DA6823

Exercise #5

Name: Moneeb Abu-Esba KVG805

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. **Be sure to cut and paste the tables here that help you answer the questions below.**

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.

Estimated Number of Clusters	
Criterion	Number of Clusters
FIRSTPEAK	7

ABC Statistics					
Number of Clusters	Logarithm of Within-Cluster SSE		Gap	Simulation Adjusted Standard Deviation	One Standard Error Adjusted Gap
	Input	Reference			
2	11.4523	13.0019	1.5496	0.00676	1.5429
3	11.2977	12.8002	1.5025	0.0101	1.4925
4	11.1473	12.6443	1.4970	0.00917	1.4879
5	11.0513	12.5402	1.4890	0.0110	1.4779
6	11.0809	12.3989	1.3180	0.0118	1.3062
7	10.9842	12.3136	1.3293	0.00652	1.3228
8	10.9311	12.2544	1.3233	0.00950	1.3138
9	10.8696	12.1663	1.2967	0.00565	1.2911

You did not answer the question. Cutting and pasting a table and expecting someone to have to d

2. 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

Estimated Number of Clusters	
Criterion	Number of Clusters
GLOBALPEAK	7

ABC Statistics

Number of Clusters	Logarithm of Within-Cluster SSE		Gap	Simulation Adjusted Standard Deviation	One Standard Error Adjusted Gap
	Input	Reference			
2	11.4523	13.0019	1.5496	0.00676	1.5429
3	11.2977	12.8002	1.5025	0.0101	1.4925
4	11.1473	12.6443	1.4970	0.00917	1.4879
5	11.0513	12.5402	1.4890	0.0110	1.4779
6	11.0809	12.3989	1.3180	0.0118	1.3062
7	10.9842	12.3136	1.3293	0.00652	1.3228
8	10.9311	12.2544	1.3233	0.00950	1.3138
9	10.8696	12.1663	1.2967	0.00565	1.2911

You did not answer the question. Cutting and pasting a table and expecting someone to have to do

3. 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?
   I chose K=5 while Firstpeak and Globalpeak chose K=7. The the local Maxima and the Gap analysis informs us where to stop I am choosing to use as my original answer as the One Standard Error Adjusted Gap - Gap = less than 0.1.

4. 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?
   The Values have large space apart and look to have decent discrimination.
5. Turn in your code, output and this report.

Your answer to #3 does not make any sense. How many clusters did you get in the k means exercise'

where is your cut and pasted table for cluster means for gap analysis? -4