

## STATISTICS

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

Ans. d) All of the mentioned.

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

Ans. a) Discrete

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

Ans. a) pdf

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

Ans. c) mean

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

Ans. a) variance

6. The \_\_\_\_ of the chi-squared distribution is twice the degrees of freedom.

Ans. a) variance

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

Ans. c) 0 and 1

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

Ans. b) bootstrap

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

Ans. b) summarized

10. Difference between boxplot and histogram.

Ans. Boxplot and histogram both are used for visualize the data. Histogram plots are better in determining the distribution of data whereas boxplots are better in determining the sets in data i.e quantiles and for visualize the outliers.

11. Metrics selection

Ans. If in a given dataset, target variable is continuous in nature then we will apply regression algorithms. And if in a given dataset, target variable is classification type (means yes or no in nature) then we will apply classification algorithms.

12. First of all we will check for null values. Then, will check for distribution of data, outliers present or not. Then, will check for multicollinearity problem. After that

again check for, how data is distributed overall. Then split the dataset into two variables. Then, standardize the data and will do train test split. And after all this process, will apply the required Algorithms according to the dataset.

13. Data of students passed in exam or fail in exam, source of income, in how much time phone battery dies.
14. Because mean is used to measure central tendency of normal distributed data whereas median is used for skewed data. So, Income may be the best example where median is a better measure than mean.
15. The chance of something happening.