/\* Creating library \*/

libname epi749 libname epi749 "C:\Users\bk658\Desktop\ILE\week3";

";

/\* Importing datasets \*/

libname demo XPORT "C:\Users\bk658\Desktop\ILE\week 3\DEMO\_I.XPT";

libname inc XPORT " C:\Users\bk658\Desktop\ILE\week 3\INQ\_I.XPT";

libname dep XPORT " C:\Users\bk658\Desktop\ILE\week 3\DPQ\_I.XPT";

libname emp XPORT " C:\Users\bk658\Desktop\ILE\week 3\ OCQ\_I.XPT";

libname home XPORT " C:\Users\bk658\Desktop\ILE\week 3\ HOQ\_I.XPT";

/\* Copying into NHANES library \*/

proc copy in = demo out = epi749; run;

proc copy in = inc out = epi749; run;

proc copy in = dep out = epi749; run;

proc copy in = emp out = epi749; run;

proc copy in = home out = epi749; run;

/\* Outputting names and dimensions of variables within each dataset \*/

proc contents data=epi749.\_ALL\_;

run;

/\* Sort datasets to merge \*/

proc sort data = epi749.demo\_i; by SEQN; RUN;

proc sort data = epi749.dpq\_i; by SEQN; RUN;

proc sort data = epi749.inq\_i; by SEQN; RUN;

proc sort data = epi749.ocq\_i; by SEQN; RUN;

proc sort data = epi749.hoq\_i; by SEQN; RUN;

/\* Merging data \*/

data epi749.comb;

MERGE epi749.demo\_i epi749.dpq\_i epi749.inq\_i epi749.ocq\_i epi749.hoq\_i;

BY SEQN;

run;

proc contents data=epi749.comb;

run; \*9971 total;

/\* Determing inclusion/exclusion \*/

data epi749.comb;

set epi749.comb;

if RIDSTATR ne 2 then exam\_exclude = 1; else exam\_exclude = 0; \*create interview/exam category;

if RIDAGEYR lt 20 then age\_exclude = 1; else age\_exclude = 0; \*create age category;

run;

proc freq data=epi749.comb;

tables age\_exclude\*exam\_exclude;

run;

/\* Creating permanent dataset based on inclusion/exclusion criteria \*/

data epi749.final;

set epi749.comb;

if exam\_exclude = 1 or age\_exclude = 1 then delete;

run;

**data** epi749.NHANES\_Lab4;

set epi749.final;

/\*

my DAG identified my minimally sufficient adjustment set to be:

age, education, employment status, race/ethnicity, residency, sex.

residency does not exist in NHANES, I will use citizenship instead

\*/

KEEP SEQN RIDSTATR RIAGENDR RIDAGEYR RIDRETH3 INDHHIN2 HOQ065 INQ300 IND310

DMDEDUC2 DMDCITZN OCD150 DPQ010 DPQ020 DPQ030 DPQ040 DPQ050 DPQ060 DPQ070

DPQ080 DPQ090 SDMVSTRA SDMVPSU WTMEC2YR;

**run**;

**proc** **contents** data=epi749.NHANES\_Lab4;

**run**;

/\* -- Preliminary data exploration & management -- \*/

**data** epi749.nhanes\_lab4;

set epi749.nhanes\_lab4;

/\* Missing data \*/

\*Set refused/don't know to missing;

if DMDEDUC2 GE **7** then call missing(DMDEDUC2);

if DMDCITZN GE **7** then call missing(DMDCITZN);

if HOQ065 GE **7** then call missing(HOQ065);

if INQ300 GE **7** then call missing(INQ300);

if OCD150 GE **7** then call missing(OCD150);

if INDHHIN2 GE **77** then call missing(INDHHIN2);

\*For all variable names starting with "dpq";

array \_dpq dpq:;

do over \_dpq;

if (\_dpq >=**7**) then

call missing(\_dpq);

end;

/\* Depression scale \*/

\*Create depression score (score will be missing if \*any\* of the items are missing);

dep\_score\_v1 = dpq010+dpq020+dpq030+dpq040+dpq050+dpq060+dpq070+dpq080+dpq090;

\*Create dichotomous depression variable;

if missing(dep\_score\_v1) then call missing(dep\_bin\_v1);

else if dep\_score\_v1 GE **5** then dep\_bin\_v1=**1**;

else dep\_bin\_v1=**0**;

/\* Wealth \*/

if missing(HOQ065) AND missing(INQ300) then call missing(wealth);

else if HOQ065 = **1** OR INQ300 = **1** then wealth = **1**;

else wealth = **0**;

/\* Age (checked - no missing)\*/

if RIDAGEYR < **40** then age\_cat = **1**;

else if RIDAGEYR < **60** then age\_cat = **2**;

else age\_cat = **3**;

/\* Race/ethnicity (checked - no missing) \*/

if RIDRETH3 < **3** then race = **1**;

else if RIDRETH3 = **3** then race = **2**;

else if RIDRETH3 = **4** then race = **3**;

else race = **4**;

/\* Income \*/

if missing(INDHHIN2) then call missing(income\_cat);

else if INDHHIN2 < **5** then income\_cat = **1**;

else income\_cat = **2**;

/\* Education \*/

if missing(DMDEDUC2) then call missing(educ\_cat);

else if DMDEDUC2 < **2** then educ\_cat = **1**;

else if DMDEDUC2 = **3** then educ\_cat = **2**;

else if DMDEDUC2 = **4** then educ\_cat = **3**;

else educ\_cat = **4**;

/\* Occupation \*/

if missing(OCD150) then call missing (job\_cat);

else if OCD150 = **1** then job\_cat = **1**;

else job\_cat = **2**;

