

HEI_distrib_wrapper Documentation (03/02/2018)

The *HEI_distrib_wrapper.sas* macro can be used to fit distributions of HEI scores using the MCMC method. It calls 3 macros which in turn call 4 other macros to produce estimates of the distribution of HEI scores for a population.

We present an example for NHANES 11-12 data, which uses the dataset *FORHEI*. This dataset includes all of the dietary components needed to calculate the HEI score using MCMC.

MCMC is a statistical method that is used when modeling complex distributions, such as the modeling of the HEI score, which requires simultaneous estimation of 21 latent variables (2 for each of the 6 episodically consumed constituents, 1 for probability of consumption and 1 for consumption day amount; 1 for each of the 8 daily consumed dietary constituents; and 1 for energy). It works by simulating the random variables many times to estimate the distributions (the “Monte Carlo” piece), and uses an algorithm to ensure the correct distribution is estimated (the “Markov Chain” piece). At least 2 24-hour recalls on a subset of participants are required to fit the MCMC model. See Zhang et al. (2011) for a statistical paper describing MCMC for the HEI-2005.

There are 10 steps to fit the distribution of the HEI-2015 score. These include:

- 1) Data preparation
- 2) Determine strata
- 3) Explore data characteristics
- 4) Find minimum values for the dietary components and replace daily consumed 0 values with $\frac{1}{2}$ the minimum amount
- 5) Find lambda values to transform data to approximate normality using Box-Cox transformation
- 6) Transform dietary data to have a mean of 0 and a variance of 2 (needed for MCMC model fitting)
- 7) Run MCMC model fitting
- 8) Generate distribution of usual intake of dietary components
- 9) HEI scoring
- 10) Calculation of standard errors

This document will describe the SAS code and output necessary to complete each step. Steps 2-10 may be executed using the *HEI_distrib_wrapper.sas* macro.

First, the macro defines where the data will be stored and read in all the macros. The user will need to put in the appropriate paths. You should create an empty folder to store the data from the macro; here it is called *dataout* and is a subfolder within the home folder. This code also assumes the macros are in a subfolder of the home folder called *macros*.

```
/** Assign libnames for the data set and macros using macro variables */  
/** Note: if using the UNIX system, you will need to replace backslashes with forward slashes */  
  
options nonotes;  
  
%global home macros titles liblabel;  
  
%let home = H:\HEI\HEI wrapper;  
%let liblabel = &home\dataout; **make sure you have created a folder here called 'dataout' (or whatever name you choose) for all output;  
  
libname data "&home\data"; **dataset is in a subfolder called 'data';
```

```

/**** Include all macros ****/
/**** This code assumes all macros are stored in a subfolder of home called 'macros' ****/

%include "&home\macros\std_cov_boxcox24hr_conday_minamt_macro_v2.01.sas";
%include "&home\macros\multivar_mcmc_macro_v1.01.sas";
%include "&home\macros\multivar_distrib_macro_v1.0.sas";
%include "&home\macros\hei2015.score_macro.sas";

%include "&home\macros\explore_episodic_n_macro.sas";
%include "&home\macros\fitdistrib_macro.sas";
%include "&home\macros\runcorese_macro.sas";

/*****
/**** STEP 1: DATA PREPARATION ****/
/*****/

```

Step 1 is data preparation. Details are given in *step1.create.data1112.sas*. Note that when using the MCMC method, a dietary component cannot be contained within another dietary component (e.g., whole fruit as part of total fruit). Therefore, the components fit in the MCMC macro are not the components of the HEI, but will be summed to calculate the HEI components. See the table below for details regarding the dietary components modeled in MCMC.

HEI Component	MCMC Variables	Type ^a	Units	Dietary Constituents
From FPED (or other food-based database)				
Total Fruits	Whole Fruit	Episodic	cup eq.	Citrus, Melons, Berries + Other Intact Fruits
	Fruit Juice	Episodic	cup eq.	Fruit Juice
Whole Fruits	Whole Fruit	Episodic	cup eq.	Citrus, Melons, Berries + Other Intact Fruits
Total Vegetables	Non-Dark Green Vegetables	Daily	cup eq.	Total Vegetables – Dark Green Vegetables ^b
	Dark Green Vegetables	Episodic	cup eq.	Dark Green Vegetables
	Legumes (Beans and Peas)	Episodic	cup eq.	Legumes (Beans and Peas)
Greens and Beans	Dark Green Vegetables	Episodic	cup eq.	Dark Green Vegetables
	Legumes (Beans and Peas)	Episodic	cup eq.	Legumes (Beans and Peas)
Whole Grains	Whole Grains	Episodic	oz. eq.	Whole Grains
Refined Grains	Refined Grains	Daily	oz. eq.	Refined Grains
Dairy	Dairy	Daily	cup eq.	Total Dairy
Total Protein Foods	Meat, Poultry and Eggs ^b	Daily	oz. eq.	Total Meat, Poultry (including organ meats and cured meats) + Eggs
	Seafood, Soy, and Nuts and Seeds	Episodic	oz. eq.	Seafood (high in n-3) + Seafood (low in n-3) + Soy + Nuts and Seeds
	Legumes (Beans and Peas)	Episodic	oz. eq.	Beans and Peas
Seafood and Plant Proteins	Seafood, Soy, and Nuts and Seeds	Episodic	oz. eq.	Seafood (high in n-3) + Seafood (low in n-3) + Soy + Nuts and Seeds
	Legumes (Beans and Peas)	Episodic	oz. eq.	Legumes (Beans and Peas)
Added Sugars	Added Sugars	Daily	tsp. eq.*	Added Sugars
From FNDDS (or other nutrient database)				
Fatty Acids	Fatty Acids	Daily	g	Total Monounsaturated Fatty Acids + Total Polyunsaturated Fatty Acids
Saturated Fats	Saturated Fats	Daily	g*	Total Saturated Fatty Acids
Sodium	Sodium	Daily	mg	Sodium
----	Energy	Daily	kcal	Total Energy

cup eq.=cup equivalents; oz. eq.=ounce equivalents; tsp. eq.= teaspoon equivalents; tsp. eq*.=added sugars are calculated in teaspoon equivalents but converted to energy in the scoring process; g=grams; g*= fatty acids are calculated in grams but converted to energy in the scoring process; mg=milligrams

^a General guidance on daily vs. episodic. Should be examined for each individual dataset.

