Jared Yu
Module 7 Discussion

In the first part regarding the various residuals, I also listed the ones that you did. I did include the initial residual, but I think that was something extra. The definitions you provided also make sense to me. I am used to writing in LaTeX, so I just included the formulas also. However, I think the important part is to explain why each are useful. You describe first standardized residuals and how they're important for finding outliers. Then, you discuss the studentized residuals, and how they're related to leverage and influential points. There's also a point about how it's regarding points that may need removal, rather than outright removing them. Afterward, you discuss the PRESS residual. I think these are an interesting concept, in particular how they're considered internal. You also relate the standardized version of the PRESS residual to studentized residuals which I thought was interesting.

I think also that your description of the difference between random errors and residuals goes with what I tried to say also. My thinking is also that the random errors are theoretically based, while the residuals are what we can derive from sample data and our fitted model. I am not sure about the difference, I tried to elaborate more on the meaning to try and make a point about each. I think what you said about one being compared to the true, while the other being compared to the fitted value makes sense also.

Your approach to answering the last bullet is also interesting to me. You first state all the assumptions and make points about what happens if they're violated. This is a good approach I feel, and I think it makes good sense for the question being asked. The first assumption is about the model being linear. I recall the book saying that we could try other predictors that are nonlinear such as x^2. If the mean is nonzero for epsilon, I am curious as to what the data would look like. Your idea about the bias term and adding a regressor is interesting to me. I think when the book discusses inconstant variance, they make many cases about what could be the reason why. I think the timeseries case is particularly interesting. Correlated errors seems like an interesting problem to me. I haven't done too much regression analysis where I am looking so carefully at the errors. You mention that some information could be missing, I suppose this could come in the form of a new predictor. The last point about uniform errors is interesting. I hadn't considered that non-normal errors would impact the R^2 metric.