

# SAS<sup>1</sup> Macros for Fitting Multivariate Measurement Error Models and Estimating Multivariate Usual Intake Distributions

August 2013

Contents	Page
Section 1: Introduction	1
SAS Macros	2
Example SAS Programs	3
References	3
Section 2: Documentation for SAS Macro MULTIVAR_MCMC	4
Section 3: Documentation for SAS Macro STD_COV_BOXCOX24HR_CONDAY_MINAMT	12
Section 4: Documentation for SAS Macro MULTIVAR_DISTRIB	15
Section 5: Documentation for SAS Macro BRR_PVALUE_CI	19
Section 6: Example SAS Programs	22
Example 1a	22
Example 1b	22
Example 1c	23

## Section 1: Introduction

For dietary surveillance in the United States, the preferred method for collecting dietary intake data is the 24-hour dietary recall. However, researchers are interested in usual intake (i.e. long-term average daily intake). Researchers are also interested in exploring multiple dietary components collectively to consider dietary patterns, and consumption of these dietary components varies. Some dietary components are consumed almost daily by almost everyone and other dietary components are consumed episodically.

Zhang et al. (2011a, 2011b) developed an approach to simultaneously model multiple dietary components and to use this model for estimation of a multivariate usual intake distribution. These statistical methods also allow for a weighted analysis of survey data such as the National Health and Nutrition Examination Survey (NHANES). These statistical methods have been implemented in the SAS macros introduced in this section and described in detail in Sections 2 through 4. Section 5 describes a macro that performs balanced repeated replication (BRR) variance estimation and calculates p-values and confidence intervals for user specified parameters. Several example SAS programs that illustrate the methods using NHANES data are introduced in this section and described in more detail in Section 6.

The SAS macros and example programs can be downloaded from the following website: <http://appliedresearch.cancer.gov/diet/usualintakes/macros.html>.

Please send your comments and questions to [usualintake@imsweb.com](mailto:usualintake@imsweb.com). We will take this feedback into consideration when developing future versions of our software and other

---

<sup>1</sup>SAS Institute, Inc., Cary, NC, USA

resources. We will not be able to send a response to each user that submits a comment or question.

## **SAS Macros**

The MULTIVAR\_MCMC macro is used to fit a multivariate measurement error model that can include episodically consumed dietary components and non-episodically consumed dietary components. The measurement error model is fit using the Markov Chain Monte Carlo (MCMC) algorithm from Zhang et al. (2011a, 2011b). This methodology and macro also allow for survey-weighted MCMC computations. Output from this macro can be saved and used in subsequent analysis steps such as the estimation of multivariate usual intake distributions. The MULTIVAR\_MCMC macro requires SAS/IML version 9.2 or higher. If trace plots are requested, the MULTIVAR\_MCMC macro uses the SAS SGPANEL procedure available in Base SAS version 9.3 or in SAS/GRAPH version 9.2.

The input data for the MULTIVAR\_MCMC macro can be prepared using the STD\_COV\_BOXCOX24HR\_CONDAY\_MINAMT macro which is used to standardize continuous covariates, transform and standardize reported nonzero amounts (e.g. from 24-hour recalls), and create variables needed for the MULTIVAR\_MCMC macro. The STD\_COV\_BOXCOX24HR\_CONDAY\_MINAMT macro requires SAS/STAT.

The calculations performed by the STD\_COV\_BOXCOX24HR\_CONDAY\_MINAMT macro are summarized as follows:

1. Continuous covariates are standardized to have a mean of 0 and a variance of 1.
2. For dietary components consumed episodically, consumption-day indicator variables are created, and the following transformation approach, used by Zhang et al. (2011a, 2011b), is performed. First, a suitable Box-Cox transformation is applied to the nonzero values. Second, the Box-Cox transformed values are centered and scaled to obtain an approximate normal distribution with a mean of 0 and a variance of 2.
3. For dietary components that are consumed nearly every day, as specified by the user, zero values are replaced with minimum amount values provided by the user or calculated as half of the observed nonzero minimum amount. Then the following transformation approach, used by Zhang et al. (2011a, 2011b), is performed for the dietary components that are non-episodically consumed. First, a suitable Box-Cox transformation is applied. Second, the Box-Cox transformed values are centered and scaled to obtain an approximate normal distribution with a mean of 0 and a variance of 2.

The MULTIVAR\_DISTRIB macro uses parameter estimates from the MULTIVAR\_MCMC macro to generate a multivariate Monte Carlo distribution of the usual intakes for the dietary components specified in the multivariate measurement error model fit using the MULTIVAR\_MCMC macro. The MULTIVAR\_DISTRIB macro also uses covariates that were stored using the macro parameter "optional\_ims\_store\_names" of the MULTIVAR\_MCMC macro. The MULTIVAR\_DISTRIB macro allows specification of 1 or 2 lists of covariates. If 2 lists of covariates are specified, the usual intake is calculated as the weighted average of the usual intake calculated using covariate list 1 and the usual intake calculated using covariate list 2. This weighted average is calculated using the values specified for the "set\_value\_for\_weight\_cov\_list1" and "set\_value\_for\_weight\_cov\_list2" macro parameters. The MULTIVAR\_DISTRIB macro uses SAS/IML and uses results stored by the MULTIVAR\_MCMC macro in a SAS/IML catalog.

The BRR\_PVALUE\_CI macro performs balanced repeated replication (BRR) variance estimation, calculates a p-value for each user specified parameter to test the null hypothesis that the parameter value is zero, and calculates a confidence interval for each user specified parameter. The macro produces an output data set including the parameter estimates, the BRR standard error estimates, the p-value used to test the null hypothesis that the parameter value is zero versus the alternative hypothesis that the parameter value is not zero, and the lower and upper confidence limits. The macro input includes a data set with one record including parameter estimates from the original data and a data set with records 1, 2, ..., B including parameter estimates from replicate analyses performed for BRR variance estimation. Optionally, Fay's BRR method can be specified.

