\* Gabriel Tellez;

\* Exam 1;

\*libname mylib '\\client\C$\Users\Jonathan\Documents\Classes\Practicum I\Exam1';

\*filename survey '\\client\C$\Users\Jonathan\Documents\Classes\Practicum I\Exam1\FA12\_Data.txt' lrecl = 65576;

libname mylib 'C:\Users\yyf335\Documents\Practicum I\Exam1';

filename survey 'C:\Users\yyf335\Documents\Practicum I\Exam1\FA12\_Data.txt' lrecl = **65576**;

\* CLEANING VARIABLES;

**data** variables;

infile survey;

\* sets of variables for PCA;

\* FIRST SET: Are BR shoppers loyal to BR?;

\* few favorite clothing brands to stick with;

\* always look for favorite brands first;

\* on several occasions have surprised myself by buying brands I usually don't;

\* usually shop at favorite stores because they have brands I like;

\* drawn to stores I normally don't shop at by sales;

\* SECOND SET: Do BR shoppers influene others in shopping habits?;

\* good at convincing others to try new things;

\* people often copy what I do or wear;

\* people come to me for advice before buying new things;

\* I prefer to shop with my friends;

\* friend's opinion of a store influence whether I shop there;

input respond\_id **1**-**7**

favstick\_agrlot **3316** favstick\_agrlit **3343** favstick\_neither **3397** favstick\_dislit **3424** favstick\_dislot **3451**

favfirst\_agrlot **3327** favfirst\_agrlit **3354** favfirst\_neither **3408** favfirst\_dislit **3435** favfirst\_dislot **3462**

surprise\_agrlot **3329** surprise\_agrlit **3356** surprise\_neither **3410** surprise\_dislit **3437** surprise\_dislot **3464**

favbrand\_agrlot **6238** favbrand\_agrlit **6280** favbrand\_neither **6364** favbrand\_dislit **6406** favbrand\_dislot **6448**

shopsale\_agrlot **6256** shopsale\_agrlit **6298** shopsale\_neither **6382** shopsale\_dislit **6424** shopsale\_dislot **6466**

convince\_agrlot **6551** convince\_agrlit **6578** convince\_neither **6632** convince\_dislit **6659** convince\_dislot **6686**

copywear\_agrlot **6555** copywear\_agrlit **6582** copywear\_neither **6636** copywear\_dislit **6663** copywear\_dislot **6690**

come4adv\_agrlot **7491** come4adv\_agrlit **7515** come4adv\_neither **7563** come4adv\_dislit **7587** come4adv\_dislot **7611**

shopfrnd\_agrlot **6262** shopfrnd\_agrlit **6304** shopfrnd\_neither **6388** shopfrnd\_dislit **6430** shopfrnd\_dislot **6472**

frndinfl\_agrlot **6241** frndinfl\_agrlit **6283** frndinfl\_neither **6367** frndinfl\_dislit **6409** frndinfl\_dislot **6451**

/\* RELEVANT VARIABLES (DRIVERS for cluster analysis): Do BR shoppers like sports (specifically college basketball)? \*/

/\* more likely to buy products from companies that sponsor sport teams/events \*/

sponsor\_agrlot **4529** sponsor\_agrlit **4610** sponsor\_neither **4772** sponsor\_dislit **4853** sponsor\_dislot **4934**

/\* try to keep abreast of changes in styles and fashion \*/

styfash\_agrlot **7488** styfash\_agrlit **7512** styfash\_neither **7560** styfash\_dislit **7584** styfash\_dislot **7608**

/\* Sports Interests - College Basketball \*/

collbball\_very **24073** collbball\_some **24074** collbball\_alit **24075** collbball\_none **24076**

/\* Sports Watched - Regular/Seasonal - College basketball in various networks \*/

/\* Viewed in last 12 months in ABC, CBS, ESPN, ESPN2, Other \*/

collbball\_abc **20277** collbball\_cbs **20278** collbball\_espn **20279** collbball\_espn2 **20280** collbball\_other **20283**

/\* Special Sports Programs & Events - Basketball \*/

/\* NCAA Men's Basketball Tournament (CBS/TBS/TNT/TRUTV) \*/

ncaatourney\_athome **21343** ncaatourney\_outside **21344** ncaatourney\_nexttime **21345**

/\* NCAA Men's Basketball Championship (CBS) \*/

ncaachamp\_athome **21346** ncaachamp\_outside **21347** ncaachamp\_nexttime **21348**

/\* DESCRIPTOR VARIABLES \*/

/\* education \*/

input edu\_noschool **2323** edu\_somegrade **2324** edu\_somehigh **2325** edu\_gradhigh **2326**

edu\_col\_less1 **2327** edu\_colyear1 **2328** edu\_colyear2 **2329** edu\_colyear3 **2330**

edu\_bachelor **2331** edu\_somegrad **2332** edu\_graddegree **2333**

/\* do some sport/exercise at least once a week \*/

dosport\_agrlot **4469** dosport\_agrlit **4550** dosport\_neither **4712** dosport\_dislit **4793** dosport\_dislot **4874**

