

Summary/Review

Exploratory Data Analysis

EDA is an approach to analyzing data sets that summarizes their main characteristics, often using visual methods. It helps you determine if the data is usable as-is, or if it needs further data cleaning.

EDA is also important in the process of identifying patterns, observing trends, and formulating hypothesis.

Common summary statistics for EDA include finding summary statistics and producing visualizations.

Feature Engineering and Variable Transformation

Transforming variables helps to meet the assumptions of statistical models. A concrete example is a linear regression, in which you may transform a predictor variable such that it has a linear relation with a target variable.

Common variable transformations are: calculating log transformations and polynomial features, encoding a categorical variable, and scaling a variable.