WORKSHEET

STATISTICS WORKSHEET-1

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

1. Bernoulli random variables take (only) the values 1 and 0.
a) True
b) False
Ans:a

2. 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?
a) Central Limit Theorem
b) Central Mean Theorem
c) Centroid Limit Theorem
d) All of the mentioned

Ans:b

4. 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
Ans:c

5. _____ random variables are used to model rates.
a) Empirical
b) Binomial
c) Poisson
d) All of the mentioned
Ans:c
6. 10. Usually replacing the standard error by its estimated value does change the CLT.
a) True
b) False
Ans:b
7. 1. 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
8. 4. Normalized data are centered at_____and have units equal to standard deviations of the original data.
a) 0
b) 5
c) 1
d) 10

Ans:c

9. 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

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

10. What do you understand by the term Normal Distribution?

Normal distribution is also known as Gaussian Distribution, is a probability distribution that is symmetric about the mean,showing that data near the mean are more frequent in occurrence than the data far from the mean. In graph form , normal distribution will appear as bell curve.

11. How do you handle missing data? What imputation techniques do you recommend?

Mean or median imputation. When data is missing at random,we can use list wise or pair wise deletion of the missing observations.

Multivariate imputation by chained equation.

12. What is A/B testing?

A/B testing is also known as split testing is a process of showing two variants of the same web page to different segments of website visitors at the same time and comparing which variant drives more conversions.

13. Is mean imputation of missing data acceptable practice?

It is a non standard ,but fairly flexible imputation algorithm.it uses random forest at its core to predict the missing data. It can be applied to both continious and categorical variables which makes it advantageous over the other imputation.

14. What is linear regression in statistics?

Linear regression is a linear approach to modelling the relationship between a scalar response and one or more explanatory variables.one explanatory variable is called as simple linear regression and more than one variable is called as multiple linear regression.

15. What are the various branches of statistics?

Descriptive statistics and inferential statistics. Both are employed in scientific analysis of data and both are equally important for the student of statistics.