### **Example SAS Programs**

The Example 1a SAS program illustrates the use of the STD\_COV\_BOXCOX24HR\_CONDAY\_MINAMT and MULTIVAR\_MCMC macros to fit a multivariate measurement error model for the Healthy Eating Index-2010 (HEI-2010) using NHANES 2003-2004 data for smokers, age 20 and older. The Example 1b SAS program illustrates the use of the MULTIVAR\_DISTRIB macro to generate a multivariate Monte Carlo distribution of usual intakes and illustrates subsequent calculations of mean usual HEI-2010 scores using NHANES 2003-2004 data for smokers, age 20 and older. These programs, Examples 1a and 1b, can also be modified to fit a multivariate measurement error model and calculate mean usual HEI-2010 scores for nonsmokers, age 20 and older.

The HEI-2010 nonsmokers and smokers analysis uses balanced repeated replication (BRR) variance estimation, so the multivariate measurement error model is fit using the original weights (i.e. weights for replicate 0) and using 16 BRR weights (i.e. weights for replicates 1, 2, ..., 16). Similarly, the mean usual HEI-2010 scores are calculated using the original weights and using the 16 BRR weights. The Example 1a and 1b programs consider replicate 0; however, these programs allow specification of a range of replicates to facilitate the calculation of results required for BRR variance estimation. Performing the required replicate analysis steps requires additional computation time, so for convenience, Example 1c utilizes an input data set that includes the mean usual HEI-2010 scores calculated using the original weights (i.e. weights for replicate 0) and using 16 BRR weights (i.e. weights for replicates 1, 2, ..., 16) and includes the differences between these means for smokers and nonsmokers. The Example 1c SAS program illustrates the use of the BRR\_PVALUE\_CI macro to calculate the BRR standard errors for the mean usual HEI-2010 scores for nonsmokers and smokers and to perform t-tests comparing mean usual HEI-2010 scores for nonsmokers versus smokers.

### **References**

- Zhang S, Krebs-Smith SM, Midthune D, Perez A, Buckman DW, Kipnis V, Freedman LS, Dodd KW, Carroll RJ. Fitting a bivariate measurement error model for episodically consumed dietary components. *Int J Biostat* 2011;7(1):Article 1.
- Zhang S, Midthune D, Guenther PM, Krebs-Smith SM, Kipnis V, Dodd KW, Buckman DW, Toozé JA, Freedman L, Carroll RJ. A new multivariate measurement error model with zero-inflated dietary data, and its application to dietary assessment. *Ann Appl Stat* 2011 Jun;5(2B): 1456-87.

## Section 2: Documentation for SAS Macro MULTIVAR\_MCMC

```

/*****
/*****
/*
/*
/*          MULTIVAR_MCMC MACRO
/*
/*****
/*          VERSION 1.0          7/7/2013
/*
/*
/*
/* The MULTIVAR_MCMC macro is used to fit a multivariate measurement error
/* model that can include episodically consumed dietary components and
/* non-episodically consumed dietary components. The measurement error
/* model is fit using the Markov Chain Monte Carlo (MCMC) algorithm from
/* Zhang et al. (2011a, 2011b). This methodology and macro also allow for
/* survey-weighted MCMC computations. Output from this macro can be saved
/* and used in subsequent analysis steps such as the estimation of
/* multivariate usual intake distributions.
/*
/* The input data for this macro can be prepared using the
/* STD_COV_BOXCOX24HR_CONDAY_MINAMT macro which is used to standardize
/* continuous covariates, transform and standardize reported nonzero amounts
/* (e.g. from 24-hour recalls), and create variables needed for the
/* MULTIVAR_MCMC macro.
/*
/* The calculations performed by the STD_COV_BOXCOX24HR_CONDAY_MINAMT macro
/* are summarized as follows:
/*
/* 1. Continuous covariates are standardized to have a mean of 0 and a
/* variance of 1.
/*
/* 2. For dietary components consumed episodically, consumption-day
/* indicator variables are created, and the following transformation
/* approach, used by Zhang et al. (2011a, 2011b), is performed.
/* First, a suitable Box-Cox transformation is applied to the nonzero
/* values. Second, the Box-Cox transformed values are centered and
/* scaled to obtain an approximate normal distribution with a mean of
/* 0 and a variance of 2.
/*
/* 3. For dietary components that are consumed nearly every day, as
/* specified by the user, zero values are replaced with minimum
/* amount values provided by the user or calculated as half of the
/* observed nonzero minimum amount. Then the following
/* transformation approach, used by Zhang et al. (2011a, 2011b), is
/* performed for the dietary components that are non-episodically
/* consumed. First, a suitable Box-Cox transformation is applied.
/* Second, the Box-Cox transformed values are centered and scaled to
/* obtain an approximate normal distribution with a mean of 0 and a
/* variance of 2.
/*
/* The MULTIVAR_MCMC macro requires SAS IML version 9.2 or higher.
/*
/* References:
/*
/* Zhang S, Krebs-Smith SM, Midthune D, Perez A, Buckman DW, Kipnis V,
/* Freedman LS, Dodd KW, Carroll RJ. Fitting a bivariate measurement error
/* model for episodically consumed dietary components. Int J Biostat
/* 2011;7(1):Article 1.

```

```

/*                                                                 */
/* 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. Ann Appl Stat 2011 Jun;5(2B): */
/* 1456-87. */
/*                                                                 */
/*                                                                 */
/* The syntax for calling the MULTIVAR_MCMC macro is: */
/*                                                                 */
/* %multivar_mcmc(data =, */
/*                   subject =, */
/*                   weight_var =, */
/*                   repeat =, */
/*                   conday_epis_vars =, */
/*                   gst_rec24hr_epis_vars =, */
/*                   gst_rec24hr_daily_vars =, */
/*                   covars_epis_prob =, */
/*                   covars_epis_amt =, */
/*                   covars_daily_amt =, */
/*                   set_seed_mcmc =, */
/*                   set_number_mcmc_iterations =, */
/*                   set_number_burn_iterations =, */
/*                   set_thin =, */
/*                   gen_inverse =, */
/*                   print =, */
/*                   titles =, */
/*                   std_print_store =, */
/*                   notes_print =, */
/*                   out_lib =, */
/*                   out_store_label =, */
/*                   set_number_saved_out_data =, */
/*                   traceplots_method1_gpath =, */
/*                   traceplots_method2_file_pdf =, */
/*                   optional_ism_store_data =, */
/*                   optional_ism_store_names = */
/*                   ); */
/*                                                                 */
/* where */
/*                                                                 */
/* "data"           Specifies an input data set that includes */
/*                  one or more observations for each subject. */
/*                  An adequate number of subjects should have */
/*                  at least two days of reported consumption */
/*                  for episodically consumed dietary */
/*                  components, so covariance matrix */
/*                  parameters can be estimated. */
/*                                                                 */
/* "subject"       Specifies a variable that uniquely */
/*                  identifies each subject. */
/*                                                                 */
/* "weight_var"    Specifies a weight variable that is used */
/*                  to create a new variable "weight_nw_sumw" */
/*                  defined as: */
/*                  the number of subjects, n, multiplied */
/*                  by the input weight value and divided

