**Machin learning QUIZ -Summer Training and Internship**

1. **Consider the following dataset. Which of these are variables are continuous?**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Player** | **Nationality** | **Career Earning** | **Average aces per match** | **Hands used for backhand** |
| **Andy Murray** | **Great Britain** | **$59,736,549** | **6.5** | **2** |
| **Novak Djokovic** | **Serbia** | **$108,218,048** | **3.6** | **2** |
| **Rafael Nadal** | **Spain** | **$82,313,302** | **2.5** | **2** |
| **Roger Federer** | **Switzerland** | **$103,990,195** | **10.1** | **1** |
| **Stan Wawrinka** | **Switzerland** | **$29,184,010** | **7.3** | **1** |

1. **Player B) Career Earning**

**C) Average aces per Match D) Hands used for backhand**

1. **Consider dataset from the previous question as a population. Find the Standard deviation of “Average aces per match”.**
2. **7.352 B) 2.711 C) 6 D) 3**
3. **Which of the following definitions describes the concept of statistical distribution?**
4. **Distribution is a mathematical function which is linked to the data**
5. **Distribution is a chart which visualizes the data**
6. **Distribution is a table which contains the data**
7. **All of these definitions describe the concept of statistical distribution**
8. **Which of these statements about the normal distribution is FALSE?**
9. **The more observations you have in your sample the better it will approximate to the normal distribution.**
10. **Usually 95.4% of the observations fall within 2 standard deviations from the mean**
11. **The mean is always zero**
12. **All are true**
13. **If you know the values of N, µ, and σ, then you know**
14. **The parameters of a sample**
15. **The statistics of a sample**
16. **The parameters of a population**
17. **The statistics of a population**
18. **Which of the following definitions describes Central Limit Theorem?**
19. **Provided enough samples are taken the sampling distribution of the sample mean will be normally distributed regardless of the population**
20. **The average of the results obtained from a large number of trials should be close to the expected value and will tend to be closer as more trials are performed**
21. **Both definitions describe Central Limit Theorem**
22. **Neither definition describe Central Limit Theorem**
23. **the Central Limit Theorem, as the number of observations in the samples grows, the sampling distribution of the sample mean becomes**
24. **thinner**
25. **broader**
26. **positively skewed**
27. **negatively skewed**
28. **taken a random sample with 121 observations from a population with 625 observations, mean=34 and standard deviation=4.1, what would be the standard deviation of that sample?**
29. **4.1**
30. **0.703**
31. **0.164**
32. **0.373**
33. **In statistical language, which of the following is a synonym of Z-Score?**
34. **Normal score**
35. **Standard score**
36. **Both are**
37. **Neither is**
38. **original distribution with mean=192 and standard deviation=8. Next, you have taken a random sample from it with number of observations in the sample=100. Find the  Z-Score of Xcrit.=193.6.**
39. **2.12**
40. **0.9772**
41. **1.6**
42. **2**
43. **In statistical hypothesis testing, the statistical significance α is typically set to:**
44. **50%**
45. **5%**
46. **95%**
47. **100%**
48. **The statistical significance α is the probability of:**
49. **Rejecting H0 when it is false**
50. **Rejecting H0 when it is true**
51. **Accepting H0 when it is false**
52. **Accepting H0 when it is true**
53. **In statistical hypothesis testing, we can reject H0 when**
54. **p-value>α**
55. **p-value<α**
56. **p-value>50%**
57. **Neither is true**
58. **If you have randomly picked a sample from the rejection region, you can:**
59. **Reject H1**
60. **Reject H0**
61. **Accept H0**
62. **Neither is true**
63. **Which of the following is NOT among the assumptions of the Z-Test?**
64. **Sample is selected at random**
65. **Observations are independent**
66. **The population`s standard deviation is known or the sample contains at least 30 observations**
67. **Those three are the assumptions of the Z-Test**
68. **In which of the following scenarios can you not apply proportion testing?**
69. **test the Hypothesis (with 97% significance level) that only 19% of the cars in France are white. You surveyed 100 random cars in France and found that 20 of them are white**
70. **test the Hypothesis (with 95% significance level) that only 5% of football clubs in England have green kits. You surveyed 60 random English football clubs and found that 4 of them have green kits**
71. **to test the Hypothesis (with 90% significance level) that only 30% of the office workers in Asia like to wear blue shirts to work. You surveyed 505 random office workers in Asia and found that 123 of them like to wear blue shirts to work**
72. **Proportion testing is possible in all of the three scenarios**
73. **Which among visualization plot gives us median, quartiles from its plot?**
74. **Histogram B) Scatter Plot C) Box and Whisker plots D) Correlation Plot**
75. **Which python library is used specially for Machine Learning?**
76. **Numpy B) Pandas C) Matplotlib D) Scikit-learn**