^b Should be rounded to 2 decimal places and set to 0.00 if value is negative.

Note that at least 50 people with 2 non-zero recalls (for the dietary component) per stratum are suggested for each dietary component. The *explore* macro (described below) will provide the number of non-zero recalls by stratum and dietary component. This information should be considered when determining if this methodology is appropriate for your dataset and when determining the number of strata.

```

/***** STEP 2: DETERMINE STRATUM *****/

```

Step 2 is to define strata. This allows the variance components to differ by strata. Note that the complex survey scheme of NHANES (age/gender sampling frame) was considered when creating the strata. The following code creates 3 strata in the NHANES 11-12 dataset, and then create one dataset for each stratum:

```

DATA HEI;
SET IMPORT.FORHEI;
IF 2 <= RIDAGEYR <= 11 then STRATUM=1;
IF RIDAGEYR >= 12 and MALE=1 then STRATUM=2;
IF RIDAGEYR >= 12 and MALE=0 then STRATUM=3;
RUN;

```

```

data HEI1 HEI2 HEI3;
set HEI;
if stratum=1 then output HEI1;
if stratum=2 then output HEI2;
if stratum=3 then output HEI3;
run;

```

```

/***** STEP 3: EXPLORE DATA CHARACTERISTICS *****/

```

Step 3 is to look at the data to determine which components are episodic and which are consumed daily (or almost daily), as well as to ensure there are at least 50 people with 2 recalls per stratum. This is done by calling the explore macro. The code below is divided by the components used in MCMC, in the usual scoring of the HEI, and may be used with other scoring methods as a step to explore the data.

```

%explore(HEI,seqn,
/*mcmc and usual n=9 */
wholefruit=FWHOLEFRT, wholegrains=G_WHOLE, refinedgrains=G_REFINED, dairy=D_TOTAL, sugars=ADD_SUGARS, satfat=TSFAT,
sodium=TSODI, kcal=TKCAL, mufapufa=MONOPOLY,
/*usual only n=5 */
totalfruit=F_TOTAL, beansandgreens=VDRKGRLEG, vegandlegumes=VTOTALLEG, seafoodplant=PFSEAPLANTLEG, allprotein=PFALLPROTLEG,
/*mcmc only n=6 */
fruitjuice=F_JUICE, nondarkgr=NONDRKGR, darkgr=V_DRKGR, legumes=V_LEGUMES, seasoynut=SEASOYNUT, mpeggs=MPEGGS)

```

The log file will have:

```

The variables considered are FWHOLEFRT G_WHOLE G_REFINED D_TOTAL ADD_SUGARS TSFAT TSODI TKCAL
MONOPOLY F_TOTAL VDRKGRLEG VTOTALLEG PFSEAPLANTLEG PFALLPROTLEG F_JUICE NONDRKGR V_DRKGR
V_LEGUMES PFSEASOYNUT PFMPEGGS
The number of variables is 20
ind_list = FWHOLEFRT_01 G_WHOLE_01 G_REFINED_01 D_TOTAL_01 ADD_SUGARS_01 TSFAT_01 TSODI_01 TKCAL_01
MONOPOLY_01 F_TOTAL_01 VDRKGRLEG_01 VTOTALLEG_01 PFSEAPLANTLEG_01 PFALLPROTLEG_01 F_JUICE_01
VNONDRKGR_01 V_DRKGR_01 V_LEGUMES_01 PFSEASOYNUT_01 PFMPEGGS_01

```

This macro will save a pdf file called Component_Characteristics.pdf in the \dataout folder. (The same table will also print in the Output window or .lst output. (Note that the “ods noresults” statement will turn off the Results Viewer and keep pdf files from automatically opening.) A selection of the output table is below. The table gives, for each stratum, the percent with 0 intakes and whether this would be

episodic (<5% is Not Episodic, 5-10% is Marginal, and >10% is Episodic), and the number with 2+ non-zero 24 hour recalls.

Component Characteristics

Used for	HEI Component	Stratum	Percent 0	Episodic	Number with 2+ non-zero recalls	Note
All	ADD_SUGARS	1	0.1	Not Episodic	1728	
		2	0.3	Not Episodic	2547	
		3	0.3	Not Episodic	2684	
	D_TOTAL	1	0.3	Not Episodic	1690	
		2	2.7	Not Episodic	2253	
		3	2.0	Not Episodic	2418	
	FWHOLEFRT	1	14.4	Episodic	977	
		2	32.8	Episodic	953	

Also, in the Output window or .lst output, the means procedure will print the sample size, number missing, minimum, maximum, and mean by stratum. These data are presented to review to make sure that the numbers appear reasonable. For example, for Stratum=1:

The MEANS Procedure STRATUM=1						
Variable	Label	N	N Miss	Minimum	Maximum	Mean
FWHOLEFRT	Whole Fruit	3722	0	0	11.5700000	0.7967894
G_WHOLE	Whole grains (oz. eq.)	3722	0	0	11.3300000	0.7640489
G_REFINED	Refined or non-whole grains (oz. eq.)	3722	0	0	43.2400000	5.3455508
D_TOTAL	Total milk, yogurt, cheese, and whey (cup eq.)	3722	0	0	9.2300000	2.1400833
ADD_SUGARS	Foods defined as added sugars (tsp. eq.)	3722	0	0	83.1100000	14.7696803
TSFAT	Total saturated fatty acids (gm)	3722	0	0	83.2320000	22.4858426
TSODI	Sodium (mg)	3722	0	37.0000000	13381.00	2720.36
TKCAL	Energy (kcal)	3722	0	63.0000000	5419.00	1772.20
MONOPOLY	MUFAs and PUFAs	3722	0	0	158.8100000	36.9165543
F_TOTAL	Total intact or cut fruits and fruit juices (cup eq.)	3722	0	0	15.1700000	1.3121252
VDRKGRLEG	Dark Green Vegetables and Legumes	3722	0	0	2.9300000	0.1354164
VTOTALLEG	Total Vegetables (inc Legumes)	3722	0	0	6.3000000	0.9059753
PFSEAPLANTLEG	Seafood and Plant Proteins (inc Legumes)	3722	0	0	19.3800000	0.9108517
PFALLPROTLEG	All Protein (inc Legumes)	3722	0	0	27.8400000	3.9964481
F_JUICE	Fruit juices, citrus and non citrus (cup eq.)	3722	0	0	13.5100000	0.5152499
VNONDRKGR	Vegetables without Greens (w/o Legumes)	3722	0	0	6.3000000	0.7705588

