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

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

3. Which of the following is incorrect with respect to use of Poisson distribution?
b) Modeling bounded count data

4. Point out the correct statement.
d) All of the mentioned

5. _____ random variables are used to model rates.
c) Poisson

6. 10. Usually replacing the standard error by its estimated value does change the CLT.
b) False

7. 1. Which of the following testing is concerned with making decisions using data?
b) Hypothesis

8. 4. Normalized data are centered at_____and have units equal to standard deviations of the original data.
a) 0

9. Which of the following statement is incorrect with respect to outliers?
c) Outliers cannot conform to the regression relationship

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

10. What do you understand by the term Normal Distribution?
Ans. A normal distribution is the proper term for a probability bell curve.
In a normal distribution the mean is zero and the standard deviation is 1. It has zero skew and a kurtosis of 3.
Normal distributions are symmetrical, but not all symmetrical distributions are normal.
In reality, most pricing distributions are not perfectly normal.

11. How do you handle missing data? What imputation techniques do you recommend?
Ans. The concept of missing data is implied in the name: it's data that is not captured for a variable for the observation in question.
Missing data reduces the statistical power of the analysis, which can distort the validity of the results.
Imputation techniques are:
1.Mean, Median and Mode
2.Time-Series Specific Methods
3.Last Observation Carried Forward (LOCF) & Next Observation Carried Backward (NOCB)
4.Linear Interpolation
5.Seasonal Adjustment with Linear Interpolation

12. What is A/B testing?
Ans.A/B testing is a popular way to test your products. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.
A/B testing works best when testing incremental changes, such as UX changes, new features, ranking, and page load times.
Here you may compare pre and post-modification results to decide whether the changes are working as desired or not.
A/B testing doesn't work well when testing major changes, like new products, new branding, or completely new user experiences.
In these cases, there may be effects that drive higher than normal engagement or emotional responses that may cause users to behave in a different manner.

13. Is mean imputation of missing data acceptable practice?
Ans.When data is missing, it may make sense to delete data. However, that may not be the most effective option.
For example, if too much information is discarded, it may not be possible to complete a reliable analysis.
Or there may be insufficient data to generate a reliable prediction for observations that have missing data.
Instead of deletion, data scientists have multiple solutions to impute the value of missing data.
Depending why the data are missing, imputation methods can deliver reasonably reliable results.

14. What is linear regression in statistics?
Ans.Linear regression is a basic and commonly used type of predictive analysis.
The overall idea of regression is to examine two things:
(1) does a set of predictor variables do a good job in predicting an outcome (dependent) variable
(2) Which variables in particular are significant predictors of the outcome variable, and in what way do

they–indicated by the magnitude
and sign of the beta estimates–impact the outcome variable
These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables.
The simplest form of the regression equation with one dependent and one independent variable is defined by the formula y = c + b*x, where y = estimated dependent variable score, c = constant, b = regression coefficient, and x = score on the independent variable.

15. What are the various branches of statistics?
Ans. Two branches, descriptive statistics and inferential statistics, comprise the field of statistics.

1.Descriptive Statistics
The branch of statistics that focuses on collecting, summarizing, and presenting a set of data.

2.Inferential Statistics
The branch of statistics that analyzes sample data to draw conclusions about a population