Print

Identifying Associations in Linear Regression with Scatter Plots

Now let's turn to regression. Suppose you want to look for an association between the predictor variable Above Ground Living Area and SalePrice. Scatter plots are two-dimensional graphs produced by plotting one continuous variable against another continuous variable within a set of coordinate axes, and they describe the relationship between the two variables. A linear association can be inferred when the general shape of a scatter plot is a straight line. Other types of possible associations appear in different shapes.

In addition to using scatter plots to examine the relationship between two continuous variables, you can also use them to detect outliers, identify trends in your data, identify the range of X and Y values, and communicate the results of a data analysis. Let's look at some scatter plot examples.

The first plot with the relatively straight line shows a positive, linear relationship between the two variables. As the values of one variable increase, the values of the other variable increase as well.

In the second scatter plot, the relationship is curved. Later, when you learn about fitting a model to the data, you'll need a quadratic or squared term in the model for this type of data.

The third scatter plot shows a cyclical relationship. You often see this type of relationship in seasonal data where the pattern repeats itself over time.

The last scatter plot displays a random pattern of points. No clear relationship exists between the variables.

Close