

Regression Diagnostics - Introduction



<https://github.com/learn-co-curriculum/dsc-regression-diagnostics-intro>



<https://github.com/learn-co-curriculum/dsc-regression-diagnostics-intro/issues/new/choose>

Introduction

In addition to model evaluation and coefficient interpretation, regression diagnostics will allow you to level up your modeling abilities. This means applying techniques to "diagnose" why a model is underperforming, then "treat" the problem using an appropriate technique.

Model Transformations

These techniques are the "treatments" for problems with your model. First we'll go over the different types of transformations, how to apply them, and how they impact the modeling results.

Linear Transformations

Linear transformations include **scaling** a variable to change its units, **shifting** a variable to change its mean, and **normalization** and **standardization** which combine both scaling and shifting.

These transformations change the distributions of the variables and therefore also the interpretation of the coefficients, but are still part of a regular linear regression process.

Extensions to Linear Models

Unfortunately, not every continuous variable can be predicted effectively using a straight line (a linear model). The goal of these kinds of transformations is to transform the variables in a non-linear way so that the relationship between the transformed variables is linear.

Extensions to linear models include **log transformations**, **interaction terms** (for when two or more variables interact in a non-additive manner), and **polynomial regression** (bringing in higher powers of variables -- squared, cubed, etc.).

Linear Regression Assumptions

Most often the reason that you will need to apply one of the transformations described above is that you have "diagnosed" a model as violating one of the assumptions of linear regression.

We'll use the acronym **LINE** to describe these assumptions. This is also a helpful device for memorizing the assumptions, since this question sometimes comes up in technical interviews!

The assumptions of linear regression are:

L: Linearity

I: Independence

N: Normality

E: Equal Variance (Homoscedasticity)

For each assumption we'll go over what it means, how to assess whether the assumption is being met (using graphs as well as statistical tests), and how to apply transformations to address assumption violations.

Summary

It's time to start learning some more advanced regression techniques!

How do you feel about this lesson?



Have specific feedback?

[Tell us here!](https://github.com/learn-co-curriculum/dsc-regression-diagnostics-intro/issues/new/choose)  [\(https://github.com/learn-co-curriculum/dsc-regression-diagnostics-intro/issues/new/choose\)](https://github.com/learn-co-curriculum/dsc-regression-diagnostics-intro/issues/new/choose)