Module 2: Cluster Analysis

Conduct Hierarchical, K means, and 2 Step models

Part 1: SPSS output (10 marks):

Using **between group linkage** and **squared Euclidean distances** conduct the following:

Using the Ontario federal electoral district data or US 2020 census data (placed in d2l) *conduct a Hierarchical cluster model using 5 variables* of your choice (that are initially in absolute whole value form). How do the groups cluster when you build a model based on 1) unrated, 2) rated, 3) z score, and then 4) rated-z score data. For this part of the assignment, you only need to show dendograms and then explain (in written format) how the dendograms change as you progress from unrated, rated, z score, and then rated-score. On each dendogram show where the group breaks are located. If the dendogram has more than 8 groups, then show only 4 groups, based on the method shown in class (eg. North vs South, and then two categories with each North and South).

Using your final model, which is based on rated-z score data, and provide descriptors for each group on your dendogram and show these descriptors on your dendogram.

What happens to the Dendogram when you change Methods to **Nearest Neighbour** and Interval to **Block**?

Run a Hierarchical Model on 15 variables of your choice where the variables have been standardized using z scores. (Variables do not need to be rated). Which variables show to be most important on your dendogram. Why do you think they are the most important.

Part 2: Mapping Results

Map your cluster results for the rated-z score model (that was based on using **between group linkage** and **squared Euclidean distances). (5 marks**)

Part 3: Explain the Map (7.5 marks).

Explain the findings on your map and why the patterns on your map exist (5 marks).

Part 1: SPSS K-means analysis (10 marks):

-Employ the same 5 variables and region that you used above in a K means analysis. How do the clusters change from unrated, rated, z score, to rated-z score (based on 4 cluster groups)?

-Based on your final Hierachical Model, based on rated-z scores, how many groups should be used in K Means?

-Based on your final model (rated-z scores) What is the optimal number of groups to be used? Show how you determined this using 2 step cluster analyses? And show bar graph for your clusters.

-Fully articulate and discuss the SPSS output with explanations for your rated-z score model.

-How do your results differ from the hierarchical method?

Part 2: K Means Map (5 Marks)

Map your K means groups for your rated-z score model.

Part 3: K Means Map Discussion (7.5 Marks).

Fully discuss and explain the patterns founds on your map.