

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
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
3. 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
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
5. _____ random variables are used to model rates.
 - a) Empirical(✓)
 - b) Binomial
 - c) Poisson
 - d) All of the mentioned
6. 10. Usually replacing the standard error by its estimated value does change the CLT.
 - a) True
 - b) False(✓)
7. Which of the following testing is concerned with making decisions using data?
 - a) Probability
 - b) Hypothesis(✓)
 - c) Causal
 - d) None of the mentioned
8. Normalized data are centered at _____ and have units equal to standard deviations of the original data.
 - a) 0(✓)
 - b) 5
 - c) 1
 - d) 10
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

Q10 and Q15 are subjective answer type questions, Answer them in your own words briefly.

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

A normal distribution is an arrangement of a data set in which most values cluster in the middle of the range and the rest taper off symmetrically toward either extreme.

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

Missing data can be dealt with in a variety of ways. I believe the most common reaction is to ignore it. Choosing to make no decision, on the other hand, indicates that your statistical programme will make the decision for you.

Your application will remove things in a listwise sequence most of the time. Depending on why and how much data is gone, listwise deletion may or may not be a good idea.

Another common strategy among those who pay attention is imputation. Imputation is the process of substituting an estimate for missing values and analysing the entire data set as if the imputed values were the true observed values.

And how would you choose that estimate? The following are some of the most prevalent methods:

Mean imputation Calculate the mean of the observed values for that variable for all non-missing people. It has the advantage of maintaining the same mean and sample size, but it also has a slew of drawbacks. Almost all of the methods described below are superior to mean imputation.

12. What is A/B testing?

A/B testing (also known as bucket testing or split-run testing) is a **user experience research methodology**. A/B tests consist of a randomized experiment that usually involves two variants (A and B), although the concept can be also extended to multiple variants of the same variable.

13. Is mean imputation of missing data acceptable practice?

Mean imputation is typically considered terrible practice since it ignores feature correlation

14. What is linear regression in statistics?

This form of analysis estimates the coefficients of the linear equation, involving one or more independent variables that best predict the value of the dependent variable. Linear regression fits a straight line or surface that minimizes the discrepancies between predicted and actual output values.

15. What are the various branches of statistics?

There are three real branches of statistics: data collection, descriptive statistics and inferential statistics

