STATISTICS

What do you understand by the term normal distribution?

In a normal distribution the mean is zero and the standard deviation is 1. A normal distribution is the proper term for a probability bell curve. Normal distributions are symmetrical, but not all symmetrical distributions are normal.

The normal distribution is a continuous probability distribution that is symmetrical on both sides of the mean, so the right side of the center is a mirror image of the left side. ... The normal distribution is often called the bell curve because the graph of its probability density looks like a bell.

How do you handle missing data? What imputations techniques do you recommend?

When dealing with missing data, we can use two primary methods to solve the error: imputation or the removal of data. The imputation method develops reasonable guesses for missing data. It's most useful when the percentage of missing data is low.

* Mean imputation. Simply calculate the mean of the observed values for that variable for all individuals who are non-missing. ...
* Substitution
* Hot deck imputation
* Cold deck imputation
* Regression imputation.
* Stochastic regression imputation.
* Interpolation and extrapolation.

What is A/B Testing?

A/B testing is a basic randomized control experiment. It is a way to compare the two versions of a variable to find out which performs better in a controlled environment.

For instance, let’s say you own a company and want to increase the sales of your product. Here, either you can use random experiments, or you can apply scientific and statistical methods. A/B testing is one of the most prominent and widely used statistical tools.

Is Mean Imputation of missing data acceptable practice?

 Imputing the mean preserves the mean of the observed data.  So if the data are [missing completely at random](https://www.theanalysisfactor.com/causes-of-missing-data/), the estimate of the mean remains unbiased. By imputing the mean, you are able to keep your sample size up to the full sample size.

What is linear regression in statistics?

Linear regression is a basic and commonly used type of predictive analysis.  The overall idea of regression is to examine two things: (1) does a set of predictor variables do a good job in predicting an outcome (dependent) variable?  (2) Which variables in particular are significant predictors of the outcome variable, These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables.  The simplest form of the regression equation with one dependent and one independent variable is defined by the formula y = c + b\*x, where y = estimated dependent variable score, c = constant, b = regression coefficient, and x = score on the independent variable.

What are the various branches of statistics?

Statistics plays a main role in the field of research. It helps us in the collection, analysis and presentation of data. In this blog post we will try to learn about the two main branches of statistics that is descriptive and inferential statistics.

Machine Learning

Which Linear regression training algorithm can we use if we have a training set with millions of features?

Batch gradient descent, stochastic gradient descent, or mini-batch gradient descent.

Which algorithm will not suffer or might suffer if the features in the training set have very different scales?

The normal equations method does not require normalizing the features, so it remains unaffected by features in the training set having very different scales.Feature scaling is required for the various gradient descent algorithms. Feature scaling will help gradient descent converge quicker.