Kilger

MKT6971 2 credit course

Practicum I second semester

Exercise #5

Name: Collin Real (yhi267)

This fifth exercise is to give you practice at clustering market segments using GAP analysis

You can likely reuse a bunch of your code from exercise 4. So instead of doing a k means clustering, you are going to use PROC HPCLUS to do a gap analysis clustering.

1. Using the same drivers you did for your k means assignment, run your PROC HPCLUS using FIRSTPEAK as your criterion. How many clusters does it say is optimum? Cut and paste the ABC Statistics table as well as the FIRSTPEAK table below.

Optimum clusters = 2

A table with numbers and letters

Description automatically generated

A table of numbers with black text

Description automatically generated

1. Repeat step #2 except use GLOBALPEAK as your criterion. How many clusters does it say is optimum? Is it the same as Step #1 above? Cut and paste the ABC Statistics table as well as the GLOBALPEAK table below

Optimum clusters = 2

A table with numbers and letters

Description automatically generated

A table of numbers with black text

Description automatically generated

1. What number of clusters did you pick on the previous K means exercise? Is it the same as the HPCLUS suggested number of clusters? How might you decide which to use?

The number of clusters I picked from the previous exercise was k=5. The k means exercise’s suggestion is not the same as the HPCLUS suggestion. Although I agree with Tibshairani, Walther, and Hastie’s conclusion that the gap analysis is the better choice due to modern computing, I don’t think we should automatically assume that it’s best for every analysis. The diagnostic statistics for cubic clustering criteria and Pseudo F statistic as well as the gap analysis should be performed and compared for each analysis. We can compare the discrimination among the clusters for each method to influence our final decision.

1. Examine the cluster means for the drivers for the result either in Step #1 or Step #2 above. Do they look like there is decent discrimination among the clusters for the driver variables?

There are two instances where the gap does not meet the 0.1 threshold for driver variables disc\_clothes\_good and dress\_to\_please\_myself. Since the HPCLUS suggested 2 clusters, these driver variables would have to be removed and replaced with other driver variables for the gap analysis. However, the K means clustering algorithms in the prior exercise produced adequate differences between the means, so replacing the driver variables for those algorithms was not necessary.