/\* find TV advertising interesting and something to talk about \*/

tvadvrt\_agrlot **5652** tvadvrt\_agrlit **5695** tvadvrt\_neither **5781** tvadvrt\_dislit **5824** tvadvrt\_dislot **5867**;

**run**;

**data** missing;

set variables;

\* array to turn missing values to zeros;

array miss(**14**,**5**)

favstick\_agrlot favstick\_agrlit favstick\_neither favstick\_dislit favstick\_dislot

favfirst\_agrlot favfirst\_agrlit favfirst\_neither favfirst\_dislit favfirst\_dislot

surprise\_agrlot surprise\_agrlit surprise\_neither surprise\_dislit surprise\_dislot

favbrand\_agrlot favbrand\_agrlit favbrand\_neither favbrand\_dislit favbrand\_dislot

shopsale\_agrlot shopsale\_agrlit shopsale\_neither shopsale\_dislit shopsale\_dislot

convince\_agrlot convince\_agrlit convince\_neither convince\_dislit convince\_dislot

copywear\_agrlot copywear\_agrlit copywear\_neither copywear\_dislit copywear\_dislot

come4adv\_agrlot come4adv\_agrlit come4adv\_neither come4adv\_dislit come4adv\_dislot

shopfrnd\_agrlot shopfrnd\_agrlit shopfrnd\_neither shopfrnd\_dislit shopfrnd\_dislot

frndinfl\_agrlot frndinfl\_agrlit frndinfl\_neither frndinfl\_dislit frndinfl\_dislot

sponsor\_agrlot sponsor\_agrlit sponsor\_neither sponsor\_dislit sponsor\_dislot

styfash\_agrlot styfash\_agrlit styfash\_neither styfash\_dislit styfash\_dislot

dosport\_agrlot dosport\_agrlit dosport\_neither dosport\_dislit dosport\_dislot

tvadvrt\_agrlot tvadvrt\_agrlit tvadvrt\_neither tvadvrt\_dislit tvadvrt\_dislot;

/\* make missing values zeros\*/

do i = **1** to **14**;

do j = **1** to **5**;

if miss(i,j) = **.** then miss(i,j) = **0**;

end;

end;

/\* make array for 14 variable sums \*/

array mysum(**14**);

/\* sum up the vars and make no mark or > 1 mark missing\*/

/\* now make each variable, being sure to ignore zeroes and > 1 \*/

do k = **1** to **14**;

mysum(k) = miss(k,**1**) + miss(k,**2**) + miss(k,**3**) + miss(k,**4**) + miss(k,**5**) ;

end;

/\* now if the variable is not zero or > 1 create var \*/

array myvar(**14**);

do m = **1** to **14**;

if mysum(m) = **1** then

myvar(m) = (miss(m,**1**)\***5**) + (miss(m,**2**)\***4**)+ (miss(m,**3**)\***3**) + (miss(m,**4**))\***2** + (miss(m,**5**)\***1**);

else

myvar(m) = **.**;

end;

/\* now make the var names pretty again \*/

favstick = myvar(**1**);

favfirst = myvar(**2**);

surprise = myvar(**3**);

favbrand = myvar(**4**);

shopwear = myvar(**5**);

convince = myvar(**6**);

copywear = myvar(**7**);

come4adv = myvar(**8**);

shopfrnd = myvar(**9**);

frndinfl = myvar(**10**);

sponsor = myvar(**11**);

styfash = myvar(**12**);

dosport = myvar(**13**);

tvadvrt = myvar(**14**);

/\* make sum variables available \*/

mysum1=mysum(**1**);

mysum2=mysum(**2**);

mysum3=mysum(**3**);

mysum4=mysum(**4**);

mysum5=mysum(**5**);

mysum6=mysum(**6**);

mysum7=mysum(**7**);

mysum8=mysum(**8**);

mysum9=mysum(**9**);

mysum10=mysum(**10**);

mysum11=mysum(**11**);

mysum12=mysum(**12**);

mysum13=mysum(**13**);

mysum14=mysum(**14**);

/\* new variable CBB\_Interest with 4 levels\*/

CBB\_Interest = **.**;

if (collbball\_very = **1**) then CBB\_Interest = **4**;

if (collbball\_some = **1**) then CBB\_Interest = **3**;

if (collbball\_alit = **1**) then CBB\_Interest = **2**;

if (collbball\_none = **1**) then CBB\_Interest = **1**;

/\* new variable Education with 5 levels \*/

Education = **.**;

