DA6823

Exercise #4

Name: Moneeb Abu-Esba

This forth exercise is to give you practice at clustering market segments using k means clustering. **Be sure that you cut and paste the tables and plots that are requested in the questions below!!!**

You can reuse the code from exercise 3. You will need to add some extra code in the PROC FACTOR procedure to create a temporary SAS file that creates a temporary SAS system file that contains all of your original variables plus the two PCA factor variables that were created in PROC FACTOR. **Watch the lecture carefully to see how to do this.**

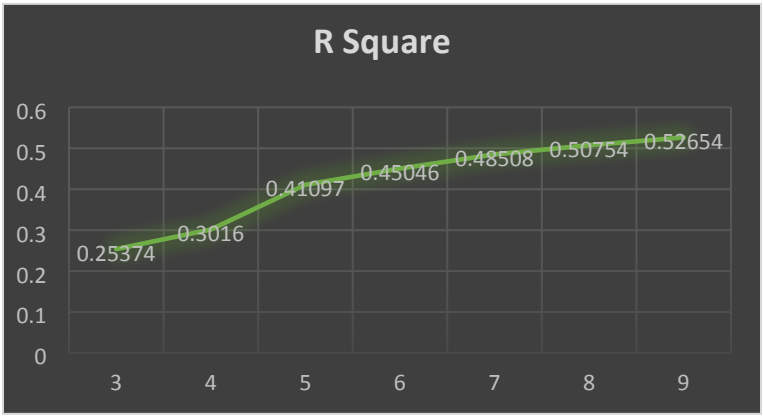
1. Run your k means cluster analysis using PROC CLUSTER using the temporary file just created above. Use all of your single driver variables plus the two factor variables you created in the PROC FACTOR procedure. Run it each time for k = 3 to k= 9 (that is 6 runs). Be sure to include these runs in your submission.

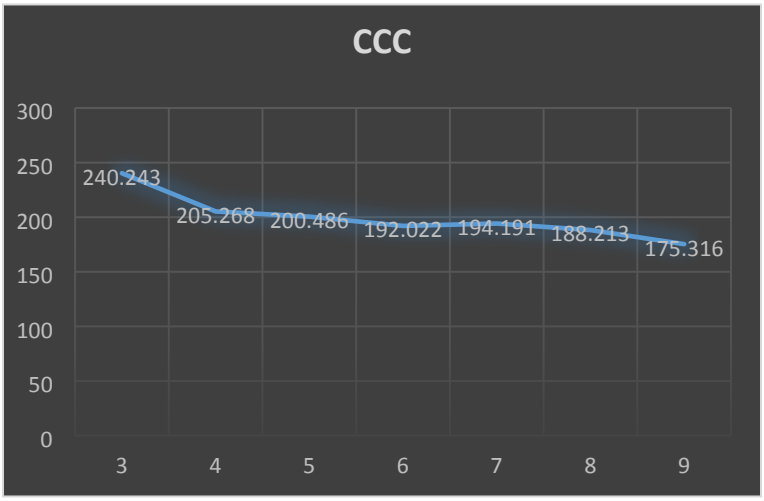
2. Build a table that contains the results for the k=3 to k=9 runs with values for the following statistics:

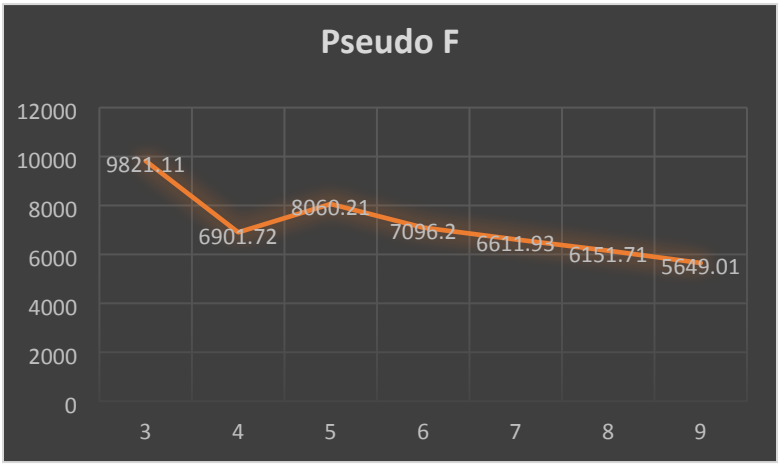
    a. Number of clusters
    b. R square
    c. CCC
    d. Pseudo F

Number of clusters	R Square	CCC	Pseudo F
3	0.25374	240.24	9821.11
4	0.3016	205.27	6901.72
5	0.41097	200.49	8060.21
6	0.45046	192.02	7096.2
7	0.48508	194.19	6611.93
8	0.50754	188.21	6151.71
9	0.52654	175.32	5649.01

3. Use whatever plotting software you like to create three different plots that the following against the number of clusters
    a. R square
    b. CCC
    c. Pseudo F







4. Examine the CCC plot and tell me what the rule is to suggest how many clusters to keep. How many does it suggest for this k means analysis. Note that it may not suggest any specific number of clusters. Tell me what you see!

   Based on the CCC plot my local maxima is when K=7 as from K=8 and k=9 the numbers are decreasing again. The first 5 clusters what about the first five clusters?

5. Examine the Pseudo F plot and tell me what the rule is to suggest how many clusters to keep. How many does it suggest for this k means analysis. Note that it may not suggest any specific number of clusters. Tell me what you see!

   The Pseudo F plot rises again at K=5 so I keep the first 3 clusters. what is this "keep"? business-

6. Examine the means of the driver variables for the solution that is suggested by the CCC plot. Do they look like this might be a good solution? Why or why not? Cut and paste the driver means table for the best solution. Note the criteria that I discuss in class for deciding if this is a decent solution or not.

   I chose Cluster six since it is my best cluster:
   Int_Change_Life: 0.402148419
   Comp_Peronal_Info: 0.146497201
   Pay_Any: 1.221337195
   Up_Tech: 3.621340431
   Gad: 1.584956183 whee are the cluster means for all drivers for all clusters?-6
   Gad_app: 1.756127154

   The Cluster Means in Cluster 6 has a difference of 0.1 or greater with some being more than 1.0. The values are spread out which shows a lot of discrimination.

7. Examine the means of the driver variables for the solution that is suggested by the Pseudo F plot. Do they look like this might be a good solution? Why or why not? Cut and paste the driver means for the best solution. Note the criteria that I discuss in class for deciding if this is a decent solution or not.

   Pseudo F suggest Cluster 5 as it is a local maxima. The Cluster means also look to be closer in value.

   what are the cluster means for all driver variables? where is the table -6

8. **If your solution above satisfies the criteria for a good solution, you are set. If neither the CCC or the Pseudo F plot suggest a best solution, try removing some of the drivers or adding one or two new drivers until you get something that satisfies the criteria for a decent solution. When you have that solution, then cut and paste this new table with number of clusters, R square, CCC and pseudo F below. Also cut and paste the driver means for this new solution below.**