```

```

/*          by the sum of the input weight values. */
/*          */
/* "repeat" Specifies a variable that indexes repeated */
/*          observations for each subject.          */
/*          */
/* "conday_epis_vars" Specifies a list of variables that */
/*          includes a consumption-day indicator */
/*          variable for each episodically consumed */
/*          dietary component.                      */
/*          */
/* "gst_rec24hr_epis_vars" Specifies a list of variables that */
/*          includes an appropriately transformed */
/*          24-hour recall variable for each */
/*          episodically consumed dietary component. */
/*          With the priors used in the MCMC, the */
/*          following transformation approach has been */
/*          used by Zhang et al. (2011a, 2011b). */
/*          First, a suitable Box-Cox transformation */
/*          is applied to the nonzero values.      */
/*          Second, the Box-Cox transformed values are */
/*          centered and scaled to obtain an */
/*          approximate normal distribution with a */
/*          mean of 0 and a variance of 2.         */
/*          */
/* "gst_rec24hr_daily_vars" Specifies a list of variables that */
/*          includes an appropriately transformed */
/*          24-hour recall variable for each daily */
/*          consumed dietary component.            */
/*          With the priors used in the MCMC, the */
/*          following transformation approach has been */
/*          used by Zhang et al. (2011a, 2011b). */
/*          First, a suitable Box-Cox transformation */
/*          is applied.                            */
/*          Second, the Box-Cox transformed values are */
/*          centered and scaled to obtain an */
/*          approximate normal distribution with a */
/*          mean of 0 and a variance of 2.         */
/*          */
/* "covars_epis_prob" Specifies a list of covariates for each of */
/*          the consumption-day indicator variables */
/*          for the episodically consumed dietary */
/*          components. If one list of covariates is */
/*          given, these covariates will be used for */
/*          all of the episodically consumed dietary */
/*          components. More generally, separate */
/*          lists of covariates can be specified for */
/*          each dietary component if each list is */
/*          preceded with the following keyword: */
/*          STARTCOVARLIST. To include an intercept, */
/*          the list of covariates should include a */
/*          constant such as the variable "constant1" */
/*          that is created by the macro and equals 1 */
/*          for each observation of the input data */
/*          set.                                    */
/*          */
/* "covars_epis_amt" Specifies a list of covariates for each of */

```

```

/* the transformed 24-hour recall variables */
/* for the episodically consumed dietary */
/* components. If one list of covariates is */
/* given, these covariates will be used for */
/* all of the episodically consumed dietary */
/* components. More generally, separate */
/* lists of covariates can be specified for */
/* each dietary component if each list is */
/* preceded with the following keyword: */
/* STARTCOVARLIST. To include an intercept, */
/* the list of covariates should include a */
/* constant such as the variable "constant1" */
/* that is created by the macro and equals 1 */
/* for each observation of the input data */
/* set. */
/*
/* "covars_daily_amt" Specifies a list of covariates for each of */
/* the transformed 24-hour recall variables */
/* for the daily consumed dietary components. */
/* If one list of covariates is given, these */
/* covariates will be used for all of the */
/* daily consumed dietary components. More */
/* generally, separate lists of covariates */
/* can be specified for each dietary */
/* component if each list is preceded with */
/* the following keyword: */
/* STARTCOVARLIST. To include an intercept, */
/* the list of covariates should include a */
/* constant such as the variable "constant1" */
/* that is created by the macro and equals 1 */
/* for each observation of the input data */
/* set. */
/*
/* "set_seed_mcmc" Specifies a seed for random number */
/* generation in the MCMC algorithm. If */
/* "set_seed_mcmc" is not specified, the SAS */
/* IML procedure will generate an initial */
/* seed value from the system clock. */
/*
/* "set_number_mcmc_iterations" Specifies the total number of iterations */
/* for the MCMC algorithm including the */
/* burn-in. The default value is 12000. */
/*
/* "set_number_burn_iterations" Specifies the burn-in value for the MCMC */
/* algorithm. The default value is 2000. */
/*
/* "set_thin" Specifies the thinning value for the MCMC */
/* algorithm. The default value is 25. */
/*
/* "gen_inverse" If "gen_inverse=n" or "gen_inverse=N" then */
/* the SAS IML "inv" function is used to */
/* compute the inverse of matrices as needed */
/* throughout the algorithm. This */
/* specification may save computational time, */
/* but the user may encounter numerical */
/* problems and an error message regarding a

```

```

/*          matrix that should be non-singular. The */
/*          default value is "y". By default, the SAS */
/*          IML "ginv" function is used to compute the */
/*          Moore-Penrose generalized inverse of */
/*          matrices as needed throughout the */
/*          algorithm. */
/*          */
/* "print"    If "print=n" or "print=N" then macro */
/*            results are not printed. The default */
/*            value is "y". */
/*          */
/* "titles"   Specifies the number of title lines to be */
/*            reserved for the user's titles. One */
/*            additional title line is used by the */
/*            macro. The default value is "0". */
/*          */
/* "std_print_store" If "std_print_store=y" or */
/*                  "std_print_store=Y" then sample standard */
/*                  deviations are printed and stored in the */
/*                  IML storage catalog. The default value is */
/*                  "n" because some analyses require more */
/*                  sophisticated methods of variance */
/*                  estimation. For example, replication */
/*                  methods such as balanced repeated */
/*                  replication (BRR) are used for analysis of */
/*                  data from a survey with complex sampling. */
/*          */
/* "notes_print" If "notes_print=n" or "notes_print=N" then */
/*               notes are not printed to the SAS log. The */
/*               default value is "y". */
/*          */
/* "out_lib"   Specifies a SAS library that the macro */
/*            uses when saving the IML storage catalog */
/*            file and the optional */
/*            "multivar_mcmc_samples_out" SAS data set. */
/*            The description for the "out_store_label" */
/*            macro parameter includes additional */
/*            details regarding the IML storage catalog. */
/*            The description for the */
/*            "set_number_saved_out_data" macro */
/*            parameter includes additional details */
/*            regarding the */
/*            "multivar_mcmc_samples_out" SAS data set. */
/*          */
/* "out_store_label" Specifies a label for the SAS IML storage */
/*                  catalog. The catalog name is defined by */
/*                  adding the prefix "iml" to the user */
/*                  supplied label. The catalog is stored in */
/*                  the library specified by the "out_lib" */
/*                  macro parameter. After deleting the */
/*                  current contents of the catalog, the */
/*                  following matrices (including vectors and */
/*                  scalars) are stored. */
/*          */
/*          "weight_nw_sumw" */
/*            The number of subjects, n, multiplied */
/*            by the input weight value and divided

