Math 6020-1: Spring 2013
Problem set 3

Due date: March 8, 2013

This problem set uses the same data as your problem set 2 did. The "Passenger Car Mileage" data file is from Carnegie Mellon's DASL library.

1. Double-check your estimate $\widehat{\boldsymbol{\Sigma}}$ for the covariance matrix $\boldsymbol{\Sigma}$. In particular, make sure that you have $\widehat{\sigma}_{0,0} \approx 491$ [this is your estimate for the variance of VOL].

2. Estimate the correlation matrix $\boldsymbol{\rho}$. Describe your estimator carefully, and explain why you think it is a reasonable estimator.

3. Identify the elements of $\widehat{\boldsymbol{\rho}}$—your estimate for the matrix $\boldsymbol{\rho}$—that are suspiciously close to one in absolute value. What do you think should be done about these terms in your linear model of the last assignment?

4. Make your answer to the previous more precise by actually carrying out a complete PCA on $\boldsymbol{\Sigma}$, and a complete PCA on $\boldsymbol{\rho}$.

5. Which PCA do you recommend here? Explain your reasoning carefully.