Stat 666
HW #6a
Due TBA

1. Conduct a principal component analysis of the Stat 2301 grade data described in Example 14.2.2 on page 484 of the text, utilized in SectIVB.sas, and found in `Gradedat.txt` on the webpage. Omit the Labquiz and the Project variables for these analyses, using only the other 6 variables.

   (a) Conduct a PCA using the covariance matrix **S** and address the following issues:

      i. Use each of the first three methods discussed in class to recommend a number of components to be retained.

      ii. Give an interpretation of the important components. (This is sometimes difficult, but give it your best shot.)

   (b) Conduct a PCA using the correlation matrix **R** and address the same issues as in part (a).

2. Conduct an exploratory factor analysis of the correlation matrix from the pollution data given in `receptor2.txt` on the webpage. The data consist of hourly ambient measurements of volatile organic compounds in the El Paso, Texas area. Your goal is to assess whether or not these compounds are being driven by some small number of pollution sources or "factors." The variable names for the 11 columns in the data set are:

   ```
   Ethylene Propane nButane x2Mp x3Mp Benzene Cyclohexane x2Mh x224Tmp Acetylene Ethane
   ```

   (a) Use the principal factor method to estimate the factor loadings. Use the "priors=smc" option to define the initial values for the communality to be equal to the squared multiple correlation with all the other variables (as discussed in class.) Make sure to justify your choice for the number of factors.

   (b) Use the Chi-squared goodness-of-fit test to assess the fit of 1-factor, 2-factor, or 3-factor models. *[Note: As it turns out, the temporal correlation among the observations is the problem here. The 2 factor model is actually a good fit. For further edification (i.e., "this is not required") see Technometrics, vol. 44, 328-337.]* Use maximum likelihood to estimate the factor loadings for the two-factor model. Use the promax rotation and interpret the factors. *[Hint: Auto exhaust is known to be high in both acetylene and ethylene, while industrial emissions in this region have larger amounts of propane and ethane.]*