

Math 736 Homework Assignment #1: **Due Wednesday, September 26th**

The R dataset “Cars93”, stored in the MASS library, provides records for 93 cars selected at random from among 1993 passenger car models that were listed in both *Consumer Reports* and the *PACE Buying Guide*. The variables included are the following:

Manufacturer: Manufacturer

Model: Model

Type: Small, Sporty, Compact, Midsize, Large, Van

Min.Price: Minimum Price (in \$1,000) - price for a basic version

Price: Midrange Price (in \$1,000) - average of 'Min.Price' and 'Max.Price'

Max.Price: Maximum Price (in \$1,000) - price for "a premium version"

MPG.city: City MPG (miles per US gallon by EPA rating)

MPG.highway: Highway MPG

AirBags: Air Bags standard. Factor: none, driver only, or driver& passenger

DriveTrain: Drive train type: rear wheel, front wheel or 4WD; (factor).

Cylinders: Number of cylinders (missing for Mazda RX-7, which has a rotary engine).

EngineSize: Engine size (litres)

Horsepower: Horsepower (maximum)

RPM: RPM (revs per minute at maximum horsepower)

Rev.per.mile: Engine revolutions per mile (in highest gear)

Man.trans.avail: Is a manual transmission version available? (yes or no, Factor).

Fuel.tank.capacity: Fuel tank capacity (US gallons)

Passengers: Passenger capacity (persons)

Length: Length (inches)

Wheelbase: Wheelbase (inches)

Width: Width (inches)

Turn.circle: U-turn space (feet)

Rear.seat.room: Rear seat room (inches) (missing for 2-seater vehicles)

Luggage.room: Luggage capacity (cubic feet) (missing for vans)

Weight: Weight (pounds)

Origin: Of non-USA or USA company origins? (factor)

Make: Combination of Manufacturer and Model (character)

Your assignment is to fit a regression model that predicts “Price” using the other variables.

Your model MAY NOT include the variables “Manufacturer”, “Model”, “Min.Price”, “Max.Price”, or “Make,” although you may reference these in your discussion. Write a brief report including the following:

- Preliminary analysis. How are the variables distributed? Are there any outliers?
- Model fitting. Which variables are the best predictors? How did you evaluate different models? Are there highly correlated variables? Are any transformations needed?
- Conclusions and interpretation. What have you learned about this dataset? Does your model seem intuitively logical? If not, why have you chosen these predictors?

Your report should be clear to *any* reader with a basic understanding of statistical models, so make sure that you explain your results and provide explanations for any figures and output that you include.