# **/\*The SAS program (HEI-2015 at the population/group level using NHANES 2011-2012 data (and FPED))**

# **Population Ratio Method.SAS\*/**

**/\*INSTRUCTIONS – complete tasks 1-4 in this section, and run these SAS codes before proceeding to the HEI-2015 scoring program that follows\*/**

/\*1. Create a folder on your computer “home folder”, and save the FPED data, NHANES data, Demographic data, and the required HEI-2015 macro in it. Specify the path to the folder. \*/

%let home = C:\Users\Documents\FPED\_NHANES; /\*In this Example, the “home” folder is in C Drive, within Documents, and is called FPED\_NHANES. \*/

/\*2. Libnames here specify the input files. \*/

libname NH “&home\NH”;

libname FPED “&home\FPED”; /\*In this Example, the FPED data are in a folder called “FPED”, and the NHANES and Demographic data are in a folder called “NH”, all saved within the “home” folder. These are SAS datasets. \*/

/\*3. Create a folder in the "Home" folder, where the output file, containing HEI-2015 component and total scores are to be exported. Specify the name of the folder. \*/

filename RES “&home\RES”; /\*In this Example, the folder is called “RES”, within the “home” folder, and the exported results will be a csv file called “result”. \*/

/\*4. Read in required HEI-2015 scoring macro. This macro must be saved within the home folder. \*/

%include ““&home\hei2015.score.macro.sas”;

/\*NOTE: Once you have completed all the steps above, all you need to do is run the SAS program below. Unless you used different names for your datasets and folders, no other action is required from you. \*/

title 'HEI-2015 scores for NHANES 2011-2012 day 1, AGE >= 2, RELIABLE DIETS, Include Pregnant and Lactating Women';

/\*Section (I): Calculations at the individual participant level to obtain variables needed to calculate HEI-2015 scores.\*/

/\*Step 1: locate the required datasets and variables \*/

\*part a: get FPED data per day;

**data** FPED;

set FPED.fped\_dr1tot\_1112;

**run**;

\*part b: get individual nutrient food intake if reliable recall status;

**data** NUTRIENT (keep=SEQN WTDRD1 DR1TKCAL DR1TSFAT DR1TALCO DR1TSODI DR1DRSTZ DR1TMFAT DR1TPFAT);

set NH.DR1TOT\_G;

if DR1DRSTZ=**1**; /\*reliable dietary recall status\*/

**run**;

\*part c: get demographic data for persons aged two and older;

**data** DEMO (keep=SEQN RIDAGEYR RIAGENDR SDDSRVYR SDMVPSU SDMVSTRA);

set NH.DEMO\_G;

if RIDAGEYR >= **2**;

**run**;

/\*Step 2: Combine the required datasets\*/

**proc** **sort** data=FPED;

by SEQN;

**run**;

**proc** **sort** data=NUTRIENT;

by SEQN;

**run**;

**proc** **sort** data=DEMO;

by SEQN;

**run**;

**data** COHORT;

merge NUTRIENT (in=N) DEMO (in=D) FPED;

by SEQN;

if N and D;

**run**;

/\*Step 3: Creates additional required variables: FWHOLEFRT, MONOPOLY, VTOTALLEG, VDRKGRLEG, PFALLPROTLEG and PFSEAPLANTLEG \*/

**data** COHORT;

set COHORT;

by SEQN;

FWHOLEFRT=DR1T\_F\_CITMLB+DR1T\_F\_OTHER;

MONOPOLY=DR1TMFAT+DR1TPFAT;

VTOTALLEG=DR1T\_V\_TOTAL+DR1T\_V\_LEGUMES;

VDRKGRLEG=DR1T\_V\_DRKGR+DR1T\_V\_LEGUMES;

PFALLPROTLEG=DR1T\_PF\_MPS\_TOTAL+DR1T\_PF\_EGGS+DR1T\_PF\_NUTSDS+DR1T\_PF\_SOY+DR1T\_PF\_LEGUMES;

PFSEAPLANTLEG=DR1T\_PF\_SEAFD\_HI+DR1T\_PF\_SEAFD\_LOW+DR1T\_PF\_NUTSDS+DR1T\_PF\_SOY+DR1T\_PF\_LEGUMES;

**run**;

/\*Section (II): Calculation of weighted means and a variance-covariance matrix and generation of a Monte Carlo

dataset, enabling standard errors to be calculated.\*/

/\*Step 1. Calculate the weighted means and the variance/covariance matrix for the dietary variables of interest. \*/

