/\*25 Cluster run of FASTCLUS\*/

libname census "/folders/myfolders/census";

proc fastclus data=census.psam\_h17\_subset1

radius=0 replace=full

converge=0 maxiter=200

maxclusters=25

OUTSTAT=census.psam\_h17\_subset1\_25clusters\_stat

OUT=census.psam\_h17\_subset1\_25clusters

distance;

id SERIALNO;

var VALP HINCP;

run;

proc sgplot;

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

title ’ACS PUMS 2013 1 YR 25-Cluster Analysis’;

title2 ’of Data Containing Property Value and Household Income’;

run;

/\*Plot TAXP distribution across the clusters\*/

proc freq data=census.psam\_h17\_subset1\_25clusters;

tables TAXP\*Cluster;

run;

/\*Join additional variables from psam\_h17 table for profiling\*/

proc sql;

title '25 Cluster aggregated information for profiling';

create table census.psam\_h17\_subset1\_25clusters\_agg as

select a.SERIALNO, a.CLUSTER, a.DISTANCE, a.HINCP, a.VALP, a.TAXP,

b.BLD, b.TEN, b.YBL, b.WIF, b.MV, b.WORKSTAT

from census.psam\_h17\_subset1\_25clusters as a left join census.psam\_h17 as b

on a.SERIALNO=b.SERIALNO;

run;

%macro sqlloop;

proc sql;

/\*Loop through 25 times to create a table for each cluster\*/

%DO k=1 %TO 25;

create table census.psam\_h17\_subset1\_cluster\_agg&k. as

select \*

from census.psam\_h17\_subset1\_25clusters\_agg

where CLUSTER=&k.;

%END;

QUIT;

%mend;

%sqlloop;

/\*Get Min and Max VALP HINCP values by cluster\*/

proc sql;

title 'Mean, Min and Max VALP HINCP values by cluster';

create table census.psam\_h17\_subset1\_25clus\_min\_max as

select CLUSTER,

Mean(HINCP) as HINCP\_Mean, Min(HINCP) as HINCP\_Min, Max(HINCP) as HINCP\_Max,

Mean(VALP) as VALP\_Mean, Min(VALP) as VALP\_Min, Max (VALP) as VALP\_Max

from census.psam\_h17\_subset1\_25clusters\_agg group by CLUSTER;

run;

/\*sort the 25 clusters by population\*/

proc sort data=census.psam\_h17\_subset1\_25clusters\_stat;

by descending OVER\_ALL;

run;

/\*print the sorted data by cluster population\*/

proc print;

var CLUSTER VALP HINCP OVER\_ALL;

WHERE \_TYPE\_ = "FREQ";

title2 '25-cluster solution sorted by population';

run;

/\*plot the seed values per cluster\*/

proc sgplot;

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

WHERE \_TYPE\_ = "SEED";

title ’ACS PUMS 2013 1 YR 25-Cluster Analysis - Means’;

title2 ’of Data Containing Property Value and Household Income’;

run;

/\* Analyze categorical variables \*/

proc freq data=CENSUS.PSAM\_H17\_SUBSET1\_CLUSTER\_AGG3;

tables TAXP / plots=(freqplot);

run;