/\* GRADE SCHOOL \*/

if (edu\_noschool = **1**) or (edu\_somegrade = **1**) then Education = **1**;

/\* HIGH SCHOOL \*/

if (edu\_somehigh = **1**) or (edu\_gradhigh = **1**) then Education = **2**;

/\* SOME COLLEGE \*/

if (edu\_col\_less1 = **1**) or (edu\_colyear1 = **1**) or (edu\_colyear2 = **1**) or (edu\_colyear3 = **1**) then Education = **3**;

/\* COLLEGE \*/

if (edu\_bachelor = **1**) then Education = **4**;

/\* GRAD SCHOOL \*/

if (edu\_somegrad = **1**) or (edu\_graddegree = **1**) then Education = **5**;

**run**;

/\* check incidence of multiple check marks \*/

**proc** **freq** data=missing;

tables

mysum1 mysum2 mysum3 mysum4 mysum5 mysum6 mysum7

mysum8 mysum9 mysum10 mysum11 mysum12 mysum13 mysum14;

**run**;

/\* do freqs for binary vars \*/

**proc** **freq** data=missing;

tables

favstick\_agrlot favstick\_agrlit favstick\_neither favstick\_dislit favstick\_dislot

favfirst\_agrlot favfirst\_agrlit favfirst\_neither favfirst\_dislit favfirst\_dislot

surprise\_agrlot surprise\_agrlit surprise\_neither surprise\_dislit surprise\_dislot

favbrand\_agrlot favbrand\_agrlit favbrand\_neither favbrand\_dislit favbrand\_dislot

shopsale\_agrlot shopsale\_agrlit shopsale\_neither shopsale\_dislit shopsale\_dislot

convince\_agrlot convince\_agrlit convince\_neither convince\_dislit convince\_dislot

copywear\_agrlot copywear\_agrlit copywear\_neither copywear\_dislit copywear\_dislot

come4adv\_agrlot come4adv\_agrlit come4adv\_neither come4adv\_dislit come4adv\_dislot

shopfrnd\_agrlot shopfrnd\_agrlit shopfrnd\_neither shopfrnd\_dislit shopfrnd\_dislot

frndinfl\_agrlot frndinfl\_agrlit frndinfl\_neither frndinfl\_dislit frndinfl\_dislot

sponsor\_agrlot sponsor\_agrlit sponsor\_neither sponsor\_dislit sponsor\_dislot

styfash\_agrlot styfash\_agrlit styfash\_neither styfash\_dislit styfash\_dislot

dosport\_agrlot dosport\_agrlit dosport\_neither dosport\_dislit dosport\_dislot

tvadvrt\_agrlot tvadvrt\_agrlit tvadvrt\_neither tvadvrt\_dislit tvadvrt\_dislot;

**run**;

/\* now sniff new variables and compare - should be approx same as cell counts \*/

**proc** **freq** data=missing;

tables favstick favfirst surprise favbrand shopwear convince copywear

come4adv shopfrnd frndinfl sponsor styfash dosport tvadvrt Education CBB\_Interest;

**run**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

\* IMPUTING MISSING OBSERVATIONS;

\* determine the pattern of missing data using NIMPUTE=0;

**proc** **mi** data=missing nimpute=**0**;

var favstick favfirst surprise favbrand shopwear convince copywear

come4adv shopfrnd frndinfl sponsor styfash dosport tvadvrt CBB\_Interest;

**run**;

\* pattern of missing data is ARBITRARY;

\* variables are classification ordinal variables;

\* available method with MI procedure = FCS Logistic Regression;

\* Single Imputation;

**proc** **mi** data=missing nimpute=**1** out=mi\_fcs seed=**54321**;

class favstick favfirst surprise favbrand shopwear convince copywear

come4adv shopfrnd frndinfl sponsor styfash dosport tvadvrt CBB\_Interest;

/\*fcs plots=trace(mean std) \*/

var favstick favfirst surprise favbrand shopwear convince copywear

come4adv shopfrnd frndinfl sponsor styfash dosport tvadvrt CBB\_Interest;

fcs logistic(favstick favfirst surprise favbrand shopwear convince copywear

come4adv shopfrnd frndinfl sponsor styfash dosport tvadvrt CBB\_Interest) nbiter=**1**;

**run**;

\* Analysis phase;

\* 6 iterations produced lowest AIC;

\*proc genmod data=mi\_fcs;

\* class favstick favfirst surprise favbrand shopwear convince copywear

come4adv shopfrnd frndinfl sponsor styfash dosport tvadvrt;

\* model favstick = favfirst surprise favbrand shopwear convince copywear

come4adv shopfrnd frndinfl sponsor styfash dosport tvadvrt;

\* by \_imputation\_;

\* ods output ParameterEstimates=gm\_fcs;

\*run;

\* Pooling phase;

