Assignment 5
Note this assignment is worth 20pts

due Tuesday October 31, 2011 at the beginning of the class.

The following questions refer to the attached dataset which may be found in the file *happiness.xls*. A description of the data may be found in your textbook on page D-4 in the appendix (6. Happiness). Please read the description of the variables in the dataset. The data was collected between 1995 and 2005. The objective of this assignment is to relate the variables in the data with life expectancy variable.

1. To do this first create scatterplots with all the potential x variables in the data.

2. Then construct a multivariate regression model using all the variables. Obtain the residuals. Perform residual check to verify the validity of the model. Is the model good? Can it be improved. What (if any) assumptions of the model are violated?

3. Using the ideas presented in the lecture 11 and chapters 10 and 11 of your book please construct the best possible regression model. Check the validity of the final model constructed.

4. Using the final model you constructed predict the average life expectancy for a country not included in the data set (Cheesia) which happened to score LSI=7, GINI=40, CORRUPT=3 and DEMOCRACY=5. Also give a 95% confidence interval for the average life expectancy in this particular country.

5. Repeat the previous prediction (including the confidence interval) for Fantasia who we all know scored LSI=2, GINI=37, CORRUPT=10 and DEMOCRACY=7 in 2008, the last time it was surveyed.