Case Studies

CONTEXT FOR USE: This flipped activity has been used to introduce multivariate regression in advanced undergraduate statistics courses. It has also been used as a final project in introductory statistics courses.

LEARNING GOALS:
- Reduce variability of students’ knowledge of simple linear regression.
- Help students recognize that simple linear regression often fails to provide a sufficient model.
- Provide a starting point for students to work through the entire process of model building and assessment in multivariate regression.

DETAILED DESCRIPTION:
Students read a two page introduction before class. This includes a description of how people estimate used car values and a quick review of the simple linear regression model. Students are asked to create a simple linear regression model that will address the question: "Are cars with lower mileage worth more?" Clearly it seems reasonable to expect to see a relationship between mileage (number of miles the car has been driven) and retail value. Students try fitting a simple linear regression model Price = 24723 – 0.17 Mileage. They also calculate the t-statistic for the slope coefficient (b1): t = -4.09 (p-value < 0.001) and the R-Sq value = 2.0%.

These results can lead to some nice review questions that students answer before class:
1) In general, what happens to price when there is one more mile on the car?
2) Does the fact that b1 is small (-0.17) mean mileage is not very important?
3) Does mileage help you predict price? What does the p-value tell you?
4) Does mileage help you predict price? What does the R-Sq value tell you?

During class students work in small groups to develop a regression model that includes several quantitative explanatory variables.

INSTRUCTOR NOTES:
By discussing these questions at the beginning of class (or by reading through students quiz responses before class), instructors can identify if any concept within simple linear regression needs review. Traditional textbook examples usually have data that fit very nicely, with high R-sq and low p-values. In this example, the simultaneously small p-value and small R-Sq value seemingly send contradictory messages about the model.

In addition, a case study gets students engaged. Students come to class seeing simple linear regression has some problems and wanting to know how to find the answer. This greater level of investment encourages greater student learning and a passion for knowing how to find a solution to the question at hand.
RESOURCES:
This case study is within the text, Practicing Statistics: Guided Investigations for a Second Statistics Course (Kuiper and Sklar 2012) and a more complete description of this activity is available at http://www.amstat.org/publications/jse/v16n3/datasets.kuiper.html.