

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-A

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 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-D

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-A

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

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

ANS- A normal distribution or Gaussian distribution refers to a probability distribution where the values of a random variable are distributed symmetrically. These values are equally distributed on the left and the right side of the central tendency. Thus, a bell-shaped curve is formed.

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

ANS- 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 statistical programme will make the decision .

The 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. The imputation techniques that I recommend are mean imputation, substitution, Hot deck imputation, cold deck imputation etc.

12. What is A/B testing?

ANS- A/B testing is a user experience research methodology. A/B tests consist of a randomized experiment with two variants, A and B. It includes application of statistical hypothesis testing or "two-sample hypothesis testing" as used in the field of statistics. A/B testing is a way to compare two versions of a single variable, typically by testing a subject's response to variant A against variant B, and determining which of the two variants is more effective

13. Is mean imputation of missing data acceptable practice?

ANS- Yes, mean imputation of missing data acceptable practice imputing the mean preserves the mean of the observed data. So if the data are missing completely at random, the estimate of the mean remains unbiased. Since most research studies are interested in the relationship among variables, mean imputation is not a good solution.

14. What is linear regression in statistics?

Ans- In statistics, linear regression is a linear approach for modelling the relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables). The case of one explanatory variable is called simple linear regression; for more than one, the process is called multiple linear regression.

15. What are the various branches of statistics?

ANS- There are various branches of statistics they are as follows

Econometric

Econometric is one of the branches of statistics where it takes parts to resolve economic models and problem .

Actuarial

Actuarial is another applied statistical branch that focuses on studying and analyzing risk in finance and insurance.

Psychometrics

Psychometrics is another interesting branch of statistics. This one focus on studying measurement technique and analyzing in the education world and psychology. This thing included attitude, personality, emotion, and many others.

Physics Statistics

Physics statistics is one of the statistical branches that focuses on solving physics science. Usually, statistics take part in measurement and calculation with particle.

Population Statistics

The population of statistics is one of the most useful branches that study about many things related to society. It has many connections with another aspect of our life, such as health, education, migration, and so on.

Official Statistics

Not just people, the government also need statistics as a tool to make the right policy. In official statistics, we learn how to measure some strategic indicator that has a large impact on society. There are many indicators that the government's need to fulfill the policy based on evidence or data.

Biostatistics

Biology statistics or more popular with names Biostatistics is another branch of statistics that used to resolve many statistical problems in biology. Actually, it's not just for biology. Biostatistics is also applied to many problems in the medical world, such as vaccine making, medicine quality control, and others. It's also often used to analyze some disease factors, medical risk, etc.

Industrial Statistics

In Industrial Statistics, we learn and study how to resolve an industrial problem such as quality control, queuing theory, optimizing variable, etc. It is very useful especially for them who work in manufacture or company management