

## Overview

Now that we've developed a model for our Ames housing data set, we want to interpret the relationships between features of homes and sale prices, and make predictions about future sale prices based on those features. But before we can use our model, we must determine whether we can trust it. Have we met the assumptions of the model? Our linear regression model assumes a linear relationship between the response and the predictors. It also requires three assumptions about the errors: they're normally distributed, have a constant variance, and are independent. Meeting these assumptions is necessary before we can trust the standard errors, confidence intervals, and p-values that are produced by our model, or simply, make any inferences from our model.

In this lesson, you learn to verify these assumptions and diagnose problems that you encounter in linear regression. You'll learn to examine residuals, identify outliers that are numerically distant from the bulk of the data, and identify influential observations that unduly affect the regression model. Finally, you learn to diagnose collinearity to avoid inflated standard errors and parameter instability in the model.

---

*Statistics 1: Introduction to ANOVA, Regression, and Logistic Regression*

Copyright © 2019 SAS Institute Inc., Cary, NC, USA. All rights reserved.

Close