dimension (factor 1)and a relativistic dimension (factor 2). Based on the research, internal marketing research departments are perceived to be the most ethical on both dimensions. 2. TRUE (b) FALSE   Answer: (b)   1. \_\_\_\_\_\_\_\_ is a class of procedures for representing perceptions and preferences of respondents spatially by means of a visual display.   a)Conjoint analysis  b)Regression analysis  c)Hybrid conjoint analysis  d)Multidimensional scaling (MDS)  Answer : (d)  18. MDS could be used for all of the marketing applications below except \_\_\_\_\_\_\_\_  A) Market segmentation - position brands and consumers in the same space and thus identify groups of consumers with relatively homogeneous perceptions.  B) Assessing advertising effectiveness-spatial maps can be used to determine whether advertising has been successful in achieving the desired brand positioning.  C)Consumer intention - how do consumer's intentions to buy the brand vary with different price levels?  D)Channel decisions - judgments on compatibility of brands with different retail outlets could lead to spatial maps useful for making channel decisions.  Answer : (c)  19. \_\_\_\_\_\_\_\_ requires that the researcher specify the purpose for which the MDS results would be used and select the brands or other stimuli to be included in the analysis.  a)Formulating the problem  b)Obtaining input data  c)Selecting an MDS procedure  d)Deciding on the number of dimensions  Answer: (d)  20. In MDS, at minimum, \_\_\_\_\_\_\_\_ brands or stimuli should be included so as to obtain a well-defined spatial map.Including more than \_\_\_\_\_\_\_\_ brands is likely to be cumbersome and may result in respondent fatigue.  a)6; 20  b)6; 25  c)8; 20  d)8; 25  Answer : (d) | | | |  | |
|  | | | | | |
|  | |  | |  | |
|  | | | |  | |
|  | | | | | |
|  | |  | |  | |
|  | | | |  | |
|  | | | | | |
|  | |  | |  | |
|  | | | |  | |
|  | | | | | |
|  | |  | |  | |
|  | | | |  | |
|  | | | | | |
|  | |  | |
|  | |  | |
|  | |  | |
|  | |  | |
|  | |  | |

Bottom of Form