Jared Yu
Module 10 Discussion

In my discussion I actually forgot to consider the all possible regressors method and only looked at the other three methods you discussed. The argument that you make seems pretty sensible for this method. That is, the individual $r_{x_i,y}$ don't matter so much, since they can be placed by some alternate subset that provides a higher R^2_{Adj}.

In the forward selection method, I also agree that $r_{x_i,y}$ has a great impact on the model, since the first regressor chosen is based on this metric. I noticed that in the subsequent regressors, when it is looking at the *partial correlations*, it seems that the value will be determined based on the sample mean of the next regressors. This is because the slope will evaluate to zero when looking at just the simple linear regression formula. The book didn't seem to go into too much detail for the third, fourth, etc. regressors, so I am not sure if the partial correlation would then turn into the multiple linear regression formula. However, if it stays in the simple linear regression form, then it would seem that the idea that regressors are chosen based only on their sample mean is kind of strange.

In the backward elimination method, you make a powerful point that I didn't notice. I tried to say that in the partial F statistic, that the simple correlation is not directly impacted. You however state that $r_{x_i,y}$ is proportional to the t value, therefore my argument is definitely wrong. Something else you mention is that regressors with the smallest simple correlation are removed. This makes sense to me also, but when I looked at the steps I wasn't able to determine directly how exactly it is related.

Regarding stepwise selection, I think we both try to say that $r_{x_i,y}$ does have a large impact, but I think we have different reasons for why. I tried to say that it was highly similar to the forward selection method, and so the partial correlation idea from before applies here also. I didn't really notice how it's also a combination of backward elimination too, where at each step it performs one step of backward elimination. When looking at the backward elimination portion, given your argument about $r_{x_i,y}$ being related to the t value, I think then that there is definitely an impact in the resulting decision, at least at each stage of the process. I am curious though if the initial $r_{x_i,y}$ calculations will have a large impact on the resulting subset. Perhaps it is instead as you say, it ultimately is more akin to the all possible regressors, where it's the value of the chosen subset, rather than each sequential subset.