Tutorials for Using R and Python for Linear Regression

Kristin Porter

2024-03-20

Regression analysis serves as a versatile tool in statistics, used for both exploring the relationships between variables and for making predictions.

When used for exploring relationships, regression helps to understand how dependent variables change in relation to one or multiple independent variables. This is how you might use regression in projects for this class. This approach is particularly valuable in research and analysis, where the objective is to identify patterns, trends, and underlying mechanisms. For instance, an analyst might use regression to explore how average hours of sleep is related to academic performance, focusing on understanding the nature and strength of the relationship.

Conversely, when used for making predictions, regression focuses on utilizing known data to forecast unknown values of the dependent variable based on the relationship derived from historical data. For example, a company might use regression analysis to predict future sales based on past advertising spending, past consumer spending and economic contextual factors, aiming to make informed decisions. Regression analysis for prediction is beyond the scope of this class. However, if this is a topic that is of interest to you

The primary difference between these two applications lies in their objectives: exploration seeks to understand and describe relationships without necessarily predicting future outcomes, while prediction aims to use those relationships to forecast unobserved values, often with an emphasis on the accuracy of these predictions.

R Tutorials

These tutorials focus on linear regression, which you would use when your dependent variable (Y) is continuous.

Linear Regression in R | A Step-by-Step Guide & Examples

A Gentle Introduction to Regression in R

Here is a tutorial for logistic regression, which you would use when your dependent variable is binary:

Logistic Regression: R Data Analysis Examples

Python Tutorials

Most python resources focus on using regression for predictive analytics - to build a model for making predictions in *new* data - different data than data used to estimate the model. However, this seems to be a decent tutorial for using linear regression for understanding the relationships between variables.

Linear Regression With Python