1. **What will be the output of the given code?**

**list= [‘a’,’e’,’i’,’o’,’u’]**

**print list [8:]**

1. **[‘a’,’e’,’i’,’o’,’u’]**
2. **[’o’,’u’]**
3. **[‘a’,’e’]**
4. **empty list**
5. **what will be you do when you have missing value in categorical variable.**
6. **Mean**
7. **Median**
8. **Most\_ Frequent(MODE)**
9. **None**
10. **what is the syntax for the data importing for CSV file using pandas?**

**A) pd.Read\_CSV B) pd.read\_csv C) pd.Read\_Table D)pd.read\_table**

1. **which among gives us strength among the two variables?**

**A) Co-Variance B) Correlation C) Both D) None**

**23.using Data Frame indexing of an integer indexing among given below?**

**A) .loc[ ] B) .iloc[ ] C) .ilx[ ] D) lx[ ]**

**24. which among is not Tree based learning?**

**A) CART B) C4.5 C) Random Forest D) SVM**

**25. When do you go for a Label Encoder?**

**A) when you are having missing value**

**B) when you have Numeric variable**

**C) when you have Categorical Variable in dataset**

**D) None of the above**

**26. when you have outlier data in dataset what to do with it ?**

**A) Remove it from dataset**

**B) Continue working with Outlier Data**

**C) None**

**27.from the stats model how you choose the variable significance?**

**A) p<=0.5**

**B) p<=0.05**

**C) p>=0.05**

**D) p>=0.5**

#### 28. Null hypothesis is a general statement that there is a relationship between two measured phenomena

#### A) True

#### B) False

#### 29. If the data observed in the sample test is highly unlikely to occur, then the null hypothesis is

#### Rejected

#### Not rejected

#### None

#### 

#### 30. ****A Pearson correlation between two variables is zero but, still their values can still be related to each other****

#### ****A) True****

#### ****B) False****

**31. Which of the following is/are one of the important step(s) to pre-process the text in NLP based projects?**

1. **Stemming**
2. **Stop word removal**
3. **Object Standardization**
4. **1 and 2**
5. **1 and 3**
6. **2 and 3**
7. **1,2 and 3**

**32. Which of the following options can be used to get global minima in k-Means Algorithm?**

1. **Try to run algorithm for different centroid initialization**
2. **Adjust number of iterations**
3. **Find out the optimal number of clusters**
4. **2 and 3**
5. **1 and 3**
6. **1 and 2**
7. **All of above**

**33. The difference between deep learning and machine learning algorithms is that there is no need of feature engineering in machine learning algorithms, whereas, it is recommended to do feature engineering first and then apply deep learning.**

**A) TRUE**

**B) FALSE**

**34. Random Forest deal with?**

**A) Regression Task B) Classification Task**

**C) Both A and B D) None**

**35. Lemmatization is the process of reducing a group of words into their lemma or dictionary form.**

**A) True**

**B) False**

**36. Which of the following is/are true about bagging trees?**

1. **In bagging trees, individual trees are independent of each other**
2. **Bagging is the method for improving the performance by aggregating the results of weak learners**

**A) 1  
B) 2  
C) 1 and 2  
D) None of these**

**37. Deep learning itself does feature engineering whereas machine learning requires manual feature engineering.**

**A) True B) False**

**38. In Neural Network the weights of the neurons are auto updated in deep networking layers**

**A) True B) False**

**39. Which of the following is a representation learning algorithm?**

**A) Neural network B) Random Forest**

**C) k-Nearest neighbor D) None of the above**