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

Ans: a) The outcome from the roll of a die

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: b) standard deviation

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. What is the difference between a boxplot and histogram?

Ans: Histograms and box plots are graphical representations for the frequency of numeric data values. They aim to describe the data and explore the central tendency and variability before using advanced statistical analysis techniques. Both

histograms and box plots allow to visually assess the central tendency, the amount of variation in the data as well as the presence of gaps, outliers or unusual data points.

10. How to select metrics?

Ans:

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

Ans:1) state the research hypothesis

2) state the null hypothesis

3)select a probability of error level

4)select and compute the test for statistical significance

5)interpret the results

13. Give examples of data that doesnot have a Gaussian distribution, nor log-normal.

Ans: when we assume a Gaussian distribution for calculating statistics. This is because the Gaussian distribution is very well understood. So much so that large parts of the field of statistics are dedicated to methods for this distribution. Thankfully, much of the data we work with in machine learning often fits a Gaussian distribution.

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

Ans: The mean is the arithmetic average, and it is probably the measure of central tendency that you are most familiar. Calculating the mean is very simple you just add up all of the values and divide by the number of observation your dataset.

The median is the middle value. It is the value that splits dataset in half making it a natural of central tendency.

15. What is the Likelihood?

Ans: Likelihood is a confusing term. Likelihood is not a probability, but is proportional to a probability; the two terms can't be used interchangeably. In this post, we will be dissecting likelihood as a concept and understand it's importance in machine learning.