Kilger

MKT6971 2 credit course

Practicum I second semester

**Note: There is more sample code on blackboard under exercise #6**

Exercise #6

Name: Collin Real (yhi267)

This sixth exercise is to give you practice at interpreting your descriptor variable cluster means across the clusters. These descriptor variables maybe binary in nature – e.g. drink coke =yes or no. Remember that you are looking for as much separation in means across clusters for each descriptor variable as possible. Also remember that no cluster solution is perfect and that some variables will have means across clusters that are close to each other. As long as there a number of variables that have decent separation across clusters (say .3 or so different) then that’s going to be a decent solution.

Use your k means code as a starting spot from exercise #4. Find the k=# cluster solution that you thought worked best for exercise #4. Then use the following code to output the cluster number for each case, **substituting your driver variables for the ones in the sample code**. Use the maxcluster=# of clusters you chose as best in exercise #4. Example if the best solution was 4 clusters then:

*All your previous exercise 4 code here then…*

proc fastclus data=clusready out=myclust maxclusters=4;

var

healthy

ecofriend

import\_attract\_opp\_sex\_scale

spend\_time\_family\_scale;

run;

Note the out=myclust which creates a temporary SAS data set called myclust. In that data set is all of your original data plus special variable called **CLUSTER**. That variable contains the cluster number (in this case a number from 1 to 4) that indicates which cluster the case or person belongs to.

Now you want to get the means for your descriptor variables by cluster. To do that first we need to sort the data by cluster number so that we can use the BY statement in PROC MEANS. Do this by placing code like this below after the fastclus code above. This will sort your data set by cluster number and output a new temporary data set mysort.

Proc sort data= myclust out=mysort;

By cluster;

Run;

And then you can produce means for your descriptor variables like follows:

Proc means data=mysort;

By cluster;

Var classic\_coke kfc\_chicken espn\_sports;

Run;

The BY statement tells SAS to group the means by cluster. Note that the means for binary variables such as classic\_coke can simply be interpreted as a proportion**.**

**One you have obtained the descriptor variable means by cluster then comment on well or not so well the clustering solution discriminates on that descriptor variable. Are the descriptor variable means close to together? Far apart from each other? Remember that farther apart is better. Tell me what you see. Also use the descriptor variables and briefly describe each of the clusters using the mean values from this exercise.**