**MDS 561 Week 1 Hands-On Accelerator**

**(To Be Submitted By Email To Instructor) john.aaron@elmhurst.edu**

Your deliverables and hands-on activities for this week are:

1. TEAM ASSIGNMENT: Each team designates one individual to set up one a free trial Webex account (or an alternative). The team meets to discuss state selection and segregation of duties for their program work. As part of project chartering create a project description, scope statement and simple business case. Migrate the charter in the PRIMMS document repository.

Results: I have set up a WebEx account for our team project & have all agreed to meet on Monday at 6pm CST.

2. INDIVIDUAL ASSIGNMENT: Download and register the Salford Predictive Modeler (SPM 8.x) product using the instructions provided on Blackboard:

Results: I have successfully installed Salford Predictive Modeler.

3. INDIVIDUAL ASSIGNMENT Download install and use the Microsoft Bayesian Belief Network application MSBNX <https://msbnx.azurewebsites.net/> :

Results:I have successfully installed Microsoft Bayesian Belief Network.

1. INDIVIDUAL ASSIGNMENT Go to: <https://www.pmi.org/membership/student> and register to become a PMI member (for $32). Once registered go to the PMI Guides and Standards tab on the PMI website. Then download the following documents to your local computer:

* A Guide to the Project Management Body of Knowledge, Sixth Edition with Agile Practice Guide, 2017
* Navigating Complexity: A Practice Guide, 2014
* Managing Change in Organizations: A Practice Guide, 2013
* The Standard for Risk Management in Portfolios, Programs and Projects, 2019
* Standard for Portfolio Management, Fourth Edition, 2017

Feel free to download additional PMI documents if you wish.

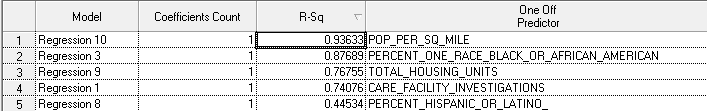
Results: I have successfully downloaded the above documents to my computer.

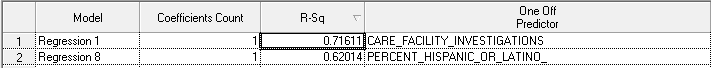
1. INDIVIDUAL ASSIGNMENT: Using the SPM regression engine and the automate feature, perform an analysis of the Wisconsin COVID County Dataset Starter to explain the following:
   1. Total Positives
   2. Total Positives Per 100,000 population
   3. Total Deaths
   4. Total Deaths Per 100,000 Population
   5. Variation across counties of key predictor variables
   6. Prediction confidence of COVID risks based upon model validation

Results:

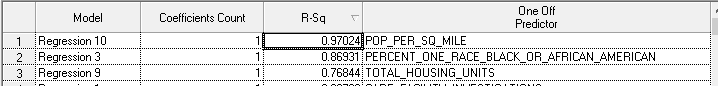
For the following exercise, I loaded in the Keep List on Blackboard. I set the test set split to 10%. Each model was ran with “ONEOFF” predictors to evaluate the impact of a single feature on the regression model for the target.

1. When the POSITIVE feature is the target variable, it appears that the POP\_PER\_SQ\_MILE, PRECENT\_ONE\_RACE\_BLACK\_OR\_AFRICAN\_AMERICAN, TOTAL\_HOUSING\_UNITS, and CARE\_FACILITY\_INVESTIGATIONS provide the most explanation on the POSITIVE feature. The next closest r-squared result is significantly lower.

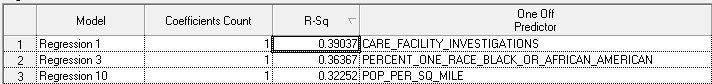


b. For POSITIVES\_PER\_100K, CARE\_FACILITY\_INVESTIGATIONS and PERCENT\_HISPANIC\_OR\_LATINO where the features that better explained positive cases per 100K population.

1. For DEATHS, POP\_PER\_SQ\_MILE, PERCENT\_ONE\_RACE\_BLACK\_OR\_AFRICAN\_AMERICAN, and TOTAL\_HOUSING\_UNITS are features that better explained the target (total deaths).

