

STATISTICS WORKSHEET

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

Ans: D

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

Ans: A

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

Ans: A

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

Ans: C

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

Ans: C

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

Ans: B

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

Ans: C

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

Ans: B

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

Ans: B

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

Ans: Both the Histogram and Boxplot helps to explore data set in an easy and understanding manner. The Box plot is used for checking outliers or compare between data sets. Whereas the Histogram is used to check the data sets distribution.

11. How to select Metrics?

Ans: We select the Metrics based on the target variable. If the target variable is very continuous variable, we used R2 squared and Mean Absolute Error, If the

target variable is in categorical, we use classification report, accuracy score, confusion metrics, F1 score, precision

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

Ans: To assess statistical significance, you would use hypothesis testing. The null hypothesis and alternate hypothesis would be stated first. Second, you'd calculate the p-value, which is the likelihood of getting the test's observed findings if the null hypothesis is true. Finally, you would select the threshold of significance (alpha) and reject the null hypothesis if the p-value is smaller than the alpha — in other words, the result is statistically significant.

13.Give examples of data that doesn't have a Gaussian Distribution, nor log-normal.

- Body temperature
- People's Heights
- Car mileage
- IQ scores

Q14. Give an example where the median is a better measure than the mean.

Ans: Income is the classic example of when to use the median instead of the mean because its distribution tends to be skewed.

Q15. What is the Likelihood?

Ans: Likelihood function is a fundamental concept in statistical inference. It indicates how likely a particular population is to produce an observed sample.