Variable	Label	N	N Miss	Minimum	Maximum	Mean
V_DRKGR	Dark green vegetables (cup eq.)	3722	0	0	2.4600000	0.0550296
V_LEGUMES	Legumes computed as vegetables (cup eq.)	3722	0	0	2.7500000	0.0803869
PFSEASOYNUT	Seafood, Soy, and Nuts (w/o Legumes)	3722	0	0	16.7200000	0.5874449
PFMPEGGS	Meat, Poultry, and Eggs (w/o Legumes)	3722	0	0	24.8500000	3.0856072

The next 5 steps are fit in the *fitdistrib.sas* macro.

```

/*****
/**** The SAS wrapper macro called fitdistrib.sas ****
/**** fits 5 steps to be used when interest is on obtaining ****
/**** distributions ****
/****
/**** The five steps are: ****
/****
/**** STEP 4: FIND MINIMUM VALUES ****
/****
/**** STEP 5: FIND LAMBDA ****
/****
/**** STEP 6: TRANSFORM DATA TO STANDARD NORMAL ****
/****
/**** STEP 7: RUN MCMC MODEL FITTING ****
/****
/**** STEP 8: RUN DISTRIBUTION MACRO ****
/****
/****

```

The %fitdistrib macro will fit steps 4-8. This is quite time consuming; it takes about ½ hour to run 1 replicate for 1 stratum using 5000 MCMC iterations and 3000 burn-ins.

Important note: At this time, the *fitdistrib* macro is set up to adjust for weekend and sequence effects, so these should be included when using the *HEI_distrib_wrapper v1.0*.

You may want to consider running replicate 0 first, then replicates 1-16. (Replicate 0 will be used to obtain the estimates of the scores; replicates 1-16 are used to obtain the standard errors.) The macro is called 3 times, 1 time for each stratum. Note that you will determine which variables to list in episodic and which in daily based on the results of the explore macro. Note that the macro variables baseweight is the variable name for the dietary weight, and the weightvarroot variable assumes that all the BRR macro variables are named RepWt_1 to RepWt_XX (where here XX=16). Also note the seed can be set to any random number, and if it is set to the number 0 SAS will use the time of day to calculate the seed.

This random number is just used to initialize the simulations, and therefore the actual number chosen is not meaningful. The code below will fit from replicate 0 (base weights) to replicate 16: (Note that if you wanted to start with running replicate 0 first to make sure the code was working correctly, you would set repliclast=0. If you then wanted to run the BRR replicates you could set replicfirst=1 and repliclast=16 to run all BRR replicates. They do not have to be run at the same time because the output datasets will be saved.)

```

*****
** Stratum 1 - children aged 2-12 **
*****

```

```

%fitdistrib
  (data      = hei1,
  stratum    = 1,

```

```

id          = seqn,
repeat      = day,
baseweight  = RepWt_0,
weightvarroot = RepWt_,
episodic    = g_whole fwwholefrt f_juice pfseasoynut v_drkgr v_legumes,
daily       = add_sugars d_total g_refined tkcal tsfat tsodi pfmpeeggs vnondrkgr monopoly,
covlisti    = male raceeth_nhbl raceeth_asian raceeth_hisp age6_11,
covlistc    = ,
sequence    = recallday2,
weekend     = weekend,
outstd      = stdcov_s1,
tranout     = tranout1,
replicfirst = 0,
repliclast  = 16,
n_mcmc_iterations = 5000,
n_burn_iterations = 3000,
thin        = 10,
folder      = mcmcdout1,
outpath     = "&liblabel",
gen_inverse = y,
print       = n,
ntitles     = 3,
store_sd    = n,
notes_print_mcmc = y,
npseudopeople = 100,
seed        = 42941)

```

See the *HEI_distrib_wrapper.sas* macro for definitions of the macro parameters.

The log file will have:

starting STEP 4: FIND MINIMUM VALUES

Starting STEP 5: FIND LAMBDA

Finding minimum and lambdas for stratum 1

starting STEP 6: TRANSFORM DATA TO STANDARD NORMAL

Standardizing covariates for stratum 1

Variables from the std_cov_boxcox24hr_conday_minamt (Stratum 1)

conday_g_whole conday_fwwholefrt conday_f_juice conday_pfseasoynut conday_v_drkgr conday_v_legumes

stdbc_g_whole stdbc_fwwholefrt stdbc_f_juice stdbc_pfseasoynut stdbc_v_drkgr stdbc_v_legumes

stdbc_add_sugars stdbc_d_total stdbc_g_refined stdbc_tkcal stdbc_tsfat stdbc_tsodi stdbc_pfmpeeggs

stdbc_vnondrkgr stdbc_monopoly

Starting STEP 7: RUN MCMC MODEL FITTING

MCMC Stratum: 1 Replicate: 0

Output Plot Location H:\Usual Intake\HEI\final code\HEI

wrapper\dataoutv11\mcmcdout1.trace.replic0.pdf

starting STEP 8: RUN DISTRIBUTION MACRO

Distrib Stratum 1 Replicate 0

MCMC Stratum: 1 Replicate: 1

Output Plot Location H:\Usual Intake\HEI\final code\HEI

wrapper\dataoutv11\mcmcdout1.trace.replic1.pdf

starting STEP 8: RUN DISTRIBUTION MACRO

Distrib Stratum 1 Replicate 1

MCMC Stratum: 1 Replicate: 2

Output Plot Location H:\Usual Intake\HEI\final code\HEI

wrapper\dataoutv11\mcmcdout1.trace.replic2.pdf.

.

etc.

