P-value

The job of the p-value is to decide whether we should accept our Null Hypothesis or reject it. **The lower the p-value, the more surprising the evidence is, the more ridiculous our null hypothesis looks**. And when we feel ridiculous about our null hypothesis we simply reject it and accept our Alternate Hypothesis.

If we found the p-value is lower than the predetermined **significance value**(often called alpha or threshold value) then we reject the null hypothesis. The alpha should always be set before an experiment to avoid bias.

Chi-Square Test

Chi-Square test is used when we perform **hypothesis testing on two categorical variables** from a single population or we can say that to **compare categorical variables** from a single population. By this we find is there any significant association between the two categorical variables.

The hypothesis being tested for chi-square is:

Null: Variable A and Variable B are independent.

Alternate: Variable A and Variable B are not independent.

Multicollinearity

Multicollinearity exists whenever an independent variable is highly correlated with one or more of the other independent variables in a multiple regression equation. **Multicollinearity** is a **problem** because it undermines the statistical significance of an independent variable.

Dummy Variables

A dummy variable is a numerical variable used in regression analysis to represent subgroups of the sample in your study.

Dummy variables are useful because they enable us to use a single regression equation to represent multiple groups.

Weight of evidence and Information Value

Weight of evidence (WOE) and Information value (IV) are simple, yet powerful techniques to perform **variable transformation and selection**. These concepts have huge connection with the logistic regression modeling technique. It is widely used in credit scoring to measure the separation of good vs bad customers.

Coarse classing

Coarse classing consolidates classes into more stable and statistically significant classes.