\*proc mianalyze parms(classvar=level)=gm\_fcs;

\* class favfirst surprise favbrand shopwear convince copywear

come4adv shopfrnd frndinfl sponsor styfash dosport tvadvrt;

\* modeleffects intercept1 favfirst surprise favbrand shopwear convince copywear

come4adv shopfrnd frndinfl sponsor styfash dosport tvadvrt;

\*run;

/\* shows all 25564 observations, 202 of them have no answers for 15 variables used \*/

**proc** **freq** data=mi\_fcs;

tables favstick favfirst surprise favbrand shopwear convince copywear

come4adv shopfrnd frndinfl sponsor styfash dosport tvadvrt CBB\_Interest;

**run**;

/\* delete rows where 15 variables are missing to get 25362 observations to work with \*/

/\* excluded Education since it has no missing observations \*/

**data** cleaned;

set mi\_fcs;

if nmiss(favstick,favfirst,surprise,favbrand,shopwear,convince,copywear,

come4adv,shopfrnd,frndinfl,sponsor,styfash,dosport,tvadvrt,CBB\_Interest) = **15** then delete;

**run**;

/\* shows the 25362 observations \*/

**proc** **freq** data=cleaned;

tables favstick favfirst surprise favbrand shopwear convince copywear come4adv

shopfrnd frndinfl sponsor styfash dosport tvadvrt CBB\_Interest Education;

**run**;

/\* save the two data files out \*/

**data** mylib.cleaned\_vars;

set cleaned;

**run**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*//\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*//\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*//\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

\* FACTOR ANALYSIS;

**data** factor\_analysis;

set mylib.cleaned\_vars;

/\* do the factor analysis - set eigenval rotate=varimax scree PCA \*/

**proc** **factor** data=factor\_analysis maxiter=**100** method=principal mineigen=**1** rotate=varimax

scree score print nfactors=**2** out=myscores;

var favstick favfirst surprise favbrand shopwear

convince copywear come4adv shopfrnd frndinfl;

**run**;

/\* rename the factor variables \*/

**data** factor\_names;

set myscores;

rename factor1 = influence;

rename factor2 = loyalty;

rename my\_id = resp\_id;

**run**;

**data** mylib.clustering;

set factor\_names;

**run**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*//\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*//\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*//\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

\* CLUSTER ANALYSIS;

**data** clusters;

set mylib.clustering;

/\* standardizing the variables \*/

**proc** **standard** data=clusters out=standard mean=**0** std=**1**;

var sponsor styfash dosport tvadvrt CBB\_Interest Education;

**run**;

/\* k-means cluster analysis \*/

/\* strict=3.0 prevents outliers from distorting the results \*/

**proc** **fastclus** data=standard maxclusters=**2** strict=**3.0**;

var influence sponsor styfash tvadvrt CBB\_Interest;

**run**;

**proc** **fastclus** data=standard maxclusters=**3** strict=**3.0**;

var influence sponsor styfash tvadvrt CBB\_Interest;

**run**;

**proc** **fastclus** data=standard maxclusters=**4** strict=**3.0**;

var influence sponsor styfash tvadvrt CBB\_Interest;

**run**;

**proc** **fastclus** data=standard maxclusters=**5** strict=**3.0**;

var influence sponsor styfash tvadvrt CBB\_Interest;

**run**;

**proc** **fastclus** data=standard maxclusters=**6** strict=**3.0**;

var influence sponsor styfash tvadvrt CBB\_Interest;

**run**;

**proc** **fastclus** data=standard maxclusters=**7** strict=**3.0**;

var influence sponsor styfash tvadvrt CBB\_Interest;

**run**;

**proc** **fastclus** data=standard maxclusters=**8** strict=**3.0**;

var influence sponsor styfash tvadvrt CBB\_Interest;

**run**;

/\* gap analysis \*/

**proc** **hpclus** data=standard maxclusters=**8** noc=abc(b=**20** minclusters=**2** align=pca criterion=firstpeak);

input influence sponsor styfash tvadvrt CBB\_Interest / level=interval;

score output = varscore;

**run**;

**proc** **contents** data=varscore;

**run**;

**proc** **freq** data=varscore;

tables \_cluster\_id\_;

**run**;

/\* unstandardized DRIVER variables for table of means \*/

/\* using diagnostic statistics, 5 clusters \*/

/\* took out factor that was used INFLUENCE \*/

**proc** **fastclus** data=clusters maxclusters=**5** strict=**3.0**;

var sponsor styfash tvadvrt CBB\_Interest;

**run**;

**proc** **hpclus** data=clusters maxclusters=**8** noc=abc(b=**20** minclusters=**2** align=pca criterion=firstpeak);

input dosport education / level=interval;

**run**;

**proc** **fastclus** data=clusters maxclusters=**5** strict=**3.0**;

var dosport education;

**run**;