

## Overview

Imagine we were interested in studying human lifespan. So many variables can potentially affect life expectancy. We might want to model lifespan using categorical predictors like race, socioeconomic status, or country of origin. Or perhaps we could use continuous predictors like the person's resting heart rate, cholesterol level, or T-cell counts.

Think about the Ames data. Several variables can affect the price of homes. Perhaps categorical variables, such as heating quality and the presence of central air, are associated with significantly different sale prices. Or maybe continuous variables, such as above ground living area and lot size, are linearly associated with sale price. Because so many potential predictors could be important in modeling sale price, we need tools to explore and help us choose which predictors might be important.

We'll start this lesson by using some graphical tools that can help us determine which predictors are likely or unlikely to be useful. These graphical explorations can be augmented with correlation analyses that describe linear relationships between potential predictors and our response variable. After we determine potential predictors, tools like ANOVA and regression will help us assess the quality of the relationship between the response and predictors.

---

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

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

Close