**run**;

/\* -- Formats -- \*/

**proc** **format**;

value depress

**1** = "Depressed"

**0** = "Not Depressed";

value wealth

**1** = "High Wealth"

**0** = "Low Wealth";

value age

**1** = "20-39"

**2** = "40-59"

**3** = "60+";

value race

**1** = "Hispanic"

**2** = "NH White"

**3** = "NH Black"

**4** = "NH Other";

value income

**1** = "<$20k"

**2** = "$20k+";

value educ

**1** = "Some HS"

**2** = "HS Grad"

**3** = "Some College"

**4** = "College Grad";

value sex

**1** = "Male"

**2** = "Female";

value citizen

**1** = "US Citizen"

**2** = "Not US Citizen";

value job

**1** = "Worked last week"

**2** = "Did not work";

**run**;

/\* -- Summary statistics -- \*/

\*Depression score;

**proc** **univariate** data=epi749.nhanes\_lab4;

var dep\_score\_v1;

hist;

**run**;

\*Depressed (y/n);

**proc** **freq** data=epi749.nhanes\_lab4;

table dep\_bin\_v1 /missing;

format dep\_score\_v1 depress.;

**run**;

\*Wealth variable;

**proc** **freq** data=epi749.nhanes\_lab4;

tables HOQ065 INQ300 wealth /missing;

tables HOQ065\*INQ300 /missing;

format wealth wealth.;

**run**; \*3% missing for home ownership, 8% missing for savings;

\*Age;

**proc** **freq** data=epi749.nhanes\_lab4;

tables age\_cat /missing;

format age\_cat age.;

**run**;

\*Race/Ethnicity;

**proc** **freq** data=epi749.nhanes\_lab4;

tables race /missing;

format race race.;

**run**;

\*Sex;

**proc** **freq** data=epi749.nhanes\_lab4;

tables RIAGENDR /missing;

format RIAGENDR sex.;

**run**;

\*Income;

**proc** **freq** data=epi749.nhanes\_lab4;

tables income\_cat /missing;

format income\_cat income.;

**run**;

\*Education;

**proc** **freq** data=epi749.nhanes\_lab4;

tables educ\_cat /missing;

format educ\_cat educ.;

**run**;

\*Employment;

**proc** **freq** data=epi749.nhanes\_lab4;

tables job\_cat /missing;

format job\_cat job.;

**run**;

\*Citizenship;

**proc** **freq** data=epi749.nhanes\_lab4;

tables DMDCITZN /missing;

format DMDCITZN citizen.;

**run**;

\*Weighted univariate statistics;

**proc** **surveymeans** data=epi749.NHANES\_Lab4;

strata SDMVSTRA;

cluster SDMVPSU;

weight WTMEC2YR;

var dep\_score\_v1;

format dep\_score\_v1 depress.;

**run**;

**proc** **surveyfreq** data=epi749.nhanes\_lab4;

strata SDMVSTRA;

cluster SDMVPSU;

weight WTMEC2YR;

table dep\_bin\_v1 wealth age\_cat race RIAGENDR income\_cat educ\_cat job\_cat DMDCITZN;

format dep\_bin\_v1 depress. wealth wealth. age\_cat age. race race. RIAGENDR sex.

income\_cat income. educ\_cat educ. job\_cat job. DMDCITZN citizen.;

**run**;

/\* -- Creating new dataset from dataset created earlier-- \*/

**data** epi749.NHANES\_Lab5;

set epi749.NHANES\_Lab4;

**run**;

**proc** **contents** data=epi749.nhanes\_lab5;

**run**;

**proc** **sort** data=epi749.nhanes\_lab5;

by wealth;

**run**;

/\* -- TABLE 1 -- \*/

/\* - Weighted Estimates - \*/

\*Unweighted continuous variables;

**proc** **univariate** data=epi749.nhanes\_lab5;

var RIDAGEYR dep\_score\_v1;

**run**;

**proc** **ttest** data=epi749.NHANES\_Lab5;

class wealth;

var RIDAGEYR dep\_score\_v1;

format wealth wealth.;

**run**;

\*Unweighted categorical variables;

**proc** **freq** data=epi749.nhanes\_lab5 ;

table age\_cat\*wealth RIAGENDR\*wealth race\*wealth income\_cat\*wealth educ\_cat\*wealth job\_cat\*wealth DMDCITZN\*wealth dep\_bin\_v1\*wealth / chisq;

format wealth wealth. age\_cat age. RIAGENDR sex. race race. income\_cat income. educ\_cat educ. job\_cat job. DMDCITZN citizen. dep\_bin\_v1 depress.;

**run**;

/\* - Weighted Estimates - \*/

\*Continuous variables;

**proc** **surveymeans** data=epi749.nhanes\_lab5 all;

by wealth;

var RIDAGEYR dep\_score\_v1;

class wealth;

strata SDMVSTRA;

cluster SDMVPSU;

weight WTMEC2YR;

format wealth wealth.;

**run**;

\*Categorical variables;

**proc** **surveyfreq** data=epi749.nhanes\_lab4;

tables age\_cat\*wealth RIAGENDR\*wealth race\*wealth income\_cat\*wealth educ\_cat\*wealth job\_cat\*wealth DMDCITZN\*wealth dep\_bin\_v1\*wealth /chisq col;

strata SDMVSTRA;

cluster SDMVPSU;

weight WTMEC2YR;

format wealth wealth. age\_cat age. RIAGENDR sex. race race. income\_cat income. educ\_cat educ. job\_cat job. DMDCITZN citizen. dep\_bin\_v1 depress.;

**run**;