Keep in mind that we will want some commentary on how well the models generalise to the test set though

Another question regarding the preprocessing section of the assignment, did you want us to comment on the quality of the data (e.g. in the tutes we checked for occurrences or 0's, NaNs and multicollinearity, etc

n the evaluation section, it states "Discussion of accuracy/validity should also consider the socio-economic nature of this data, and if or how this impacts the required accuracy and/or utility of the resultant models." What does this mean and what kind of information is required to discuss. And how does this differ from the last part about the ethical concerns of the socio-economic data.

Really we just want you to think about what sort of accuracy would be needed for this sort of model. When you report things like RMSE and R^2 rather than just saying "values of X, Y and Z were obtained, which shows the model is awesome/shithouse" consider a bit more if this sort of level of accuracy is appropriate given that such a model could be used to make funding and/or policy decisions.

You'd certainly expect to see worse RMSE on the training set becuase that's the whole idea of regularisation, trade off some accuracy on the training data for a simpler and hopefully better generalising mode.

Ideally performance improves on the testing set - but it may not. You should see better performance on the validation set though

hanks Simon. In that case should we only be explaining lambda for this dot point?

Are there any other parameters I should consider?

Screenshot 2024-03-23 at 2.29.33 pm.png

Lambda mostly.

Other details (like R^2, RMSE, etc) are more in the performance section. Similarly, p-values, etc, probably fit elsewehre.

"better" if often difficult to quantify, as it depends on what you want - maybe accuracy is all you care about, maybe you need a simple model, maybe not violating assumptions is the most important thing in your setting, maybe there's some other criteria.

All we want you to do is look at the various models and their various properties and comment on that. If they're all crap, tell us why; if they're all the same, tell us why. Just be sure to explain and justify what ever conclusion you reac

QUESTION 2