The Output window or .lst output will show the means for the weights and usual intake variables. For Step 4, it will show the dietary components and the minimum amount and then the lambdas for the Box-Cox transformation:

STEPS 4 & 5: FIND MINIMUM VALUES & LAMBDAS

Minimum Values

Obs	component	STRATUM	_TYPE_	_FREQ_	minamount	tran_paramindex
1	g_whole	1	0	2383	0.0050	1
2	fwholefrt	1	0	2674	0.0050	2
3	f_juice	1	0	2598	0.0050	3
4	pfseasoynut	1	0	1680	0.0050	4
5	v_drkgr	1	0	560	0.0050	5
6	v_legumes	1	0	693	0.0050	6
7	add_sugars	1	0	3708	0.0050	7
8	d_total	1	0	3667	0.0050	8
9	g_refined	1	0	3700	0.0150	9
10	tkcal	1	0	3722	31.5000	10
11	tsfat	1	0	3719	0.1935	11
12	tsodi	1	0	3722	18.5000	12
13	pfmpeegs	1	0	3563	0.0050	13
14	vnondrkgr	1	0	3433	0.0050	14
15	monopoly	1	0	3719	0.5700	15

STEPS 4 & 5: FIND MINIMUM VALUES & LAMBDAS

Lambdas

Obs	STRATUM	tran_lambda	component	tran_paramindex
1	1	0.13	G_WHOLE	1
2	1	0.32	FWHOLEFRT	2
3	1	0.27	F_JUICE	3
4	1	0.13	PFSEASOYNUT	4
5	1	0.29	V_DRKGR	5
6	1	0.31	V_LEGUMES	6
7	1	0.38	ADD_SUGARS	7
8	1	0.50	D_TOTAL	8
9	1	0.36	G_REFINED	9
10	1	0.42	TKCAL	10
11	1	0.29	TSFAT	11
12	1	0.28	TSODI	12
13	1	0.39	PFMPEGGS	13

Obs	STRATUM	tran_lambda	component	tran_paramindex
14	1	0.27	VNONDRKGR	14
15	1	0.33	MONOPOLY	15

For Step 6, it will show the names of the standardized variables (with the prefix “stdbc_”). You can check that these were properly transformed by checking that the mean is approximately 0 and the standard deviation is 1.41 (square root of 2) for the daily consumed dietary components. (Note that if there were any continuous covariates they should have a mean of 0 and a variance/standard deviation of 1.)

STEP 6: TRANSFORM DATA TO STANDARD NORMAL

Consumption-Day Indicators for Episodically Consumed Components and Standardized Recall Variables

The MEANS Procedure						
Variable	N	N Miss	Mean	Std Dev	Minimum	Maximum
conday_g_whole	3722	0	0.6402472	0.4799923	0	1.0000000
conday_fwholefrt	3722	0	0.7184310	0.4498247	0	1.0000000
conday_f_juice	3722	0	0.6980118	0.4591818	0	1.0000000
conday_pfseasoynut	3722	0	0.4513702	0.4976964	0	1.0000000
conday_v_drkgr	3722	0	0.1504567	0.3575666	0	1.0000000
conday_v_legumes	3722	0	0.1861902	0.3893124	0	1.0000000
stdbc_g_whole	3722	0	7.827042E-17	1.1315039	-5.1022711	4.6731373
stdbc_fwholefrt	3722	0	5.345297E-17	1.1986286	-3.2013357	5.6818388
stdbc_f_juice	3722	0	8.037036E-16	1.1814658	-2.9293549	6.5189573
stdbc_pfseasoynut	3722	0	-3.33127E-16	0.9499714	-3.4601380	4.2323977
stdbc_v_drkgr	3722	0	4.185559E-16	0.5481395	-2.7454367	4.1601517
stdbc_v_legumes	3722	0	-1.18122E-16	0.6098713	-3.1744620	4.4619396
stdbc_add_sugars	3722	0	-7.22188E-15	1.4142136	-4.6057560	5.1142867
stdbc_d_total	3722	0	-3.39045E-15	1.4142136	-3.6040096	4.6201758
stdbc_g_refined	3722	0	-1.99971E-15	1.4142136	-5.2608263	7.3478956
stdbc_tkcal	3722	0	2.023481E-14	1.4142136	-6.6011327	5.5240453
stdbc_tsfat	3722	0	-3.83525E-15	1.4142136	-6.6089902	4.5034352
stdbc_tsodi	3722	0	7.028589E-15	1.4142136	-7.7273551	6.6143281
stdbc_pfmpeggs	3722	0	2.205651E-15	1.4142136	-3.3329944	5.4553905
stdbc_vnondrkgr	3722	0	1.030879E-15	1.4142136	-3.1259091	4.2173868
stdbc_monopoly	3722	0	3.648643E-16	1.4142136	-6.0517931	5.4455088

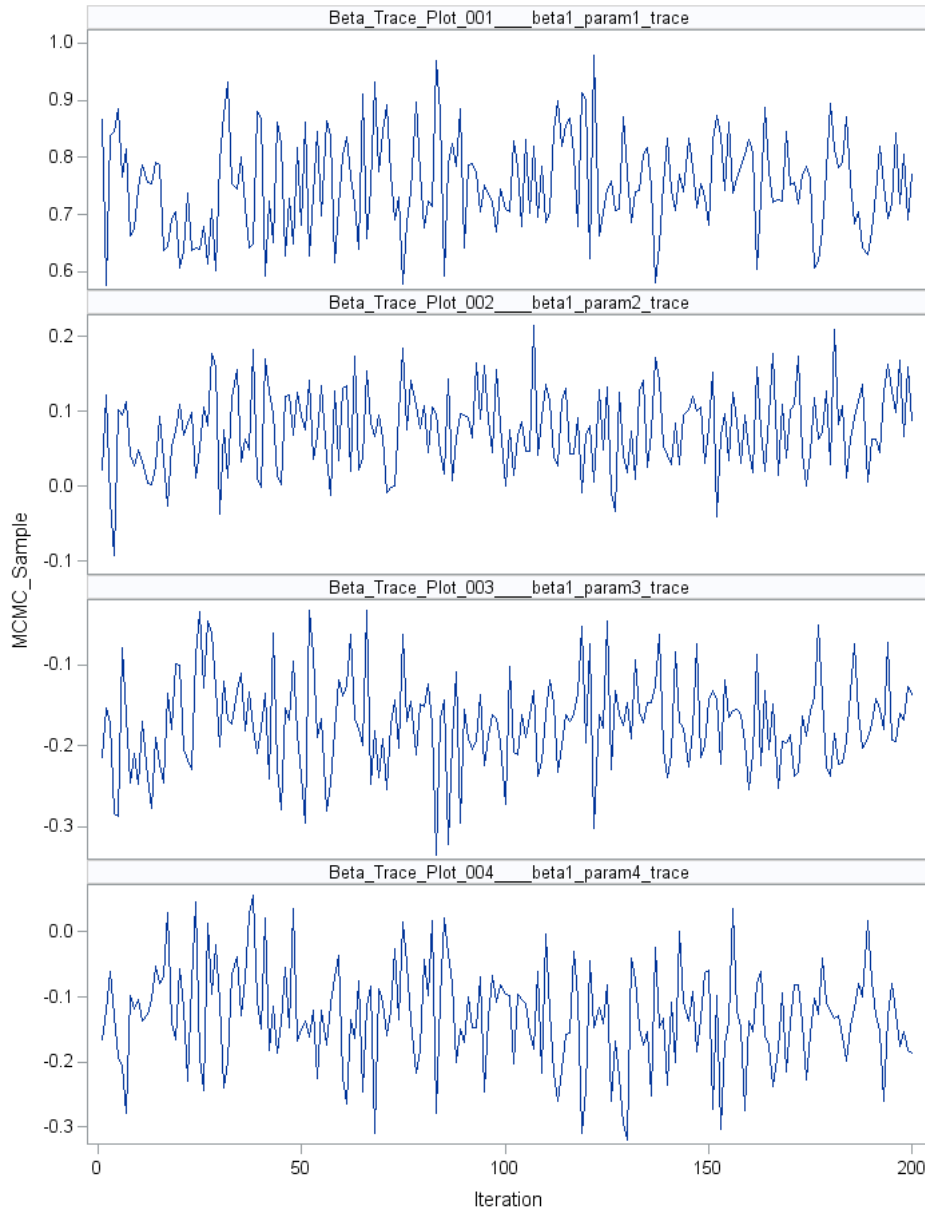
Next the variables for back-transformation print. These are for information only.