```

```

/*          by the sum of the input weight values. */
/* "num_epis_diet_comp" */
/*          The number of episodically consumed */
/*          dietary components. */
/* "num_daily_diet_comp" */
/*          The number of daily consumed dietary */
/*          components. */
/* "num_rows_covmat" */
/*          The number of rows (or columns) of the */
/*          LxL covariance matrices, sigmau_mean */
/*          and sigmae_mean, where */
/*           $L = (2 * \text{num\_epis\_diet\_comp})$  */
/*           $+ \text{num\_daily\_diet\_comp}.$  */
/* "betaL_mean", ..., "betaL_mean" */
/* "sigmau_mean" */
/* "sigmae_mean" */
/*          Sample means calculated using */
/*          generated values from MCMC iterations */
/*          selected according to the */
/*          "set_number_burn_iterations" and */
/*          "set_thin" macro parameters. The rows */
/*          of the LxL covariance matrices, */
/*          sigmau_mean and sigmae_mean, are */
/*          arranged so the episodically consumed */
/*          dietary components precede the daily */
/*          consumed dietary components. For */
/*          example, 3 episodically consumed */
/*          dietary components and 2 daily */
/*          consumed dietary components yield */
/*           $(2*3 + 2 = 8 = L)$  rows sorted as: */
/*          1. consumption probability (epis 1), */
/*          2. amount consumed (epis 1), */
/*          3. consumption probability (epis 2), */
/*          4. amount consumed (epis 2), */
/*          5. consumption probability (epis 3), */
/*          6. amount consumed (epis 3), */
/*          7. amount consumed (daily 1), */
/*          8. amount consumed (daily 2), */
/* "betaL_std", ..., "betaL_std" */
/* "sigmau_std" */
/* "sigmae_std" */
/*          Sample standard deviations calculated */
/*          using generated values from MCMC */
/*          iterations selected according to the */
/*          "set_number_burn_iterations" and */
/*          "set_thin" macro parameters. These */
/*          sample standard deviations are stored */
/*          when "std_print_store=y" or */
/*          "std_print_store=Y". */
/* "set_number_saved_out_data" Specifies the number of iterations to save */
/*          in a SAS output data set */
/*          "multivar_mcmc_samples_out" that is */
/*          saved in the library specified by the */
/*          "out_lib" macro parameter. The data set */
/*          includes MCMC samples for the Beta,

```

```

/*          Sigma_u, and Sigma_e parameters with          */
/*          values selected from among the iterations    */
/*          specified according to the                   */
/*          "set_number_burn_iterations" and             */
/*          "set_thin" macro parameters. The default   */
/*          value is 0.                                  */
/*
/* "traceplots_method1_gpath" Specifies a valid SAS fileref to indicate
/* a folder used for storing the MCMC trace
/* plot files. The fileref is used by the
/* "gpath" option of the "ods listing"
/* statement in the SAS Output Delivery
/* System (ODS). Each trace plot file
/* includes up to 4 trace plots
/* (i.e. 4 panels). Example file names
/* include:
/*          "Beta_Trace_Plot_Panels4_Image.png"
/*          "Sigma_u_Trace_Plot_Panels4_Image.png"
/*          "Sigma_e_Trace_Plot_Panels4_Image.png"
/* and SAS ODS will add an image index number
/* to ensure that each file produced has a
/* unique name. MCMC trace plots are
/* produced for all of the Beta, Sigma_u, and
/* Sigma_e parameters. The plots include
/* MCMC samples selected according to the
/* "set_number_burn_iterations" and
/* "set_thin" macro parameters. If no value
/* is specified, these trace plot files are
/* not produced. Trace plots can be produced
/* using another method as described for the
/* "traceplots_method2_file_pdf" macro
/* parameter.
/*
/* "traceplots_method2_file_pdf" Specifies a file name with a "pdf"
/* extension, such as
/* "example1.traceplot.pdf". This pdf file
/* will include the MCMC trace plots for all
/* of the Beta, Sigma_u, and Sigma_e
/* parameters. The plots include MCMC
/* samples selected according to the
/* "set_number_burn_iterations" and
/* "set_thin" macro parameters. If no value
/* is specified, this trace plot file is not
/* produced. Trace plots can be produced
/* using another method as described for the
/* "traceplots_method1_gpath" macro
/* parameter. The pdf file produced by trace
/* plot method 2 tends to have a much larger
/* file size when compared to the set of
/* files produced by trace plot method 1.
/*
/* "optional_iml_store_data" Specifies an optional SAS data set that
/* includes useful variables to be stored in
/* the IML storage catalog. The names of
/* these variables must be specified in the
/* "optional_iml_store_names" macro
/*

```

```
/*          parameter.          */
/*          */
/* "optional_ims_store_names" Specifies a list of optional matrices */
/*          (including vectors and scalars) to be          */
/*          stored in the IML storage catalog. This          */
/*          macro parameter can be used in conjunction      */
/*          with the "optional_ims_store_data" macro        */
/*          parameter, and it can be used by advanced      */
/*          users that want to save values that are        */
/*          defined in the SAS IML procedure within        */
/*          the macro.          */
/*          */
/*****/
/*****/
/*****/
```

