

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. 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. In this article, we will further discuss the similarities and differences between these two tools.

Both histograms and box plots are used to explore and present the data in an easy and understandable manner. Histograms are preferred to determine the underlying [probability distribution](#) of a data. Box plots on the other hand are more useful when comparing between several data sets. They are less detailed than histograms and take up less space.

11. How to select metrics?

ANS-

1. Good metrics are important to your company growth and objectives. Your key metrics should always be closely tied to your primary objective. ...
2. Good metrics can be improved. Good metrics measure progress, which means there, needs to be room for improvement. ...
3. Good metrics inspire action.

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

ANS- Steps in Testing for Statistical Significance

1. State the Research Hypothesis.
2. State the Null Hypothesis.
3. Select a probability of error level (alpha level)
4. Select and compute the test for statistical significance.
5. Interpret the results.

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

ANS - Exponential distributions do not have a log-normal distribution or a Gaussian distribution. In fact, any type of data that is categorical will not have these distributions as well. Example: Duration of a phone car, time until the next earthquake, etc.

Any type of categorical data won't have a Gaussian distribution or lognormal distribution. Exponential distributions - eg. The amount of time that a car battery lasts or the amount of time until an earthquake occurs.

14. 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. The median indicates that half of all incomes fall below 27581, and half are above it. For these data, the mean overestimates where most household incomes fall

15. 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.

Let $P(X; T)$ be the distribution of a random vector X , where T is the vector of parameters of the distribution. If X_o is the observed realization of vector X , an outcome of an experiment, then the function

The likelihood function itself is not probability (nor density) because its argument is the parameter T of the distribution, not the random (vector) variable X itself. For example, the sum (or integral) of the likelihood function over all possible values of T should not be equal to 1.

Even if the set of all possible values of the vector T is discrete, the likelihood function still may be continuous (as far as the set of parameters T is continuous).