STEP 6: TRANSFORM DATA TO STANDARD NORMAL

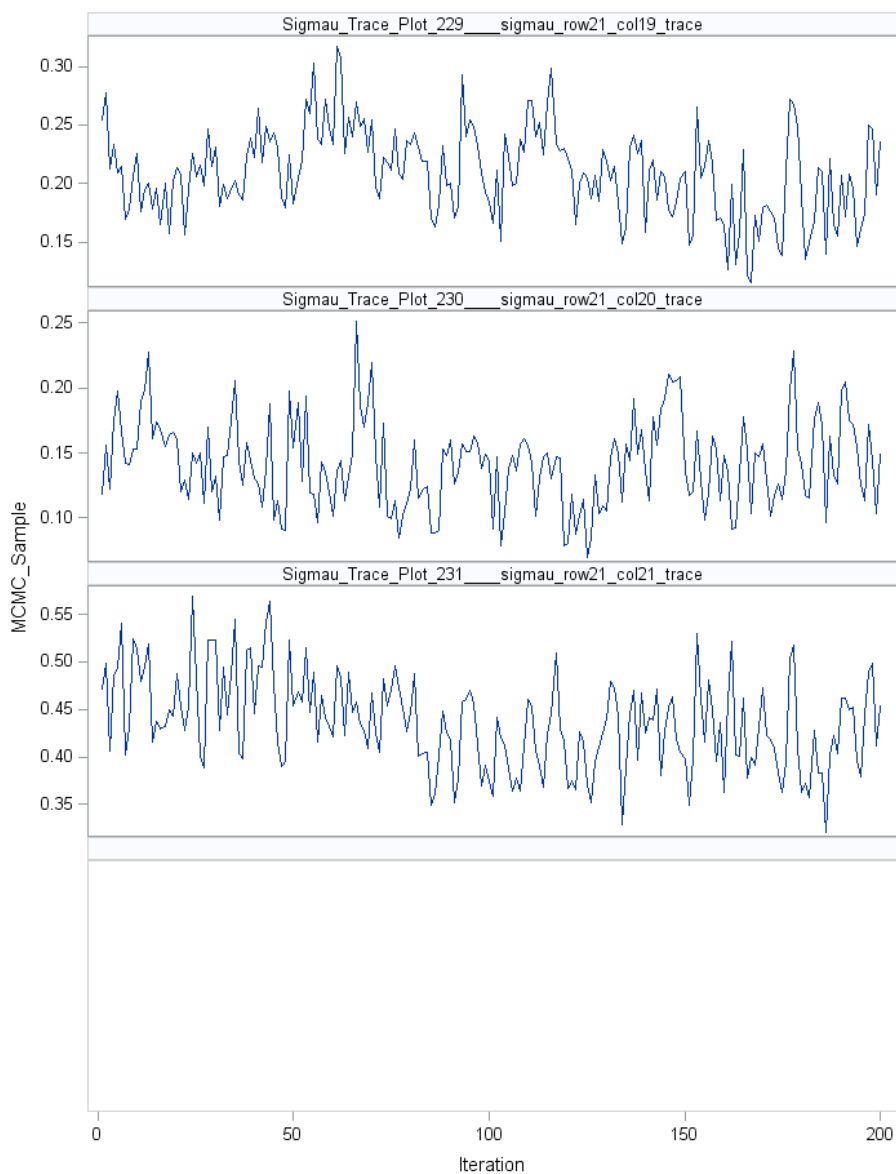
Variables for Back-Transformation

Obs	tran_lambda	tran_paramindex	tran_center	tran_scale	minamount
1	0.13	1	-0.1668	0.64644	0.0050
2	0.32	2	-0.2017	0.68951	0.0050
3	0.27	3	-0.6476	0.67864	0.0050
4	0.13	4	-0.3765	0.89262	0.0050
5	0.29	5	-1.1220	0.51695	0.0050
6	0.31	6	-0.9388	0.47669	0.0050
7	0.38	7	4.2416	1.41601	0.0050
8	0.50	8	0.7422	0.72162	0.0050
9	0.36	9	2.0769	0.80638	0.0150
10	0.42	10	51.7651	6.14734	31.5000
11	0.29	11	4.8125	0.92588	0.1935
12	0.28	12	28.4783	2.87727	18.5000
13	0.39	13	1.0413	0.98431	0.0050
14	0.27	14	-0.6035	0.70839	0.0050
15	0.33	15	6.6550	1.18445	0.5700

