STA 471 Practice Problems 3

1. a. After a regression model has been fit to a data set and the investigator is satisfied the model adequately summarizes the relationship between the variables in question, why bother performing diagnostics using the residuals?
b. Why does one check for normality, constant variance, etc. based on the residuals? Why not use the Y-data?
c. Explain the relationships amongst $\widehat{e_i}, r_i$, and t_i.
d. True or False? Externally studentized residuals are uncorrelated since they are based on $\widehat{\sigma^2}_{(i)}$.
e. What is the difference between outliers and influential observations?

2. For the first-order autoregressive model, show that the correlation between e_t and e_{t-1} is ρ. What is the correlation between e_t and e_{t-2}? Please explain.

3. Please explain what is meant by a "maximum" model. How does one go about determining the "maximum" model?

4. a. Is it better to use $R^2_{adjusted}$, $\widehat{\sigma}$, C_p, AIC_c, or BIC as the criterion to select the "best" regression model from a collection of possible models? Please explain.
b. Under what scenarios is R^2 a viable criterion to compare different regression models?