### Section 3: Documentation for SAS Macro STD\_COV\_BOXCOX24HR\_CONDAY\_MINAMT

```

/*****
/*****
/*
/*          STD_COV_BOXCOX24HR_CONDAY_MINAMT MACRO          */
/*
/*****
/*          VERSION 1.0          7/7/2013          */
/*
/*
/*
/* The STD_COV_BOXCOX24HR_CONDAY_MINAMT macro is used to standardize
/* continuous covariates, transform and standardize reported nonzero amounts
/* (e.g. from 24-hour recalls), and create variables needed for the
/* MULTIVAR_MCMC macro.
/*
/* Continuous covariates are standardized to have a mean of 0 and a variance
/* of 1, and these new standardized variables are named by adding the prefix
/* "std_" to the original variable name.
/*
/* For dietary components consumed episodically, consumption-day indicator
/* variables are created, and these new variables are named by adding the
/* prefix "conday_" to the original variable name. Also, the following
/* transformation approach, used by Zhang et al. (2011a, 2011b), is
/* performed. First, a suitable Box-Cox transformation is applied to the
/* nonzero values. Second, the Box-Cox transformed values are centered and
/* scaled to obtain an approximate normal distribution with a mean of 0 and
/* a variance of 2. These new variables are named by adding the prefix
/* "stdbc_" to the original variable name.
/*
/* For dietary components that are consumed nearly every day, as specified
/* by the user, zero values are replaced with minimum amount values provided
/* by the user or calculated as half of the observed nonzero minimum amount.
/* Then the following transformation approach, used by Zhang et al. (2011a,
/* 2011b), is performed for the dietary components that are non-episodically
/* consumed. First, a suitable Box-Cox transformation is applied. Second,
/* the Box-Cox transformed values are centered and scaled to obtain an
/* approximate normal distribution with a mean of 0 and a variance of 2.
/* These new variables are named by adding the prefix "stdbc_" to the
/* original variable name.
/*
/* References:
/*
/* Zhang S, Krebs-Smith SM, Midthune D, Perez A, Buckman DW, Kipnis V,
/* Freedman LS, Dodd KW, Carroll RJ. Fitting a bivariate measurement error
/* model for episodically consumed dietary components. Int J Biostat
/* 2011;7(1):Article 1.
/*
/* 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. Ann Appl Stat 2011 Jun;5(2B):
/* 1456-87.
/*
/*
/* The syntax for calling the STD_COV_BOXCOX24HR_CONDAY_MINAMT macro is:

```

```

/*                                                                    */
/* %std_cov_boxcox24hr_conday_minamt(data                            =,    */
/*                                prestand_continuous_covars         =,    */
/*                                rec24hr_epis_vars                  =,    */
/*                                rec24hr_daily_vars                  =,    */
/*                                boxcox_tran_lambda_data            =,    */
/*                                minamount_data                    =,    */
/*                                print                               =,    */
/*                                titles                             =     */
/*                                );                                  */
/*                                                                    */
/* where                                                                */
/*                                                                    */
/* "data"                      Specifies an input data set that includes */
/*                              one or more observations for each subject. */
/*                                                                    */
/* "prestand_continuous_covars" Specifies a list of continuous covariates */
/*                              that will be standardized to have a mean  */
/*                              of 0 and a variance of 1.                  */
/*                                                                    */
/* "rec24hr_epis_vars"         Specifies a list of 24-hour recall        */
/*                              variables for dietary components consumed  */
/*                              episodically.                              */
/*                                                                    */
/* "rec24hr_daily_vars"        Specifies a list of 24-hour recall        */
/*                              variables for dietary components consumed  */
/*                              every day or nearly every day.           */
/*                                                                    */
/* "boxcox_tran_lambda_data"   Specifies an input data set that includes */
/*                              the following two variables:              */
/*                              "tran_paramindex"                       */
/*                              an index value of 1, 2, ..., Q where     */
/*                              Q is the total number of dietary          */
/*                              components specified. The values 1,     */
/*                              2, ..., Q should be assigned according    */
/*                              to the order of the dietary components   */
/*                              as specified for the                      */
/*                              "rec24hr_epis_vars" and                  */
/*                              "rec24hr_daily_vars" macro parameters,  */
/*                              and the "rec24hr_epis_vars" dietary      */
/*                              components should precede the            */
/*                              "rec24hr_daily_vars" dietary             */
/*                              components.                               */
/*                              "tran_lambda"                             */
/*                              a Box-Cox transformation parameter       */
/*                              value for the corresponding dietary       */
/*                              component.                               */
/*                              The records in this data set should be   */
/*                              sorted according to the order described  */
/*                              for the "tran_paramindex" variable.      */
/*                                                                    */
/* "minamount_data"           Specifies an optional input data set that  */
/*                              includes the following two variables:    */
/*                              "tran_paramindex"                       */
/*                              an index value of 1, 2, ..., Q where     */
/*                              Q is the total number of dietary          */

```

```

/*          components specified.  The values 1,      */
/*          2, ..., Q should be assigned according  */
/*          to the order described for the          */
/*          "boxcox_tran_lambda_data" macro        */
/*          parameter.                             */
/*          "minamount"                            */
/*          a minimum amount value for the         */
/*          corresponding dietary component.       */
/*          The records in this data set should be */
/*          sorted according to the order described */
/*          for the "tran_paramindex" variable.    */
/*          */
/* "print"    If "print=y" or "print=Y" then macro  */
/*            results are printed.  The default value is */
/*            "y".                                       */
/*          */
/* "titles"  Specifies the number of title lines to be */
/*            reserved for the user's titles.  Two    */
/*            additional title lines are used by the  */
/*            macro.  The default value is "0".      */
/*          */
/*****
/*
/* Macro Output:
/*
/* The new variables and the original variables are saved in a SAS data set
/* named "stdcov_stdbc24hr_conday_out" which can be used as the input data
/* set for the MULTIVAR_MCMC macro.
/*
/* The following global macro variables are declared and used to output
/* variable lists that can be used to specify the lists of input variables
/* needed for the MULTIVAR_MCMC macro:
/*   conday_var_list
/*   stdbc_epis_var_list
/*   stdbc_daily_var_list
/*   std_continuous_covar_list.
/*
/* The macro also saves the following variables in a SAS data set named
/* "backtran_out" which can be used in subsequent analysis steps that
/* require back-transformation:
/*   tran_paramindex tran_lambda tran_center tran_scale minamount.
/*
/*****
/*****
/*****