Step 8 is the MCMC fitting process. Beta trace plots will be produced for each beta and each parameter and will be written to pdf files. In our example there are 21 betas (one for each daily consumed component ($N=9$) + two for each episodically consumed component ($N=6 \times 2=12$), and 8 parameters (5 covariates + weekend + recall day + intercept). These plots should not exhibit any pattern, but should show random variation. For example:



There will also be plots for Sigmau and Sigmae for each variance component (row1, col1 through row 21, col21). These should also look fairly random, for example:



Next, the stored matrices will print:

STEP 8: RUN DISTRIBUTION MACRO Stratum 1 Replicate 0

***** Output from the Show Storage Statement *****

Contents of storage library = LIBLABEL.IMLMCMCDOUT1_REPLIC0

Matrices:

AGE6_11	BETA10_MEAN
BETA11_MEAN	BETA12_MEAN
BETA13_MEAN	BETA14_MEAN
BETA15_MEAN	BETA16_MEAN
BETA17_MEAN	BETA18_MEAN
BETA19_MEAN	BETA1_MEAN
BETA20_MEAN	BETA21_MEAN
BETA2_MEAN	BETA3_MEAN
BETA4_MEAN	BETA5_MEAN
BETA6_MEAN	BETA7_MEAN
BETA8_MEAN	BETA9_MEAN
CONSTANT1	MALE
MINAMOUNT	NUM_DAILY_DIET_COMP
NUM_EPIS_DIET_COMP	NUM_ROWS_COVMAT
RACEETH_ASIAN	RACEETH_HISP
RACEETH_NHBL	RECALLDAY2
SIGMAE_MEAN	SIGMAU_MEAN
TRAN_CENTER	TRAN_LAMBDA
TRAN_PARAMINDEX	TRAN_SCALE
WEEKEND	WEIGHT_NW_SUMW

Modules:

And the stored values and dimensions will print.

***** Stored Values and Dimensions *****

NUM_EPIS_DIET_COMP	NUM_DAILY_DIET_COMP	NUM_ROWS_COVMAT
6	9	21
Number of Individuals		
1982		

Macro Parameter: set_number_monte_carlo_rand_obs 100

This information will repeat for each stratum x replicate combination. For example, for NHANES 11-12, there are 3 strata and 17 replicates, leading to 51 sets of beta trace plots, sigmau trace plots, sigmae trace plots, and information on the stored matrices and stored values and dimensions.

A dataset called 'mc_t_distrib_out' is saved for each stratum and each replicate using the naming convention *liblabel.distmc&stratum.&replicnum*. For example, for replicate 0, stratum 1, the file is *liblabel.distmc01*. The dataset will include usual intake variables mc_t1, mc_t2, ..., mc_t15, one for each dietary component in the order they are specified.

```

/*****
/**** The SAS wrapper macro called runscorese.sas
/**** scores the simulated distribution data for the HEI-2015
/**** and also calculates SEs using BRR
/****
/**** STEP 9: SCORE EACH REPLICATE
/****
/**** STEP 10: CALCULATE STANDARD ERRORS USING BRR
/****
/*****/

```

In Steps 9 and 10, the %runscorese macro scores each replicate and then takes all the replicates to get the point estimates (from the 0 run) and the standard errors (here from the 16 BRR replicates).

```

%runscorese(outpath="&liblabel", replicfirst=0, replicfirstse=1, repliclast=16, finaldataset=tdheistat16,
episodic=g_whole fwwholefrt f_juice pfseasoynut v_drkgr v_legumes,
daily=add_sugars d_total g_refined tkcal tsfat tsodi pfmpepgs vnondrkgr monopoly,
wholefruit=fwholefrt, beansandgreens=VDRKGRLEG, wholegrains=G_WHOLE, refinedgrains=G_REFINED,
dairy=D_TOTAL, sugars=ADD_SUGARS, satfat=tsfat, sodium=tsodi,
kcal=TKCAL, totalfruit=F_TOTAL, vegandlegumes=VTOTALLEG, seafoodplant=PFSEAPLANTLEG, allprotein=PFALLPROTLEG,
mufapufa=MONOPOLY, fruitjuice=F_JUICE, nondarkgr=VNONDRKGR, darkgr=V_DRKGR, legumes=V_LEGUMES, seasoynut=SEASOYNUT,
mpepgs=PFMPEPGS);

```

In Step 9, first the components of the HEI score are created from the usual intake of the 15 dietary components that were just predicted. For each replicate, the macro will print the means and percentiles. Note that the first step is to create a dataset that sets the distmc output datasets. (If three are not used, the runscorese macro will need to be modified.)