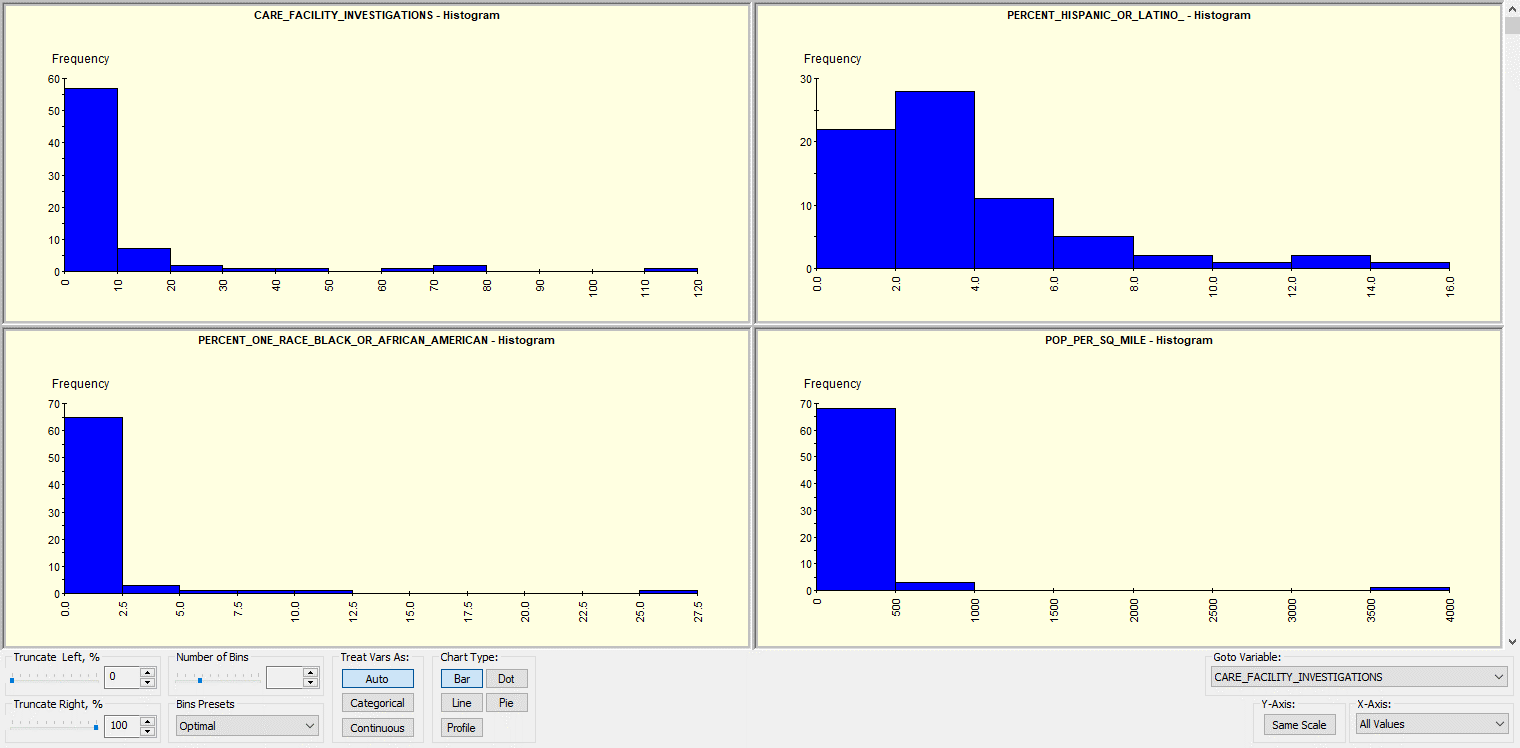


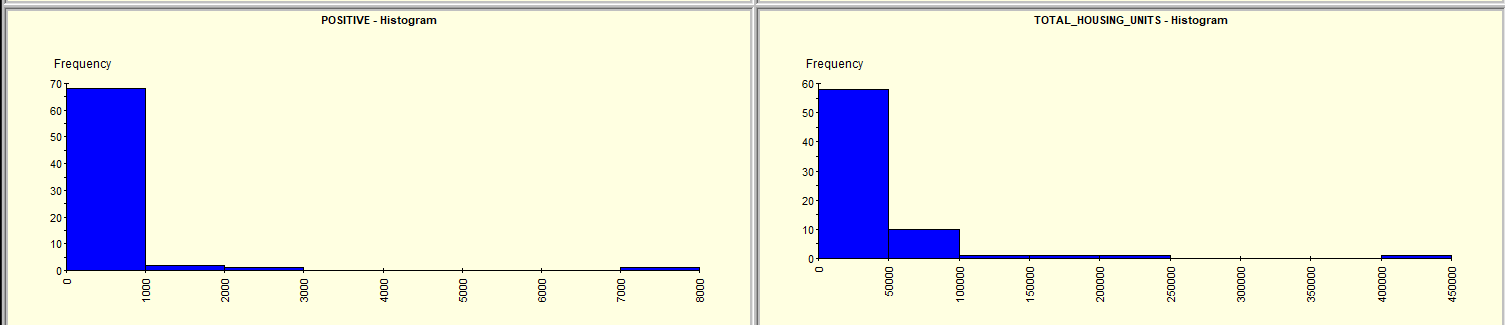
1. DEATHS\_PER\_100k as the target is mostly explained by the following features: CARE\_FACILITY\_INESTIGATIONS, PERCENT\_ONE\_RACE\_BLACK\_OR\_AFRICAN and POP\_PER\_SQ\_MILE. I think it is interesting to not that the R-squared value is significantly less then what we see in the previous targets. There’s definitely some other factors that contribute to Deaths per 100K or it has a low signal.



1. The following are the frequency distributions for the above most relevant features. From these graphs, you can see how the values of the features vary throughout the dataset.

You can see that the PERCENT\_HISPANIC\_OR\_LATINO has a greater variance among the counties then the other features.





1. I selected positive cases as my target variable and only included the 6 variables noted above. Since our dataset is so small, its not surprising that our model evaluation statistics significantly vary between the train and test set. As you can see the R-Sq for the train set is .95 while the test set shows -20.85.

