Jared Yu
Module 5
Discussion Reply

In my first reply, I noted also there are a few differences between your model and the one that I had written out. I will quote my reply from the first discussion:

"I agree with your methodology for including indicator variables to replace the categorical variables in the model. My model is almost the same with some minor differences…I also swapped around the variable indices x_1 and x_3 to avoid using something like x_{11} and x_{12} to help make the indicator variable easier to notice. The result is that I used x_3 and x_4 in place of x_{11} and x_{12}, while also moving the quantitative variable to x_1. A major difference however is that I notice you included all the interaction terms. I wasn't sure if this was needed so I skipped it. I mentioned however that an assumption is that the first-order model is sufficient, while also stating that additional terms could be added to reflect interaction effects in the model."

I also agree with your methodology for creating separate regression models to compare the different levels of x_1 and x_2. The difference that I notice is that since you include interaction terms, your regression models are slightly larger than mine, but there are still only six total models to coincide with the six total combinations of various levels between x_1 and x_2.

Something else that I notice is that our hypotheses tests are different. Regarding the first part of the question, I interpreted it as asking us the impact of x_3 on y. I read this as looking at the partial F test where only x_3 is being tested for significance, given the other regressors. From my understanding, you seem to be testing all the other regressors being 0 and looking at if having only x_3 is sufficient.

Furthermore, for the second part of the question, I interpreted it as asking whether any of the multiple regression models (i.e., the six that were redesigned) would also be able to be sufficient for a partial F test regarding the variable x_3. In other words, I would perform the same partial F test as in the first part of the question, but on different "full models (FM)" where the FM depended on the variables included in each of the six redesigned regression models. For example, in the second regression model, the FM would include β_0, β_{12}, β_3, and β_5.

To be honest, I still can't tell if either of our interpretations are correct. However, looking at your work it seems like the response that you've created makes sense given your own understanding of the problem.