The log file specifies how the dietary components correspond to usual intakes. Recall that the order output in the distribmc files is the order specified of episodic and daily variables:

```

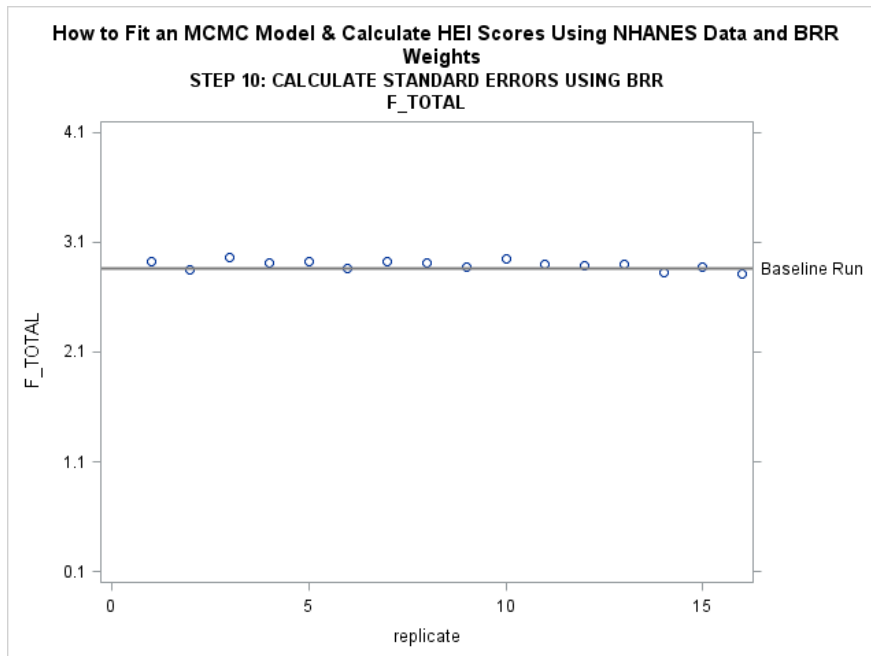
starting STEP 9: SCORE EACH REPLICATE
usintake_g_whole= mc_t1;usintake_fwwholefrt= mc_t2;usintake_f_juice= mc_t3;usintake_pfseasoynut=
mc_t4;usintake_v_drkgr= mc_t5;usintake_v_legumes= mc_t6;usintake_add_sugars= mc_t7;usintake_d_total=
mc_t8;usintake_g_refined= mc_t9;usintake_tkcal= mc_t10;usintake_tsfat= mc_t11;usintake_tsodi=
mc_t12;usintake_pfmpepgs= mc_t13;usintake_vnondrkgr= mc_t14;usintake_monopoly= mc_t15;
Replicate 0 COUNT OF COMPONENTS + TOTALs should be 15: FOUND 15
Replicate 1 COUNT OF COMPONENTS + TOTALs should be 15: FOUND 15
Replicate 2 COUNT OF COMPONENTS + TOTALs should be 15: FOUND 15
Replicate 3 COUNT OF COMPONENTS + TOTALs should be 15: FOUND 15
Replicate 4 COUNT OF COMPONENTS + TOTALs should be 15: FOUND 15
Replicate 5 COUNT OF COMPONENTS + TOTALs should be 15: FOUND 15
Replicate 6 COUNT OF COMPONENTS + TOTALs should be 15: FOUND 15
Replicate 7 COUNT OF COMPONENTS + TOTALs should be 15: FOUND 15
Replicate 8 COUNT OF COMPONENTS + TOTALs should be 15: FOUND 15
Replicate 9 COUNT OF COMPONENTS + TOTALs should be 15: FOUND 15
Replicate 10 COUNT OF COMPONENTS + TOTALs should be 15: FOUND 15
Replicate 11 COUNT OF COMPONENTS + TOTALs should be 15: FOUND 15
Replicate 12 COUNT OF COMPONENTS + TOTALs should be 15: FOUND 15
Replicate 13 COUNT OF COMPONENTS + TOTALs should be 15: FOUND 15
Replicate 14 COUNT OF COMPONENTS + TOTALs should be 15: FOUND 15
Replicate 15 COUNT OF COMPONENTS + TOTALs should be 15: FOUND 15
Replicate 16 COUNT OF COMPONENTS + TOTALs should be 15: FOUND 15
starting STEP 10: CALCULATE STANDARD ERRORS USING BRR

```

The Output window or .lst output will show the means for the weights and usual intake variables. For example, for replicate 0:

HEI Score Statistics for Replicate Number 0													
Obs component	replicate	_TYPE_	_FREQ	Mean_ score	p1	p5	p10	p25	p50	p75	p90	p95	p99
1 VTOTALLEG	0	0	793500	3.14	0.98	1.36	1.61	2.16	3.01	4.14	5.00	5.00	5.00
2 VDRKGRLEG	0	0	793500	2.46	0.21	0.45	0.65	1.17	2.12	3.68	5.00	5.00	5.00
3 F_TOTAL	0	0	793500	2.87	0.11	0.36	0.64	1.40	2.79	4.70	5.00	5.00	5.00
4 fwwholefrt	0	0	793500	3.15	0.03	0.21	0.49	1.47	3.46	5.00	5.00	5.00	5.00
5 G_WHOLE	0	0	793500	2.95	0.13	0.44	0.73	1.46	2.55	3.96	5.69	7.00	10.00
6 D_TOTAL	0	0	793500	6.52	1.40	2.36	3.02	4.42	6.45	8.98	10.00	10.00	10.00
7 PFALLPROTLEG	0	0	793500	4.58	2.37	3.05	3.48	4.31	5.00	5.00	5.00	5.00	5.00
8 PFSEAPLANTLEG	0	0	793500	3.62	0.59	1.10	1.49	2.41	3.98	5.00	5.00	5.00	5.00
9 MONOPOLY_tsfat	0	0	793500	4.63	0.00	1.06	1.69	2.88	4.40	6.17	7.98	9.18	10.00
10 tsodi	0	0	793500	4.36	0.00	0.00	0.71	2.63	4.49	6.13	7.48	8.23	9.55
11 G_REFINED	0	0	793500	5.90	0.00	1.05	2.34	4.17	6.02	7.87	9.47	10.00	10.00
12 tsfat	0	0	793500	6.12	0.69	2.42	3.29	4.69	6.19	7.66	8.97	9.73	10.00
13 ADD_SUGARS	0	0	793500	6.27	0.00	1.01	2.48	4.58	6.58	8.33	9.72	10.00	10.00
14 total	0	0	793500	56.56	32.09	38.87	42.55	49.04	56.43	64.01	70.90	74.82	81.31
15 TKCAL	0	0	793500	2104.54	1043.31	1270.08	1406.45	1664.34	2010.00	2449.10	2944.54	3264.79	3892.33

The macro also creates a pdf of plots of all the replicates and the baseline run in a file called *hei_plots.pdf*. The replicate values should be close to the baseline run. Here is the example for total fruit:



The contents of the final dataset are also printed:

How to Fit an MCMC Model & Calculate HEI Scores Using NHANES Data and BRR Weights
Description of the data set with HEI Scores and Statistics

The CONTENTS Procedure			
Data Set Name	LIBLABEL.TDHEISTAT16SE	Observations	15
Member Type	DATA	Variables	25
Engine	V9	Indexes	0
Created	09/03/2017 22:07:34	Observation Length	224
Last Modified	09/03/2017 22:07:34	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	WINDOWS_64		
Encoding	wlatin1 Western (Windows)		

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Label
1	component	Char	32	\$32.	HEI Component
22	lcl_95	Num	8		Lower 95% CI Limit
2	mean_score	Num	8		Mean
12	mean_score_se	Num	8		SE Mean
3	p1	Num	8		1st Percentile
4	p5	Num	8		5th Percentile
5	p10	Num	8		10th Percentile
6	p25	Num	8		25th Percentile
7	p50	Num	8		50th Percentile
8	p75	Num	8		75th Percentile
9	p90	Num	8		90th Percentile
10	p95	Num	8		95th Percentile
11	p99	Num	8		99th Percentile
15	p10_se	Num	8		SE 10th Percentile
13	p1_se	Num	8		SE 1st Percentile
16	p25_se	Num	8		SE 25th Percentile
17	p50_se	Num	8		SE 50th Percentile
14	p5_se	Num	8		SE 5th Percentile
18	p75_se	Num	8		SE 75th Percentile
19	p90_se	Num	8		SE 90th Percentile
20	p95_se	Num	8		SE 95th Percentile
21	p99_se	Num	8		SE 99th Percentile
25	pvalue	Num	8		p-value
23	ucl_95	Num	8		Upper 95% CI Limit
24	waldchisq	Num	8		Wald chi square

