APPLIED STATISTICS
TUTORIAL 4
Questions revised from "The Statistical Sleuth"

Question 1 (revised based on ex 9.15 from "The Statistical Sleuth")
The data on corn yields and rainfall, discussed briefly in class, can be found in the file "corn.csv". This file contains the corn yield (bu/acre), rainfall (in/year), and the corresponding year.

a) Plot corn yield versus rainfall.
b) Fit the multiple linear regression of corn yield on rain and rain^2.
c) Plot the residuals versus year. Is there pattern evident in this plot? What does it mean?
d) Fit the multiple regression of corn yield on rain, rain^2, and year. How do the coefficients of rain and rain^2 differ from those in the estimated model in (b)? How does the estimate of σ differ? How do the standard errors of the coefficients differ? Describe the effect of an increase of one inch of rainfall on the mean yield over the range of rainfalls and years.
e) Fit the multiple regression of corn yield on rain, rain^2, year, and year$\times$rain. Is the coefficient of the interaction term significantly different from zero? Interpret the interaction term?

Question 2 (revised based on ex 9.20 from "The Statistical Sleuth")
The Kentucky Derby is a 1.25 mile horse race held annually at the Churchill Downs race track in Louisville, Kentucky. The file "derby.csv" contains the data on the year of the race, the winning horse, the conditions of the track, and the average speed (in feet per second) of the winner, for years 1896-2000. The track conditions have been grouped into three categories: fast, good, and, slow. Model the mean winning time as a function of year and track conditions. (Hint: The first thing you should do is look at a plot of winning speed versus year. What does a curved plot suggest?)

Question 3 (revised based on ex 10.09 from "The Statistical Sleuth")
As part of a study of the effects of predatory intertidal crab species on snail populations, researchers measured the mean closing forces and the propodus heights of the claws on several crabs of three species. This data is contained in the file "crab.csv".

a) Fit a regression model of log(force) on log(height) and species, allow for an interaction between log(height) and species. Let Hemigrapsus nududus be the baseline species, i.e., do not use an indicator variable for this species.
b) What is the p-value for the test of the hypothesis that the slope in the regression of log(force) on log(height) is the same for Lophopanopeus bellus as it is for Hemigrapsus nududus?