STATS 500, HOMEWORK #2, due Wednesday, Sept, 29 3PM

1. Using the dataset `gala` we discussed in class.

   (a) Fit a regression model with "Endemics" as the response and "Area", "Elevation", "Nearest", "Scruz", "Adjacent" as predictors. Give a short summary of what you find. please also provide a boxplot of the residual.

   (b) Which observation has the largest absolute residual? Please give the case number.

   (c) Compute the mean and median of the residuals. Explain what the difference between the mean and the median indicates. Should you worry about this difference here?

   (d) Compute the correlation of the absolute values of the residuals and the fitted values. Plot the absolute values of the residuals against the fitted values. Please comment on what you have learned from this correlation and the plot.

   Hints: Useful R functions for the homework: library(), data(), lm(), summary(), residuals(), fitted(), which.max(), abs(), mean(), median() and cor().

2. Let r be the sample correlation of X and Y; both are continuous variabels. Also let $sd(X)$ and $sd(Y)$ be their sample standard deviation. Now let $\hat{\beta}_x$ be the LS estimated slope when regressing Y on X. Please give an equation that relates r to $\hat{\beta}_x$.

   Hints: Your equation should contain $sd(X)$ and $sd(Y)$.

3. For the problem above, if we now regress X on Y and let the estimated LS slope be $\hat{\beta}_y$. Would $\hat{\beta}_y = 1/\hat{\beta}_x$?

4. On page 35 of the class notes # 3 (STATS 500 Note-3), you were asked to perform a t-test of $H_0 : \beta_{ddpi} = 0.5$, vs. $H_A : \beta_{ddpi} \neq 0.5$. Please report your test outcome here. Please also explain why the p-value for your t-test should be the same as the p-value obtained by the F-test on page 35 (value = 0.6475).