Finally, the HEI scores and percentiles with SE will be printed:

HEI Scores														
Mean HEI Score		Percentiles									95% CI		P-Value	
HEI Component	Score (se)	p1	p5	p10	p25	p50	p75	p90	p95	p99	Lower	Upper	Wald Chi-square	P-value
Total Fruits	2.87 (0.07)	0.11 (0.03)	0.37 (0.06)	0.64 (0.08)	1.41 (0.10)	2.80 (0.11)	4.70 (0.18)	5.00 (0.00)	5.00 (0.00)	5.00 (0.00)	2.74	3.01	1766.80	<.0001
Whole Fruits	3.16 (0.10)	0.03 (0.02)	0.22 (0.07)	0.50 (0.11)	1.48 (0.17)	3.47 (0.20)	5.00 (0.00)	5.00 (0.00)	5.00 (0.00)	5.00 (0.00)	2.96	3.36	1015.91	<.0001
Total Vegetables	3.12 (0.04)	0.99 (0.11)	1.37 (0.09)	1.62 (0.07)	2.15 (0.05)	2.98 (0.05)	4.09 (0.08)	5.00 (0.00)	5.00 (0.00)	5.00 (0.00)	3.04	3.20	6318.24	<.0001
Greens and Beans	2.45 (0.10)	0.21 (0.06)	0.45 (0.08)	0.64 (0.08)	1.15 (0.09)	2.10 (0.13)	3.67 (0.20)	5.00 (0.00)	5.00 (0.00)	5.00 (0.00)	2.26	2.64	646.42	<.0001
Whole Grains	2.95 (0.10)	0.13 (0.04)	0.44 (0.07)	0.73 (0.08)	1.46 (0.09)	2.56 (0.10)	3.96 (0.14)	5.66 (0.26)	6.95 (0.40)	10.00 (0.53)	2.74	3.15	831.88	<.0001
Dairy	6.52 (0.08)	1.40 (0.10)	2.38 (0.10)	3.05 (0.11)	4.45 (0.11)	6.45 (0.11)	8.95 (0.12)	10.00 (0.00)	10.00 (0.00)	10.00 (0.00)	6.35	6.69	6069.83	<.0001
Total Protein Foods	4.57 (0.04)	2.34 (0.23)	3.03 (0.17)	3.46 (0.13)	4.30 (0.09)	5.00 (0.00)	5.00 (0.00)	5.00 (0.00)	5.00 (0.00)	5.00 (0.00)	4.49	4.66	11116.77	<.0001
Seafood and Plant Proteins	3.62 (0.10)	0.57 (0.13)	1.07 (0.17)	1.47 (0.17)	2.40 (0.16)	3.98 (0.17)	5.00 (0.00)	5.00 (0.00)	5.00 (0.00)	5.00 (0.00)	3.43	3.81	1446.82	<.0001
Fatty Acids	4.62 (0.10)	0.00 (0.14)	1.07 (0.15)	1.70 (0.14)	2.89 (0.11)	4.40 (0.11)	6.15 (0.13)	7.94 (0.19)	9.13 (0.23)	10.00 (0.00)	4.41	4.83	1939.83	<.0001
Refined Grains	5.91 (0.10)	0.00 (0.00)	1.33 (0.26)	2.53 (0.18)	4.24 (0.13)	6.00 (0.10)	7.78 (0.12)	9.33 (0.18)	10.00 (0.01)	10.00 (0.00)	5.71	6.10	3708.96	<.0001
Sodium	4.36 (0.06)	0.00 (0.00)	0.00 (0.00)	0.71 (0.11)	2.64 (0.07)	4.49 (0.07)	6.14 (0.10)	7.48 (0.15)	8.23 (0.18)	9.55 (0.24)	4.24	4.49	4915.19	<.0001
Added Sugars	6.28 (0.08)	0.00 (0.00)	1.22 (0.31)	2.61 (0.25)	4.62 (0.15)	6.56 (0.07)	8.30 (0.09)	9.71 (0.11)	10.00 (0.00)	10.00 (0.00)	6.12	6.44	6147.98	<.0001
Saturated Fats	6.12	0.76	2.46	3.32	4.70	6.19	7.64	8.94	9.70	10.00	5.82	6.42	1673.48	<.0001

Mean HEI Score		Percentiles									95% CI		P-Value	
HEI Component	Score (se)	p1	p5	p10	p25	p50	p75	p90	p95	p99	Lower	Upper	Wald Chi-square	P-value
	(0.15)	(0.33)	(0.27)	(0.24)	(0.18)	(0.15)	(0.13)	(0.14)	(0.15)	(0.00)				
Total Score	56.55 (0.58)	32.46 (0.63)	39.18 (0.53)	42.81 (0.58)	49.17 (0.57)	56.40 (0.61)	63.84 (0.66)	70.64 (0.71)	74.51 (0.75)	80.95 (0.79)	55.38	57.72	9368.28	<.0001
Total Energy	2103 (12.6)	1041 (35.4)	1269 (27.6)	1406 (23.7)	1665 (17.4)	2011 (16.3)	2450 (20.8)	2942 (35.2)	3253 (52.4)	3851 (97.2)	2078	2129	27809.00	<.0001

The Wald chi-squares and associated p-values indicate whether the scores are significantly different from zero.

These are also saved in a pdf file called *hei_scores_SE.pdf*.

Congratulations you have just fit an MCMC model!

Reference:

Zhang S, Midthune D, Guenther PM, Krebs-Smith SM, Kipnis V, Dodd KW, Buckman DW, Tooze JA, Freedman L, Carroll RJ. A new multivariate measurement error model with zero-inflated dietary data, and its application to dietary assessment. *Annals of Applied Statistics*. 2011;5:1456-1487. PMID: PMC3145332