Dani DiTomasso
Professor Eleish
Data Analytics
11 October 2024

Assignment 2 (Lab 03) Written Responses

2.1 - Linear Model of EPI Results 2024

Most significant variable to EPI:

Based on the p-values of each of the 5 variables (found in chart when running a summary of lin.mod.epi2024), all variables are over a p-value of 0.05 meaning none of the p-values are statistically significant to EPI. However, the variable with the smallest p-value is WRS with a value of 0.363, meaning it is the most significant out of the 5 variables.

Most significant variable to Asia-Pacific Region:
Based on the p-values of each of the 5 variables (found in chart when running a summary of lin.mod.AP), all variables are over a p-value of 0.05 meaning none of the p-values are statistically significant to EPI of the Asia-Pacific region. However, the variable with the smallest p-value is WWT with a value of 0.4906, meaning it is the most significant out of the 5 variables

Which model is a better fit?

The EPI model is a better fit, due to the fact that the p-values for the 5 variables selected, although above the value of 0.05, are all closer to being statistically significant compared to the Asia-Pacific Region.

3.1 - kNN model for Global West, Eastern Europe, and Latin America & Caribbean

I believe that the second kNN model is better, due to the fact that there is less drastic fluctuation between points between k-values.

4.1 - k-means clustering for Former Soviet States, Sub-Saharan Africa, and Asia-Pacific (group 1) and Global West, Eastern Europe, and Latin America & Caribbean (group 2)

In both regional groups, 3.0 is the best value for ks since it is where the "elbow" of the graph is, and both groups have 3 different regions.