**data** ONE;

set COHORT;

array comp (**14**) DR1TKCAL VTOTALLEG VDRKGRLEG DR1T\_F\_TOTAL FWHOLEFRT DR1T\_G\_WHOLE DR1T\_D\_TOTAL

PFALLPROTLEG PFSEAPLANTLEG MONOPOLY DR1TSFAT DR1TSODI DR1T\_G\_REFINED DR1T\_ADD\_SUGARS;

\* Turn each variable into an observation of the single variable VBL ;

\* keeping track of the order with the dum\_num variable;

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

VBL = comp(i);

dum\_num = i;

output;

end;

**run**;

**data** ONE;

set ONE;

\* Create dummies with the same names as the original variables;

\* The i-th dummy gets a 1 if the observation is associated with the

\* i-th original variable;

array comp (**14**) DR1TKCAL VTOTALLEG VDRKGRLEG DR1T\_F\_TOTAL FWHOLEFRT DR1T\_G\_WHOLE DR1T\_D\_TOTAL

PFALLPROTLEG PFSEAPLANTLEG MONOPOLY DR1TSFAT DR1TSODI DR1T\_G\_REFINED DR1T\_ADD\_SUGARS;

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

if dum\_num = i then comp(i) = **1**;

else comp(i) = **0**;

end;

drop i dum\_num;

**run**;

\* Now run PROC (SURVEY)REG with dependent variable VBL and all the dummies;

\* as predictors. Force the regression to not include an intercept, and ask;

\* for the covariance matrix of the parameter estimates. ;

**proc** **surveyreg** data=ONE;

strata SDMVSTRA;

cluster SDMVPSU;

weight WTDRD1;

model VBL= DR1TKCAL VTOTALLEG VDRKGRLEG DR1T\_F\_TOTAL FWHOLEFRT DR1T\_G\_WHOLE DR1T\_D\_TOTAL

PFALLPROTLEG PFSEAPLANTLEG MONOPOLY DR1TSFAT DR1TSODI DR1T\_G\_REFINED DR1T\_ADD\_SUGARS/noint covb;

\* Output the covariance matrix we wanted all along;

ods output COVB=CSD\_COV;

title2 "Tricking SURVEYREG into giving us the covariance matrix of means";

**run**;

\* this proc print is for verification and is not necessary;

\* uncomment next three lines if more information is desired;

**proc** **print** data= CSD\_COV;

title2 "Printout of csd\_cov dataset -uses complex survey info";

**run**;

\*If curious about results then remove the noprint option and more ;

\* output will be displayed;

**proc** **means** data= COHORT n min max mean ; \*noprint;

weight WTDRD1;

var DR1TKCAL VTOTALLEG VDRKGRLEG DR1T\_F\_TOTAL FWHOLEFRT DR1T\_G\_WHOLE DR1T\_D\_TOTAL

PFALLPROTLEG PFSEAPLANTLEG MONOPOLY DR1TSFAT DR1TSODI DR1T\_G\_REFINED DR1T\_ADD\_SUGARS;

title2 'look at weighted means';

output out=WTDM mean= ;

**run**;

**data** COVDATA (drop=Parameter);

set CSD\_COV;

\_TYPE\_='COV ';

\_NAME\_=Parameter;

**run**;

**data** WTDM (drop=\_TYPE\_ \_FREQ\_);

set WTDM;

**run**;

**data** WTDM;

set WTDM;

\_TYPE\_='MEAN';

**run**;

**data** COVDATA;

set COVDATA WTDM;

**run**;

\* this proc print is for verification and is not necessary;

\* uncomment next three lines if more information is desired;

**proc** **print** data=COVDATA;

title2 'input to simnorml';

**run**;

/\*Step 2. In this step, a Monte Carlo data set with 10,000 rows is generated using the means

and variance/covariance matrix from step 1\*/

\*seed value may be changed using a random number generator;

**proc** **simnormal** data=COVDATA(type=cov) numreal=**10000** seed=**51230077** outseed out=SIM\_DATA;

var DR1TKCAL VTOTALLEG VDRKGRLEG DR1T\_F\_TOTAL FWHOLEFRT DR1T\_G\_WHOLE DR1T\_D\_TOTAL

PFALLPROTLEG PFSEAPLANTLEG MONOPOLY DR1TSFAT DR1TSODI DR1T\_G\_REFINED DR1T\_ADD\_SUGARS;

**run**;

\* this proc means is for verification and is not necessary;

\* uncomment next four lines if more information is desired;

**proc** **means** data=SIM\_DATA n nmiss min max mean stddev;

var DR1TKCAL VTOTALLEG VDRKGRLEG DR1T\_F\_TOTAL FWHOLEFRT DR1T\_G\_WHOLE DR1T\_D\_TOTAL

PFALLPROTLEG PFSEAPLANTLEG MONOPOLY DR1TSFAT DR1TSODI DR1T\_G\_REFINED DR1T\_ADD\_SUGARS;

title2 "Distributions of Simulated Data";

