

STATISTICS WORKSHEET-1

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. Usually replacing the standard error by its estimated value does change the CLT.
b) False
7. Which of the following testing is concerned with making decisions using data?
b) Hypothesis
8. 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

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?

Ans _The Normal Distribution is the most widely known and used of all distribution, because the normal distribution approximates many natural phenomena so well, it has developed into a standard of reference for many probability problems.

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 your statistical programme will make the decision for you.

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.

Imputation techniques-

- Mean imputation
- Substitution
- Hot deck imputation
- Cold deck imputation
- Regression imputation
- Interpolation and extrapolation
- Stochastic regression imputation
- Single or Multiple Imputation

12. What is A/B testing?

Ans_ A/B testing, also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable (web page, page element, etc.) are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drives business metrics.

13. Is mean imputation of missing data acceptable practice?

Ans_ Mean imputation reduces the variance of the imputed variables. Mean imputation shrinks standard errors, which invalidates most hypothesis tests and the calculation of confidence interval. Mean imputation does not preserve relationships between variables such as correlations

14. What is linear regression in statistics?

Ans_ Linear Regression -Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable,

and the other is considered to be a dependent variable. For example, a modeler might want to relate the weights of individuals to their heights using a linear regression model.

A linear regression line has an equation of the form

$$Y = a + bX,$$

where X is the explanatory variable and Y is the dependent variable. The slope of the line is b,

and a is the intercept (the value of y when x = 0).

15. What are the various branches of statistics?

Ans-

- Data Collection
- Data Presentation
- Data Analysis
- Interpretation of Data