

STATISTICS WORKSHEET

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

Ans-a- true

2. Which of the following theorem states that the distribution of averages of iid variables, normalized, becomes that of a standard normal as the sample size increases?

Ans- a- central limit theorem.

3. Which of the following is incorrect with respect to use of poisson distribution?

Ans-b- modeling unbounded count data.

4. Point out the correct statement.

Ans- a- the exponent of a normally distributed random variables follows what is called the log-normal distribution.

5. _____ random variables are used to model rates.

Ans- c- poisson.

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

Ans- b- false.

7. Which of the following testing is concerned with making decisions using data?

Ans- b- hypothesis.

8. Normalized data are centered at _____ and have units equal to standard deviations of the original data.

Ans-a-0

9. Which of the following statement is incorrect with respect to outliers?

Ans- c- outliers cannot conform to the regression relationship.

10. What do you understand by the term normal distribution?

Ans- the normal distribution is a continuous probability distribution that is symmetrical around its mean, most of the observation cluster around the central peak and the probabilities for values further away from the mean taper off equally in both directions.

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

Ans- 1. Mean or median imputations.

2. multivariate imputation by chained equations mice assumes that the missing data are missing at random (MAR).

3. random forest.

12. what is A/B testing?

Ans- A/B testing also known as split testing, or bucket testing.

12. Is mean imputation of missing data acceptable practice?

Ans- imputing the mean preserves the mean of the observed data. So if the data are missing completely at random, the estimates of the mean remain unbiased, and it is not a good solution.

13. 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.

14. What are the various branches of statistics?

Ans- There are three real branches of statistics

1. Data collection.
2. Descriptive statistics.
3. Inferential statistics.