**run**;

\* this proc print is for verification and is not necessary;

\* uncomment next three lines if more information is desired;

**proc** **print** data= SIM\_DATA (obs=**20**);

title2 "Listing of 20 Records from Simulated Data";

**run**;

/\*Section (III): Application of the HEI-2015 scoring algorithm.\*/

%***HEI2015*** (indat= SIM\_DATA,

kcal= DR1TKCAL,

vtotalleg= VTOTALLEG,

vdrkgrleg= VDRKGRLEG,

f\_total= DR1T\_F\_TOTAL,

fwholefrt= FWHOLEFRT,

g\_whole= DR1T\_G\_WHOLE,

d\_total= DR1T\_D\_TOTAL,

pfallprotleg= PFALLPROTLEG,

pfseaplantleg= PFSEAPLANTLEG,

monopoly= MONOPOLY,

satfat= DR1TSFAT,

sodium= DR1TSODI,

g\_refined= DR1T\_G\_REFINED,

add\_sugars= DR1T\_ADD\_SUGARS,

outdat= AFTERMAC);

\* this proc means is for verification and is not necessary;

\* uncomment next four lines if more information is desired;

\*proc means data=AFTERMAC n nmiss min max mean stddev;

\* var VEGDEN GRBNDEN FRTDEN WHFRDEN WGRNDEN DAIRYDEN

PROTDEN SEAPLDEN FARATIO SODDEN RGDEN SFAT\_PERC ADDSUG\_PERC;

\* title2 'after hei 2015 scoring macro';

\*run;

/\*Section (IV): Calculation of HEI-2015 component and total scores and their confidence intervals.\*/

/\*Step 1. This step uses univariate and means procedures to compute total and component scores and their standard errors. \*/

**proc** **univariate** data=AFTERMAC noprint;

var HEI2015C1\_TOTALVEG HEI2015C2\_GREEN\_AND\_BEAN HEI2015C3\_TOTALFRUIT HEI2015C4\_WHOLEFRUIT HEI2015C5\_WHOLEGRAIN

HEI2015C6\_TOTALDAIRY HEI2015C7\_TOTPROT HEI2015C8\_SEAPLANT\_PROT HEI2015C9\_FATTYACID HEI2015C10\_SODIUM

HEI2015C11\_REFINEDGRAIN HEI2015C12\_SFAT HEI2015C13\_ADDSUG HEI2015\_TOTAL\_SCORE;

output out=ci pctlpts=**2.5** **97.5** pctlpre=h1\_ h2\_ h3\_ h4\_ h5\_ h6\_ h7\_ h8\_ h9\_ h10\_ h11\_ h12\_ h13\_ totscore\_;

**run**;

**proc** **means** data=AFTERMAC noprint;

var HEI2015C1\_TOTALVEG HEI2015C2\_GREEN\_AND\_BEAN HEI2015C3\_TOTALFRUIT HEI2015C4\_WHOLEFRUIT HEI2015C5\_WHOLEGRAIN

HEI2015C6\_TOTALDAIRY HEI2015C7\_TOTPROT HEI2015C8\_SEAPLANT\_PROT HEI2015C9\_FATTYACID HEI2015C10\_SODIUM

HEI2015C11\_REFINEDGRAIN HEI2015C12\_SFAT HEI2015C13\_ADDSUG HEI2015\_TOTAL\_SCORE;

output out=stat min=h1\_min h2\_min h3\_min h4\_min h5\_min h6\_min h7\_min h8\_min h9\_min h10\_min h11\_min h12\_min h13\_min totscore\_min

max=h1\_max h2\_max h3\_max h4\_max h5\_max h6\_max h7\_max h8\_max h9\_max h10\_max h11\_max h12\_max h13\_max totscore\_max

mean=h1\_mean h2\_mean h3\_mean h4\_mean h5\_mean h6\_mean h7\_mean h8\_mean h9\_mean h10\_mean h11\_mean h12\_mean h13\_mean totscore\_mean

stddev= h1\_stddev h2\_stddev h3\_stddev h4\_stddev h5\_stddev h6\_stddev h7\_stddev h8\_stddev h9\_stddev h10\_stddev h11\_stddev h12\_stddev h13\_stddev totscore\_stddev;

**run**;

/\*Step 2. This step prepares the results for display\*/

\*add a key variable to merge two datasets - each with only one observation;

**data** CI;

set CI;

key=**1**;

**run**;

**data** STAT;

set STAT;

key=**1**;

**run**;

**data** ALL;

merge CI STAT;

by key;

