Stat 704, Homework 2

1. Complete problems 1.36, 2.52 and 2.57(a)

2. **Grade Point Average** (Use SAS): 1.19(a-c), 1.23 (instead of 1.23a, discuss SAS's residual diagnostic plots), 2.4(a-c), 2.13(a,b), 2.23(a-d)

3. **Water Quality data** (Use SAS). The Excel file on your website contains 2009 water quality data from numerous monitoring stations throughout South Carolina. Counts of bacterial colonies were recorded for each of Fecal Coliform, Enterococci, and Escherichia Coli. At the time, SC DHEC (Department of Health and Environmental Control) was exploring whether to replace the water quality standard for E Coli with another bacterial parameter (Fecal Coliform was eventually chosen as the new standard).

   - Load the Excel file from our website into SAS and extract records from the Congaree River watershed as a separate data set; these records all start with the substring "C-". Plot E Coli (x-axis) vs Fecal Coliform (y-axis) and comment.

   - We will regress log(Fecal Coliform) on log(E Coli), since the former was proposed as a more reliable measure of E Coli; transform both responses and plot them against each other. Comment on whether a linear model seems appropriate here. Is variation in Fecal Coliform constant for all levels of E Coli?

   - Regress log(Fecal Coliform) on log(E Coli) and plot the fitted line–use ln, not $\log_{10}$. Comment on the least squares parameter estimates and least squares line. Interpret the test on the slope parameter. Is there a strong linear association?

   - Suppose a stream sample has an E Coli count of 100. Construct a 90% CI for mean log Fecal Coliform count for that stream sample. Now construct a 90% prediction interval for a single Fecal Coliform count for that same stream sample. Exponentiate the endpoints of the CI and PI and interpret.

   - If the 2009 daily maximum threshold for E Coli was 250 colonies/sample, what would you recommend as a new Fecal Coliform standard?

4. **Crime Rate** (Use R): 1.28