

Introduction to Statistical Modeling

Daniel Kaplan & Karen Saxe, Macalester College

Used Car Prices

[used-cars]

Car prices vary. They vary according to the model of car, the optional features in the car, the geographical location, and the respective bargaining abilities of the buyer and the seller.

We are going to investigate the influence of three variables on the asking price for used cars:

- Model year
- Mileage
- Geographical location

We have picked these variables because they are relatively easy to measure and there are web sites that allow us quickly to collect information about them. Of course, using just these variables leaves off a lot of information such as the physical condition of the car, whether it was properly maintained, and so on.

For this project, go to a used car web site. One site that seems easy to use is www.cars.com. Pick a particular model of car that is of interest to you. Also, pick a few scattered geographical locations. (At www.cars.com you can specify a zip code, and restrict your search to cars within a certain distance of that zip code.)

| Year | Vehicle | Price ↓ | Mileage |
|------|-------------------------------------|----------|---------|
| 2006 | Mazda Miata MX-5 | \$23,995 | 969 |
| 2004 | Mazda Miata MX-5 | \$19,977 | 10,198 |
| 2004 | Mazda Miata MX-5 | \$16,999 | 20,560 |
| 2002 | Mazda Miata MX-5 SE | \$15,995 | 18,826 |
| 2003 | Mazda Miata MX-5 | \$15,995 | 15,947 |
| 2002 | Mazda Miata MX-5 | \$14,995 | 44,215 |
| 2002 | Mazda Miata MX-5 LS | \$14,500 | 18,000 |
| 2002 | Mazda Miata MX-5 | \$14,500 | 1,800 |
| 2001 | Mazda Miata MX-5 | \$13,995 | 70,136 |
| 2002 | Mazda Miata MX-5 | \$12,995 | 43,978 |
| 1999 | Mazda Miata MX-5 | \$9,500 | 63,000 |
| 1999 | Mazda Miata MX-5 | \$9,500 | 63,000 |
| 1991 | Mazda Miata MX-5 | \$4,200 | 90,000 |

Price and other information about used Mazda Miatas in the Saint Paul, Minnesota area from www.cars.com.

For each location, use the web site to find prices and the other variables for a couple of dozen cars. Record these in a spreadsheet with five variables: price, model year, mileage, location, model name. (The model name will be the same for all

your data. Recording it in the spreadsheet will help in combining data for different types of cars.)

Write a one- or two-page report accounting for used car prices. Imagine that the client for your report is a car rental agency that has a steady stream of used cars to sell and wants to figure out the best age and mileage and location at which to sell them. You do NOT need to figure out this optimization problem; your job is just to give helpful information about the structure of used car prices so that someone else can solve the optimization problem.

You can assume that your reader is technically and mathematically sophisticated but is not knowledgeable about statistical modeling, so you will have to explain how to interpret model coefficients and so on. Include statistical graphics and model reports as you think appropriate.

Some questions you may want to think about in writing your report:

- What is the relationship between price and model year?
- What is the relationship between price and mileage?
- What is the relationship between model year and mileage?
- Are outliers playing an influential role?

- What is the typical variability in price around a model value?
- What role, if any, does the geographical location play in determining the price?
- How reliable are your estimates of the relationships among price, model year, mileage, and location?

There will be three phases to this project:

1. Collect your data and write your report.
2. We will select a few reports and ask you to write a commentary or critique.
3. Toward the end of the semester, you will have a chance to reconsider your report and describe how you would do things differently if you were starting over.

Write your report using a conventional word processing program. Then upload your file to the Moodle “Statistical Assignments” page.