```

## Section 4: Documentation for SAS Macro MULTIVAR\_DISTRIB

```

/*****
/*****
/*
/*
/*          MULTIVAR_DISTRIB MACRO
/*
/*****
/*          VERSION 1.0          7/11/2013
/*
/*
/*
/* The MULTIVAR_DISTRIB macro uses parameter estimates from the
/* MULTIVAR_MCMC macro to generate a multivariate Monte Carlo distribution
/* of the usual intakes for the dietary components specified in the
/* multivariate measurement error model fit using the MULTIVAR_MCMC macro.
/* The MULTIVAR_DISTRIB macro also uses covariates that were stored using
/* the macro parameter "optional_iml_store_names" of the MULTIVAR_MCMC
/* macro. The MULTIVAR_DISTRIB macro allows specification of 1 or 2 lists
/* of covariates. If 2 lists of covariates are specified, the usual intake
/* is calculated as the weighted average of the usual intake calculated
/* using covariate list 1 and the usual intake calculated using covariate
/* list 2. This weighted average is calculated using the values specified
/* for the "set_value_for_weight_cov_list1" and
/* "set_value_for_weight_cov_list2" macro parameters.
/*
/*
/* References:
/*
/* Zhang S, Krebs-Smith SM, Midthune D, Perez A, Buckman DW, Kipnis V,
/* Freedman LS, Dodd KW, Carroll RJ. Fitting a bivariate measurement error
/* model for episodically consumed dietary components. Int J Biostat
/* 2011;7(1):Article 1.
/*
/* 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. Ann Appl Stat 2011 Jun;5(2B):
/* 1456-87.
/*
/*
/* The syntax for calling the MULTIVAR_DISTRIB macro is:
/*
/* %multivar_distrib(multivar_mcmc_out_lib          =,
/*                   multivar_mcmc_out_store_label =,
/*                   t_weightavg_covariates_list1  =,
/*                   t_weightavg_covariates_list2  =,
/*                   set_value_for_weight_cov_list1 =,
/*                   set_value_for_weight_cov_list2 =,
/*                   set_seed_distrib              =,
/*                   set_number_monte_carlo_rand_obs =,
/*                   print                          =
/*                   );
/*
/* where
/*
/* "multivar_mcmc_out_lib"          Specifies the same SAS library that
/*                                  was specified for the "out_lib" macro

```

```

/* parameter of the MULTIVAR_MCMC macro. */
/* The MULTIVAR_DISTRIB macro will use */
/* the parameter estimates and */
/* covariates that were stored by the */
/* MULTIVAR_MCMC macro. Covariates can */
/* be stored using the */
/* "optional_aml_store_names" macro */
/* parameter of the MULTIVAR_MCMC macro. */
/*
/* "multivar_mcmc_out_store_label" Specifies the same label that was */
/* specified for the "out_store_label" */
/* macro parameter of the MULTIVAR_MCMC */
/* macro. The MULTIVAR_DISTRIB macro */
/* will use the parameter estimates and */
/* covariates that were stored by the */
/* MULTIVAR_MCMC macro. Covariates can */
/* be stored using the */
/* "optional_aml_store_names" macro */
/* parameter of the MULTIVAR_MCMC macro. */
/*
/* "t_weightavg_covariates_list1" Specifies a list of covariates that */
/* correspond to the beta parameter */
/* estimates from the multivariate */
/* measurement error model fit using the */
/* MULTIVAR_MCMC macro. The */
/* calculations in the MULTIVAR_DISTRIB */
/* macro are based on the assumption */
/* that the multivariate measurement */
/* error model was fit using the same */
/* set of covariates for each part of */
/* the model. The MULTIVAR_DISTRIB */
/* macro creates "constant0" which can */
/* be used as a variable that equals 0 */
/* for each observation, and the macro */
/* creates "constant1" which can be used */
/* as a variable that equals 1 for each */
/* observation. "Constant0" and */
/* "constant1" can be used when */
/* specifying covariates of interest. */
/* If the optional macro parameter */
/* "t_weightavg_covariates_list2" is */
/* used to specify a second list of */
/* covariates, then usual intake is */
/* calculated as the weighted average of */
/* the usual intake calculated using */
/* covariate list 1 and the usual intake */
/* calculated using covariate list 2. */
/* This weighted average is calculated */
/* using the values specified for the */
/* "set_value_for_weight_cov_list1" and */
/* "set_value_for_weight_cov_list2" */
/* macro parameters. If the second list */
/* of covariates is not specified, then */
/* the Monte Carlo distribution of usual */
/* intakes is only based on the */
/* covariates specified for the */

```

```

/*          "t_weightavg_covariates_list1" macro */
/*          parameter. */
/*          */
/* "t_weightavg_covariates_list2" Specifies an optional second list of */
/* covariates. If this list of */
/* covariates is specified, the usual */
/* intake is calculated as the weighted */
/* average of the usual intake */
/* calculated using covariate list 1 and */
/* the usual intake calculated using */
/* covariate list 2. This weighted */
/* average is calculated using the */
/* values specified for the */
/* "set_value_for_weight_cov_list1" and */
/* "set_value_for_weight_cov_list2" */
/* macro parameters. */
/*          */
/* "set_value_for_weight_cov_list1" Specifies a value that is used to */
/* calculate a weighted average when 2 */
/* covariate lists are provided as */
/* described for the */
/* "t_weightavg_covariates_list1" and */
/* "t_weightavg_covariates_list2" macro */
/* parameters. The value specified is */
/* used as the numerator of the weight. */
/* The denominator of the weight is */
/* defined as the sum of the values from */
/* the "set_value_for_weight_cov_list1" */
/* and "set_value_for_weight_cov_list2" */
/* macro parameters. The default value */
/* is 1 which corresponds to the */
/* situation where the */
/* "t_weightavg_covariates_list1" macro */
/* parameter is used to provide */
/* covariate list 1 and the */
/* "t_weightavg_covariates_list2" macro */
/* parameter is not used. */
/*          */
/* "set_value_for_weight_cov_list2" Specifies a value that is used to */
/* calculate a weighted average when 2 */
/* covariate lists are provided as */
/* described for the */
/* "t_weightavg_covariates_list1" and */
/* "t_weightavg_covariates_list2" macro */
/* parameters. The value specified is */
/* used as the numerator of the weight. */
/* The denominator of the weight is */
/* defined as the sum of the values from */
/* the "set_value_for_weight_cov_list1" */
/* and "set_value_for_weight_cov_list2" */
/* macro parameters. The default value */
/* is 0 which corresponds to the */
/* situation where the */
/* "t_weightavg_covariates_list2" macro */
/* parameter is not used. */
/*          */

