**Name: Ben Karabinus**

**Question 1: Poisson/quasi-poisson/negative binomial model comprehension check**

1. How does quasipoisson Regression loosen the distributional assumption of the model residuals in the Poisson model? How does the negative binomial model do so?

Quasipoissoon Regression loosens the distribution assumptions of the Poisson Regression model by taking the variance of the data to be a constant multiple of the mean.

Negative Binomial Regression loosens the distribution assumptions of the Poisson Regression model by allowing the variance of count data to grow as a multiple of the mean of the data.

1. Comparing the Poisson regression model with the quasipoisson regression model, are the coefficient estimates noticeably different? What about the standard errors?

Comparing the Poisson and Quasipoisson regression models the coefficient estimates are not noticeably different, in fact in the example given in the asynch the coefficient estimates are identical.

When comparing the standard errors of the two models the there is a noticeable difference.

1. Where in the output would you look to see if there is evidence of overdispersion in the Poisson regression model? How could you evaluate if the quasipossion model or the negative binomial model improve it?

Assessing the output of the Poisson Regression model one could examine the residual deviance and degrees of freedom in the model output. To evaluate if the Quasipoisson Regression model or the Negative Binomial regression model are more effective at correcting for overdispersion one can examine the ratio of the residual deviance over degrees of freedom. A ratio closer to one is better.

**Question 2: Project check-in**

Now that you have your project topic, I want to check in with you about where you are in your work.

1. What is your project topic?

My project topic is home value prediction using gradient boosted machines. Currently I’m leaning towards using the XGBoost library.

1. What are the data requirements of your project topic? That is, what properties must your data have to use this method? You don’t need to have identified a data set to answer this question.

One of the key strengths of gradient boosting models is that they perform well on all types of data. Continuous data, categorical data, dirty data that hasn’t been processed prior to modeling. A step that is required for XGBoost is ensuring that predictors have been converted into a matrix of numerical values. Categorical predictors must be one hot encoded to fulfill this requirement. There are a few different ways to do this (i.e., model.matrix). I haven’t decided which method I will use for preparing my data yet. (there are a few cool methods that I’ve never tried, but might attempt to use)

1. What data set do you plan to use? If your data is publicly available, please provide a link to it. If you have not yet chosen a data set, what kind of data set do you want to obtain?

I plan to use a combined data set of estimated home values by region in the US (outcome) and socioeconomic factors (predictors).

The outcome is from the Zillow Home Value Index (ZHVI) available at: <https://www.zillow.com/research/data/>

The predictors will be pulled from Bureau of Economic Analysis: https://www.bea.gov/

1. What R package/s do you anticipate using for your method?

I’m leaning towards the xgboost package, but the gbm package is my second choice.

1. What resources are you using to learn more about your chosen method? Please list these resources here (e.g., links to websites, books, book chapters, links to videos, etc.). The reason I’m asking you this question is that I would like to build a list of resources for different methods to share; because you are a student, you are the best judge of what materials are most helpful!

University of Cincinnati Business Analytics R programming Guide (Awesome Resource!): <https://uc-r.github.io/gbm_regression#learn>

Datacamp: <https://www.datacamp.com/community/tutorials/xgboost-in-python>

GitHub Awesome XGBoost: <https://github.com/dmlc/xgboost/tree/master/demo>

Analytics Vidhya: <https://www.analyticsvidhya.com/blog/2016/03/complete-guide-parameter-tuning-xgboost-with-codes-python/>

Some examples are in Python. These examples are primarily for better overall understanding.