**K-means clustering:** We are using SASHLEP.IRIS dataset for better understanding and easy results comparison purposes.

/\*Standardization\*/

**proc** **standard** data = sashelp.iris out = iris\_out mean = **0** std = **1**;

var SepalLength SepalWidth PetalLength PetalWidth;

**run**;

/\*Kmeans clustering method\*/

**proc** **fastclus** data = iris\_out out = kmeansout maxc = **3** noprint;

var SepalLength SepalWidth PetalLength PetalWidth;

**run**;

**proc** **freq** data = kmeansout;

tables Species\*Cluster/norow nocol nopercent;

**quit**;

**Hierarchical clustering:**

/\*Standardization\*/

**proc** **standard** data = sashelp.iris out = iris\_out mean = **0** std = **1**;

var SepalLength SepalWidth PetalLength PetalWidth;

**run**;

/\*Ward's minimum variance cluster method \*/

**proc** **cluster** data = iris\_out outtree = tree method = ward print = **15** ccc pseudo;

var SepalLength SepalWidth PetalLength PetalWidth;

copy Species;

**run**;

/\* After looking into CCC graph and R-square, 3 clusters seems ideal

select n= 3 in below code after choosing 3 clusters are optimum\*/

**proc** **tree** noprint out = wardsout n=**3**;

copy Species SepalLength SepalWidth PetalLength PetalWidth;

**run**;

**proc** **freq** data = wardsout;

tables Species\*Cluster/norow nocol nopercent;

**quit**;

**Elongated multinomial clustering:**

/\* Elongated Multinormal clusters \*/

**data** elongate;

keep x y;

ma=**8**; mb=**0**; link generate;

ma=**6**; mb=**8**; link generate;

stop;

generate:

do i=**1** to **50**;

a=rannor(**7**)\***6**+ma;

b=rannor(**7**)+mb;

x=a-b;

y=a+b;

output;

end;

return;

**proc** **fastclus** data=elongate out=out maxc=**2** noprint;

**run**;

%***modstyle***(name=ClusterStyle2,parent=Statistical,type=CLM,

markers=Circle Triangle circlefilled);

ods listing style=ClusterStyle;

**proc** **sgplot**;

scatter y=y x=x / group=cluster;

title ’FASTCLUS Analysis’;

title2 ’of Data Containing Parallel Elongated Clusters’;

**run**;

/\* Convert using canonical variables \*/

**proc** **aceclus** data=elongate out=ace p=**.1**;

var x y;

title ’ACECLUS Analysis’;

title2 ’of Data Containing Parallel Elongated Clusters’;

**run**;

**proc** **sgplot**;

scatter y=can2 x=can1;

title ’Data Containing Parallel Elongated Clusters’;

title2 ’After Transformation by PROC ACECLUS’;

**run**;

/\* Applying clustering on canonical variables \*/

**proc** **cluster** data=ace outtree=tree method=ward noprint;

var can1 can2;

copy x y;

**run**;

**proc** **tree** noprint out=out n=**2**;

copy x y;

**run**;

**proc** **sgplot**;

scatter y=y x=x / group=cluster;

title ’Ward’’s Minimum Variance Cluster Analysis’;

title2 ’of Data Containing Parallel Elongated Clusters’;

title3 ’After Transformation by PROC ACECLUS’;

**run**;