**run**;

**data** RESULT (keep=score slabel min max mean stderr lowerci upperci);

set ALL;

score='HEI2015x1 ';

slabel='HEI-2015 COMPONENT 1 TOTAL VEGETABLES';

min=h1\_min;

max=h1\_max;

mean=h1\_mean;

stderr=h1\_stddev;

lowerci=h1\_2\_5;

upperci=h1\_97\_5;

output RESULT;

score='HEI2015x2';

slabel='HEI-2015 COMPONENT 2 GREENS AND BEANS';

min=h2\_min;

max=h2\_max;

mean=h2\_mean;

stderr=h2\_stddev;

lowerci=h2\_2\_5;

upperci=h2\_97\_5;

output RESULT;

score='HEI2015x3';

slabel='HEI-2015 COMPONENT 3 TOTAL FRUIT';

min=h3\_min;

max=h3\_max;

mean=h3\_mean;

stderr=h3\_stddev;

lowerci=h3\_2\_5;

upperci=h3\_97\_5;

output RESULT;

score='HEI2015x4';

slabel='HEI-2015 COMPONENT 4 WHOLE FRUIT';

min=h4\_min;

max=h4\_max;

mean=h4\_mean;

stderr=h4\_stddev;

lowerci=h4\_2\_5;

upperci=h4\_97\_5;

output RESULT;

score='HEI2015x5';

slabel='HEI-2015 COMPONENT 5 WHOLE GRAINS';

min=h5\_min;

max=h5\_max;

mean=h5\_mean;

stderr=h5\_stddev;

lowerci=h5\_2\_5;

upperci=h5\_97\_5;

output RESULT;

score='HEI2015x6';

slabel='HEI-2015 COMPONENT 6 DAIRY';

min=h6\_min;

max=h6\_max;

mean=h6\_mean;

stderr=h6\_stddev;

lowerci=h6\_2\_5;

upperci=h6\_97\_5;

output RESULT;

score='HEI2015x7';

slabel='HEI-2015 COMPONENT 7 TOTAL PROTEIN FOODS';

min=h7\_min;

max=h7\_max;

mean=h7\_mean;

stderr=h7\_stddev;

lowerci=h7\_2\_5;

upperci=h7\_97\_5;

output RESULT;

score='HEI2015x8';

slabel='HEI-2015 COMPONENT 8 SEAFOOD AND PLANT PROTEIN';

min=h8\_min;

max=h8\_max;

mean=h8\_mean;

stderr=h8\_stddev;

lowerci=h8\_2\_5;

upperci=h8\_97\_5;

output RESULT;

score='HEI2015x9';

slabel='HEI-2015 COMPONENT 9 FATTY ACID RATIO';

min=h9\_min;

max=h9\_max;

mean=h9\_mean;

stderr=h9\_stddev;

lowerci=h9\_2\_5;

upperci=h9\_97\_5;

output RESULT;

score='HEI2015x10';

slabel='HEI-2015 COMPONENT 10 SODIUM';

min=h10\_min;

max=h10\_max;

mean=h10\_mean;

stderr=h10\_stddev;

lowerci=h10\_2\_5;

upperci=h10\_97\_5;

output RESULT;

score='HEI2015x11';

slabel='HEI-2015 COMPONENT 11 REFINED GRAINS';

min=h11\_min;

max=h11\_max;

mean=h11\_mean;

stderr=h11\_stddev;

lowerci=h11\_2\_5;

upperci=h11\_97\_5;

output RESULT;

score='HEI2015x12';

slabel='HEI-2015 COMPONENT 12 SATURATED FAT';

min=h12\_min;

max=h12\_max;

mean=h12\_mean;

stderr=h12\_stddev;

lowerci=h12\_2\_5;

upperci=h12\_97\_5;

output RESULT;

score='HEI2015x13';

slabel='HEI-2015 COMPONENT 13 ADDED SUGAR';

min=h13\_min;

max=h13\_max;

mean=h13\_mean;

stderr=h13\_stddev;

lowerci=h13\_2\_5;

upperci=h13\_97\_5;

output RESULT;

score='TOTAL HEI 2015';

slabel='TOTAL HEI-2015 SCORE';

min=totscore\_min;

max=totscore\_max;

mean=totscore\_mean;

stderr=totscore\_stddev;

lowerci=totscore\_2\_5;

upperci=totscore\_97\_5;

output RESULT;

**run**;

/\*Step 3 - This step displays the results\*/

**proc** **print** data=RESULT;

id score;

var slabel min max mean stderr lowerci upperci;

title2 'complex survey design population method - mean and confidence interval of HEI-2015 using NH 11/12 data';

**run**;

/\*Step 4 - This step provides an option to export the results into a CSV file that can be opened in Excel.\*/

**proc** **export** data=RESULT

file=RES

dbms=csv

replace;

**run**;