```

```

/* "set_seed_distrib"                Specifies a seed for random number */
/*                                  generation for the Monte Carlo method */
/*                                  used in the macro. If */
/*                                  "set_seed_distrib" is not specified, */
/*                                  the SAS IML procedure will generate */
/*                                  an initial seed value from the system */
/*                                  clock. */
/*                                  */
/* "set_number_monte_carlo_rand_obs" Specifies the number of observations */
/*                                  of the random effects vector to */
/*                                  generate for each individual. The */
/*                                  number of observations in the Monte */
/*                                  Carlo data set is the product of the */
/*                                  "number of individuals" and the */
/*                                  specified value for the */
/*                                  "set_number_monte_carlo_rand_obs" */
/*                                  macro parameter. The default value */
/*                                  is 500. */
/*                                  */
/* "print"                            If "print=n" or "print=N" then macro */
/*                                  information is not printed. */
/*                                  Otherwise the macro prints */
/*                                  information regarding the IML storage */
/*                                  catalog used and prints some stored */
/*                                  and specified values. The default */
/*                                  value is "y". */
/*                                  */
/***** */
/* Macro Output Data Set:  mc_t_distrib_out */
/* The MULTIVAR_DISTRIB macro produces a SAS data set "mc_t_distrib_out" */
/* that includes usual intake variables named: */
/*      mc_t1  mc_t2  ...  mc_tp */
/* where p = num_epis_diet_comp + num_daily_diet_comp, and the dietary */
/* component order (1, 2, ..., p) is equivalent to the order used when the */
/* multivariate measurement error model was fit using the MULTIVAR_MCMC */
/* macro. The "mc_t_distrib_out" data set also includes the variable: */
/*      weight_nw_sumw */
/* created and stored by the MULTIVAR_MCMC macro (see the MULTIVAR_MCMC */
/* "weight_var" macro parameter for details). */
/* The number of observations in this output data set is the product of the */
/* "number of individuals" and the specified value for the */
/* "set_number_monte_carlo_rand_obs" macro parameter. */
/*****
/*****
/*****

```

## Section 5: Documentation for SAS Macro BRR\_PVALUE\_CI

```

/*****
/*****
/*
/*
/*          BRR_PVALUE_CI MACRO
/*
/*****
/*          VERSION 1.0          8/12/2013
/*
/*
/* The BRR_PVALUE_CI macro performs balanced repeated replication (BRR)
/* variance estimation, calculates a p-value for each user specified
/* parameter to test the null hypothesis that the parameter value is zero,
/* and calculates a confidence interval for each user specified parameter.
/* The macro produces an output data set including the parameter estimates,
/* the BRR standard error estimates, the p-value used to test the null
/* hypothesis that the parameter value is zero versus the alternative
/* hypothesis that the parameter value is not zero, and the lower and upper
/* confidence limits.
/*
/* The macro input includes a data set with one record including parameter
/* estimates from the original data and a data set with records 1, 2, ..., B
/* including parameter estimates from replicate analyses performed for BRR
/* variance estimation.  Optionally, Fay's BRR method can be specified.
/*
/* Reference:
/*
/* Korn EL, Graubard BI. Analysis of Health Surveys, John Wiley & Sons,
/* Inc., New York, 1999.
/*
/* The syntax for calling the BRR_PVALUE_CI macro is:
/*
/* %brr_pvalue_ci(data_orig_estimates_lrec =,
/*                data_brr_estimates_brecs =,
/*                param_estimate_names     =,
/*                set_f_method_fay         =,
/*                set_confidence_level     =,
/*                set_degrees_freedom      =,
/*                print                     =,
/*                titles                    =
/*                );
/*
/* where
/*
/* "data_orig_estimates_lrec"  Specifies an input data set that includes
/*                             one record with parameter estimates from
/*                             the original data.
/*
/* "data_brr_estimates_brecs" Specifies an input data set that includes
/*                             records 1, 2, ..., B with parameter
/*                             estimates from replicate analyses
/*                             performed for balanced repeated
/*                             replication (BRR) variance estimation.
/*                             For example, record 1 includes the
/*                             parameter estimates obtained from

```





## Section 6: Example SAS Programs

The examples discussed in this section can be downloaded from the following website: <http://riskfactor.cancer.gov/diet/usualintakes/macros.html>. These examples are introduced in Section 1 and described in detail in this section.

### Example 1a

The Example 1a SAS program illustrates the use of the `STD_COV_BOXCOX24HR_CONDAY_MINAMT` and `MULTIVAR_MCMC` macros to fit a multivariate measurement error model for the Healthy Eating Index-2010 (HEI-2010) using NHANES 2003-2004 data for smokers, age 20 and older. This program can also be modified to fit a multivariate measurement error model for nonsmokers, age 20 and older.

The HEI-2010 nonsmokers and smokers analysis uses balanced repeated replication (BRR) variance estimation, so the multivariate measurement error model is fit using the original weights (i.e. weights for replicate 0) and using 16 BRR weights (i.e. weights for replicates 1, 2, ..., 16). The Example 1a program includes the following `replicfirst` and `repliclast` macro variables which allow specification of a range of replicates; although, this example program only considers replicate 0:

```
%let replicfirst = 0;
%let repliclast  = 0;.
```

To specify a range of replicates, such as replicates 1, 2, 3, and 4, the following specification would be used:

```
%let replicfirst = 1;
%let repliclast  = 4;.
```

This example requires SAS/IML version 9.2 or higher and the following files:

- `std_cov_boxcox24hr_conday_minamt_macro_v1.0.sas`
- `multivar_mcmc_macro_v1.0.sas`
- `data.hei2010.nonsmok.smok.ag20p.xpt`
- `bclam.hei2010.nonsmok.smok.ag20p.xpt`.

