

# STAT 428/528, HW1

For this homework, you will read in the data set `cars2.txt` located at

`http://www.math.unm.edu/~james/cars2.txt`

You can either use, for example,

```
x <- read.table("http://www.math.unm.edu/~james/cars2.txt",header=T)
```

or download the data and read it in from your own computer (or lab computer).

The dataset has variables for the model year of the car, the price of the car, the mileage, and the title status (salvage, clean, etc.). The title status won't be used for this homework. The data was collected from Craigslist in Albuquerque in 2014.

For this homework, answer the following questions in a typed report.

1. Make a scatterplot matrix for the variables **year**, **price**, and **mileage**. Do not include the title status in this plot. Comment on any trends you see in the plots.
2. Fit the following three regression models, and write out the regression equation for each model

`price = year + mileage`

`price = year`

`price = mileage`

Your regression equation should give the expected price as a function of the predictor, including the intercept term. For example, the regression equation used in class for the length of the chile peppers using both width and thickness was

$$\text{length} = 3.2619 + 2.1134 \times \text{width} + 0.8183 \times \text{thickness}$$

3. What is the predicted price for a 2005 car with 200,000 miles using all three models? Show how you obtain this answer using the regression equations.
4. Although neither mileage nor year are significant at the  $\alpha = .05$  level for any of the models, based on the output, do you think mileage or year is the better predictor? Justify your answer?
5. Repeat problems 1–4 using only cars with model years later than 1980. (This might change whether or not any variables are significant.)