16/05/2023

**Python Worksheet answers**

Q1) Which of the following operators is used to calculate remainder in a division?

A1) % (C)

Q2) In python 2//3 is equal to?

A2) 0.666 (A)

Q3) In python, 6<<2 is equal to?

A3) 24 (C)

Q4) In python, 6&2 will give which of the following as output?

A4) 2 (A)

Q5) In python, 6|2 will give which of the following as output?

A5) 6 (D)

Q6) What does the finally keyword denotes in python?

A6) It is used to mark the end of the code (A)

Q7) What does raise keyword is used for in python?

A7) It is used to raise an exception (A)

Q8) Which of the following is a common use case of yield keyword in python?

A8) in defining a generator (C)

Q9) Which of the following are the valid variable names?

A9) 1abc (B) and abc2 (C)

Q10) Which of the following are the keywords in python?

A10) all to the above (D)

A11, 12, 13,14, 15 answered in Python workbook.

**Statistics Worksheet answers**

Q1) Bernoulli random variables take (only) the values 1 and 0.

A1) True (A)

Q2) Which of the following theorem states that the distribution of averages of iid variables, properly normalized, becomes that of a standard normal as the sample size increases?

A2) Central Limit Theorem (A)

Q3) Which of the following is incorrect with respect to use of Poisson distribution?

A3) Modeling bounded count data (B)

Q4) Point out the correct statement.

a) The exponent of a normally distributed random variables follows what is called the log- normal distribution.

b) Sums of normally distributed random variables are again normally distributed even if the variables are dependent.

c) The square of a standard normal random variable follows what is called chi-squared distribution.

**D) All of the mentioned. A4) D**

Q5) \_\_\_\_\_\_ random variables are used to model rates.

A5) Poisson (C)

Q6) Usually replacing the standard error by its estimated value does change the CLT.

A6) False (B)

Q7) Which of the following testing is concerned with making decisions using data.

A7) Hypothesis (B)

Q8) Normalized data are centered at\_\_\_\_\_\_and have units equal to standard deviations of the original data.

A8) 0 (A)

Q9) Which of the following statement is incorrect with respect to outliers?

A9) Outliers cannot conform to the regression relationship. (C)

Q10) What do you understand by the term Normal Distribution?

A10) A normal distribution is a type of continuous probability distribution in which most data points cluster towards the middle of the range. Most of the data points lies between -3 and +3 in the graphical presentation. In the normal distribution the chances of the outliers are very less.

Q11) How do you handle missing data? What imputation techniques do you recommend?

A11) The missing data can checked through isnull.sum(). Missing data can be handled through various techniques. If the missing data are few the same rows can be dropped for the data. But if the missing data is in large number and the attributes are important the data the same can not be dropped. The missing data can be filled with simple imputer. The integer data can be filled with strategy of mean and mode and the string data can be filled through strategy of most frequent.

Q12) What is A/B testing?

A12) A/B testing data science is a scientific method of comparing two versions of a website, app or advertisement to determine which one performs better. It's also known as split testing.

Q13) Is mean imputation of missing data acceptable practice?

A13) Mean imputation is not an often-acceptable practice, it is considered one of the most terrible practices in Data Science. One of the reasons it is not acceptable practice is that it ignores the feature correlation which invalidates most hypothesis tests.

Q14) What is linear regression in statistics?

A14) Simple linear regression is a regression model that estimates the relationship between one independent variable and one dependent variable using the straight line. The variable which is predicted is dependent variable and the variable which is used to predict the dependent variable is known as independent variable.

Q15) What are the various branches of statistics?

A15) There are basically 2 types or branches of statistics and they are

1 – Descriptive Statistics

2 – Inferential Statistics

Descriptive Statistics – There are basically 2 types

A – Central Tendency – has 3 methods.

1. Mean
2. Median
3. Mode

B – Despertion Data – There are basically 4 methods.

1. Lang
2. Percentile
3. Variance
4. Standard Deviation
5. Skew

Inferential Statistics – It does not have direct values

There are basic methods are as follows

1. Hypothesis Testing
2. T-test
3. Z-test
4. Co-relation test
5. Chi-square