The example program uses the SAS `CIMPORT` procedure to import the data file and the file of transformation parameter values. The example program uses the `STD_COV_BOXCOX24HR_CONDAY_MINAMT` macro to transform and standardize reported nonzero amounts and create variables needed for the `MULTIVAR_MCMC` macro. Within the Example 1a program, another SAS macro, the `FIT_MODELS_REPLICATE_LOOP` macro, is defined and used to call the `MULTIVAR_MCMC` macro for each of the replicates specified in the definitions of the `replicfirst` and `repliclast` macro variables. Results from Example 1a are saved and used in Example 1b.

### Example 1b

The Example 1b SAS program illustrates the use of the `MULTIVAR_DISTRIB` macro to generate a multivariate Monte Carlo distribution of usual intakes and illustrates subsequent calculations of mean usual Healthy Eating Index-2010 (HEI-2010) scores using NHANES 2003-2004 data for smokers, age 20 and older. This program can also be modified to calculate mean usual HEI-2010 scores for nonsmokers, age 20 and older.

The HEI-2010 nonsmokers and smokers analysis uses balanced repeated replication (BRR) variance estimation, so the mean usual HEI-2010 scores are calculated using the original weights (i.e. weights for replicate 0) and using 16 BRR weights (i.e. weights for replicates 1, 2, ..., 16). The Example 1b program includes the following `replicfirst` and `repliclast` macro variables which allow specification of a range of replicates; although, this example program only considers replicate 0:

```
%let replicfirst = 0;  
%let repliclast  = 0;.
```

To consider a range of replicates, additional SAS programming would allow the results to be saved and utilized efficiently. Example 1b is a simplified program developed for illustrative purposes.

This example requires the following stored results and files:

- results from example 1a stored in a SAS/IML catalog
- `multivar_distrib_macro_v1.0.sas`
- `hei2010.beanspeas.allocation.macro.sas`  
(see <http://riskfactor.cancer.gov/tools/hei/>)
- `hei2010.score.macro.sas`  
(see <http://riskfactor.cancer.gov/tools/hei/>).

The Example 1b program will utilize results stored from Example 1a, so Example 1a should be completed first. Within the Example 1b program, another SAS macro, the `DISTRIB_HEI_REPLICATE_LOOP` macro, is defined and used to call the `MULTIVAR_DISTRIB` macro and calculate mean usual HEI-2010 scores for each of the replicates specified in the above definition of the `replicfirst` and `repliclast` macro variables.

The macro call for the `MULTIVAR_DISTRIB` macro includes two lists of covariates defined such that covariate list 1 specifies a week day by setting the weekend indicator covariate to a value of 0 and covariate list 2 specifies a weekend by setting the weekend indicator covariate to a value of 1. Additional details are provided in the Example 1b program. Using the multivariate Monte Carlo distribution from the `MULTIVAR_DISTRIB` macro, the Example 1b program uses a SAS macro, `hei2010.beanspeas.allocation.macro.sas` (see <http://riskfactor.cancer.gov/tools/hei/>) to allocate beans and peas (i.e. legumes). Then, the Example 1b program calculates the usual HEI-2010 scores using the multivariate Monte Carlo distribution of usual intakes and a SAS macro, `hei2010.score.macro.sas` (see <http://riskfactor.cancer.gov/tools/hei/>). Finally, the mean usual HEI-2010 scores are calculated.

Since the `MULTIVAR_MCMC` and the `MULTIVAR_DISTRIB` macros use algorithms that require random number generation, these macros will produce slightly different results depending on the starting seed values used. This concept is illustrated by comparison of the mean usual HEI-2010 scores for smokers produced in Example 1b for replicate 0 with the corresponding values used in the input data set for Example 1c.

### **Example 1c**

The Example 1c SAS program illustrates the use of the `BRR_PVALUE_CI` macro to calculate the standard errors for the mean usual HEI-2010 scores for nonsmokers and smokers and to perform t-tests comparing mean usual HEI-2010 scores for nonsmokers versus smokers. The input data set includes the mean usual HEI-2010 scores calculated using NHANES 2003-2004 data for nonsmokers and smokers, age 20 and older. The HEI-2010 nonsmokers and smokers analysis uses balanced repeated replication (BRR)

variance estimation. The input data set includes mean usual HEI-2010 scores calculated using the original weights (i.e. weights for replicate 0) and using 16 BRR weights (i.e. weights for replicates 1, 2, ..., 16) and includes the differences between these means for smokers and nonsmokers. Examples 1a and 1b illustrate the calculation of the mean usual HEI-2010 scores.

This example requires the following files:

- `brr_pvalue_ci_macro_v1.0.sas`
- `means.hei2010.nonsmok.smok.ag20p.xpt.`

The Example 1c program uses the SAS CIMPORT procedure to import the data file of intermediate results needed for this example. Using this input data, the example program creates a data set with one record including parameter estimates from the original data and a data set with records 1, 2, ..., B including parameter estimates from replicate analyses performed for BRR variance estimation.

The Example 1c program deletes replicates 12 and 14 due to unusable parameter estimates of covariance matrix parameters in the multivariate measurement error model for smokers. Fourteen replicates are available for the analysis, and the `BRR_PVALUE_CI` macro call specifies the value 14 for the "set\_degrees\_freedom" macro parameter. If all 16 replicates had yielded usable results, the degrees of freedom would have been limited to 15 since the NHANES 2003-2004 data includes 30 primary sampling units (PSUs) and 15 strata (i.e.  $30-15=15$ ). The degrees of freedom associated with replication variance estimators is discussed by Korn and Graubard (1999, *Analysis of Health Surveys*, pg. 34) and in the documentation for the SAS survey procedures.

The Example 1c output includes the BRR standard error estimates for the mean usual HEI-2010 scores for smokers, the BRR standard error estimates for the mean usual HEI-2010 scores for nonsmokers, and the p-values for the t-tests comparing mean usual HEI-2010 scores for nonsmokers versus smokers. These t-test p-values are printed along with the BRR standard error estimates for the differences between these means for smokers and nonsmokers.