STATISTICS WORKSHEET- 6

Q1 to Q9 have only one correct answer. Choose the correct option to answer your question.

1. Which of the following can be considered as random variable?

d) All of the mentioned

2. Which of the following random variable that take on only a countable number of possibilities?

a) Discrete

3. Which of the following function is associated with a continuous random variable?

a) pdf

4. The expected value or _____ of a random variable is the center of its distribution.

c) mean

5. Which of the following of a random variable is not a measure of spread?

d) all of the mentioned

6. The _____ of the Chi-squared distribution is twice the degrees of freedom.

a) variance

7. The beta distribution is the default prior for parameters between _____

c) 0 and 1

8. Which of the following tool is used for constructing confidence intervals and calculating standard errors for difficult statistics?

b) bootstrap

9. Data that summarize all observations in a category are called _____ data.

b) summarized

Q10and Q15 are subjective answer type questions, Answer them in your own words briefly.

10. What is the difference between a boxplot and histogram?

A10. Both Boxplot and Histogram are visualisation tools used to understand the distribution of data.

Histograms are bar charts that show the frequency of a numerical variable's values and are used to approximate the probability distribution of the given variable. It allows you to quickly understand the shape of the distribution, the variation, and potential outliers.

Boxplots communicate different aspects of the distribution of data. While you can't see the shape of the distribution through a box plot, you can gather other information like the quartiles, the range, and outliers. Boxplots are especially useful when you want to compare multiple charts at the same time because they take up less space than histograms.

11. How to select metrics?

A11. A Metric is a measure that is quantifiable.

Metrics are selected on the base of the type of data .

12. How do you assess the statistical significance of an insight?

A12. Statistical significance is a measure of the probability of the null hypothesis being true compared to the acceptable level of uncertainty regarding the true answer. It is denoted by alpha.

For 95% accuracy the significance level will be 0.05% ie., 5% of error is acceptable

13. Give examples of data that does not have a Gaussian distribution, nor log-normal.

A13. Categorical data does not have a Gaussian distribution , nor log-normal

Examples:

Winners of a race

People with a doctors degree

14. Give an example where the median is a better measure than the mean.

A14. When a distribution is skewed, the median does a better job of describing the center of the distribution than the mean.

Example:

Income level of people of a country

15. What is the Likelihood?

A15. Likelihood is the chance of occurrence of an event

Example: Tossing a coin