

STATISTICS WORKSHEET-1

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question. (I have given only answers)

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.
A) True
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.
C) 1
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?

Normal Distribution is a bell-shaped frequency distribution curve which helps describe all the possible values a random variable can take within a given range with most of the distribution area is in the middle and few are in the tails, at the extremes.

- 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?

Multiple imputation is considered a good approach for data sets with a large amount of missing data. Instead of substituting a single value for each missing data point, the missing values are exchanged for values that encompass the natural variability and uncertainty of the right values.

The Imputation techniques are,

Numerical Variables

Mean/Median imputation

Arbitrary value imputation

End of tail imputation

Mode imputation

Categorical Variables

Frequent Category imputation

Adding a "Missing" Category

Both

Complete Case Analysis

Adding a "Missing" Indicator

Random Sample Imputation

12. What is A/B testing?

A/B testing statistics made simple. A guide that will clear up some of the more confusing concepts while providing you with a solid framework to AB test effectively.

And while there has been a lot of exceptional content written on AB testing statistics, I've found that most of these articles are either overly simplistic or they get very complex without anchoring each concept to a bigger picture.

the statistics of AB testing within a linear, easy-to-follow narrative. It will cover everything you need to use AB testing software effectively and I will make A/B Testing statistics simple.

13. Is mean imputation of missing data acceptable practice?

Mean imputation is the replacement of a missing observation with the mean of the non-missing observations for that variable.

Bad practice in general

If just estimating means: mean imputation preserves the mean of the observed data

Leads to an underestimate of the standard deviation

Distorts relationships between variables by “pulling” estimates of the correlation toward zero

14. What is linear regression 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

Statistics can be divided into 2 branches, they are

- **Descriptive Statistics.** Descriptive statistics is the first part of statistics that deals with the collection of data.
- **Inferential Statistics.** The inference statistics are techniques that enable statisticians to use the information.