Math 6020-1: Spring 2013
Problem set 2

Due date: February 15, 2013

Be sure that you keep a copy of your analysis for your later use, in addition to the copy that you turn in for grade. This is important, because you will need this analysis later.

1. Download the "Passenger Car Mileage" data file from Carnegie Mellon's DASL library. This is based on 1991 EPA emissions data, and is publicly available.

2. Use the data to write a linear model for predicting the fuel efficiency of a car, based on the model,

$$\text{MPG} = \beta_0 + \beta_1 \text{VOL} + \beta_2 \text{HP} + \beta_3 \text{SP} + \beta_4 \text{WT} + \epsilon,$$

   where MPG denotes "miles per gallon," VOL denotes "passenger cab volume in cubic feet," HP denotes "engine horse power," SP denotes "top speed in mph," WT denotes "the vehicle weight in hundreds of pounds," and ϵ is "noise."

3. Find point estimates for $\beta_0, \ldots, \beta_4$, using the usual least-squares estimators of math 6010.

4. Assume the homeoscadastic normal–error model for the noise, and use it to construct confidence intervals in order to decide which, if any, of the explanatory variables are redundant.

5. What is your least-squares predictor of the MPG of a car whose vital statistics are: VOL= 91 ft^3; HP= 85; SP= 72 mph; and WT= 5320 pounds?

6. Let $\boldsymbol{X}$ denote the vector (VOL , HP , SP , WT) that corresponds to a randomly-selected vehicle in 1991. How would you estimate $\boldsymbol{\Sigma} := \text{Cov}(\boldsymbol{X})$? Perform this estimation, and explain [in as much detail as you can] your reasoning.