Statistics Worksheet-1

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

A. True

B. False

Ans : True

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

A. Central Limit Theorem
B. Central Mean Theorem
C. Centroid Limit Theorem
D. All of the mentioned

Ans : A

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

A. Modeling event/time data
B. Modeling bounded count data
C. Modeling contingency tables
D. All of the mentioned

Ans : B

Q4) Point out the correct statement.

A. The exponent of a normally distributed random variables follows what is called the log- normaldistribution
B. Sums of normally distributed random variables are again normally distributed even if the variablesare dependent
C. The square of a standard normal random variable follows what is called chi-squareddistribution
D. All of the mentioned

Ans : D

Q5) _____random variables are used to model rates.

A. Empirical
B. Binomial

C. Poisson
D. All of the mentioned

Ans : C

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

A. True
B. False

Ans : B

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

A. Probability
B. Hypothesis
C. Causal
D. None of the mentioned

Ans : B

Q8) Normalized data are centered at_____and have units equal to standard deviations of theoriginal data.

A. 0
B. 5
C. 1
D. 10

Ans : A

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

A. Outliers can have varying degrees of influence
B. Outliers can be the result of spurious or real processes
C. Outliers cannot conform to the regression relationship
D. None of the mentioned

Ans : C

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

Ans : Normal distribution is a bell shaped curve of various random variables having symmetrical data around mean . Mean is centered at 0 and have standard deviation +1 or -1.

Normal distribution is a continuous probability distribution. It is also called Gaussian curve as discovered by Carl Gauss.

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

Ans : One way of handling missing values is the deletion of the rows or columns having null values. If any columns/rows have more than half of the values as null then you can drop the entire column/row.

Various imputation techniques :

- Mean imputation
- Substitution
- Hot deck imputation
- Cold deck imputation
- Regression imputation

Q12) What is A/B testing?

Ans : A/B testing refers to the experiment where two or more variations are compared against each other to determine which one performs better for a given goal.

Q13) Is mean imputation of missing data acceptable practice?
Ans : No, it is not as good . Other techniques have advantage over it .
Mean imputation is a popular imputation technique but there are serious problems with mean imputation. The variance of a mean-imputed variable is always biased downward from the variance of the un-imputed variable. This bias affects standard errors, confidence intervals, and other inferential statistics.

Q14) What is linear regression in statistics?
Ans : It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables.
It is a best fit line where maximum points are covered . The linear regression model provides a sloped straight line representing the relationship between the dependent variables and independent variables.

Q15) What are the various branches of statistics?
Ans : Statistics has mainly 2 branches

1. Desriptive Statistics
2. Inferential Statistics

Further desrciptive statistics divided into 2 parts :
(i) Central Tendency : Inside it mean , median and mode calculated.

(ii) Dispersion of data : Inside it Range, Variance, Standard Deviation, Skew, Percentile etc. are calculated.

Under Inferential Statistics various hypothesis testing are done:

Tests are T Test, Z Test, Regression Test, Chi-square Test, Anova Test etc.