SAS Macros for Estimating Ratios of Usual Intakes of Dietary Components

Contents: Page

Section 1: Introduction

The primary dietary assessment method used in dietary surveillance is the 24-hour dietary recall. Since 24-hour dietary recalls measure intake on a given day, they have substantial day-to-day variation. Researchers, however, are often interested in the *usual* intake of a dietary component, generally defined as the long-term daily average intake. To estimate population distributions of *usual* intakes from 24-hour dietary recalls, special methods are needed.

In addition, it is often of interest to estimate population distributions of the ratios of usual intakes of dietary components. Examples include the percent of usual energy intake from saturated fat and the usual intake of sodium (mg) per 1000 kcal of usual energy intake. Estimating such distributions from 24-hour dietary recalls also requires special methods.

A set of five SAS macros has been developed to estimate population distributions of ratios of usual intakes of dietary components that are consumed every day from repeated 24-hour dietary recalls. Example programs showing how to use the macros are also provided.

The macros fit a bivariate nonlinear mixed model for two dietary components, use Monte Carlo methods to create a bivariate distribution of usual intake of the two components, and calculate percentiles of the population distribution of the ratio of usual intakes. The macros allow specification of sampling weights for complex survey data. Two or more 24-hour recalls per subject are required for at least a portion of the sample.

The macros call the SAS NLMIXED and IML procedures, so SAS/STAT and SAS/IML are required.

For a complete description of the methodology, see:

> Freedman LS, et al. The population distribution of ratios of usual intakes of dietary components that are consumed every day can be estimated from repeated 24-hour recalls. J Nutr. 2010;140:111-116.

Note that the macros can also be used to estimate ratios of usual intakes of dietary components when one of the components is episodically consumed and the other is consumed every day. A manuscript describing this methodology has been prepared but has not yet been published. We do not recommend using the macros for this purpose at this time, and we have not provided any example programs showing how to do so.

SAS Macros

The SAS macros are briefly described here. More detailed information is provided in the following sections.

1) NLMixed_Univariate (nlmixed_univariate.macro.v1.1.sas): SAS macro NLMixed_Univariate fits a univariate nonlinear mixed model for a single dietary

component. The primary purpose of this macro is to provide initial parameter estimates for macro NLMixed_Bivariate, which fits a bivariate model for two dietary components.

2) NLMixed_Bivariate (nlmixed_bivariate.macro.v1.1.sas): SAS macro NLMixed_Bivariate fits a bivariate nonlinear mixed model for two dietary components. The macro estimates parameters of the distribution of usual intakes, which are then used by macro Distrib_Bivariate to create a Monte Carlo distribution of usual intakes. The macro requires the user to provide initital parameter estimates, which can be obtained using macro NLMixed_Univariate.

3) Distrib_Bivariate (distrib_bivariate.macro.v1.1.sas): SAS macro Distrib_Bivariate creates a Monte Carlo (bivariate) distribution of the usual intakes of two dietary components. The macro requires the user to provide parameters of the distribution of usual intakes, which can be obtained using macro NLMixed_Bivariate.

4) Percentiles_Survey (percentiles_survey.macro.v1.1.sas): SAS macro Percentiles_Survey calculates the mean, standard deviation and percentiles of the population distribution of usual intake. It can also calculate the probability of being below a user-specified cut-point in the distribution. The macro requires the user to provide a population distribution of usual intake, which can be obtained using macro Distrib_Bivariate.

5) BoxCox_Survey (boxcox_survey.macro.v1.1.sas): Supplementary macro that chooses a Box-Cox transformation to transform a variable to approximate normality. The macro chooses the Box-Cox transformation that minimizes the mean square error about the best straight line on a QQ plot.

Example SAS Programs

The example programs show how to call the macros to estimate the population distribution of the percent of usual energy from saturated fat (i.e., 100 × {usual saturated fat intake (kcal)} / {usual total energy intake (kcal)}. The programs use data from the National Health and Nutrition Examination Survey (NHANES), 2001-2004. The data can be found at: www.riskfactor.cancer.gov/diet/usualintakes/macros.html

Note that there are “libname” and “%include” statements at the beginning of each program. These statements must be modified to refer to the appropriate data libraries and macros.

To estimate the population distribution of percent of usual energy from saturated fat, run the following four programs in the following order:

1) example1.nlmixed_univariate.sfat.sas – This program calls macro BoxCox_Survey to calculate the best Box-Cox transformation for saturated fat intake (g/day), and then calls macro NLMixed_Univariate to fit a univariate nonlinear mixed model for saturated fat. The estimated model parameters will be used as initial parameter estimates for a bivariate model.

2) example2.nlmixed_univariate.energy.sas – This program calls macro BoxCox_Survey to calculate the best Box-Cox transformation for energy intake (kcal/day), and then calls macro NLMixed_Univariate to fit a univariate nonlinear mixed model for energy. The estimated parameters will be used as initial parameter estimates for a bivariate model.

3) example3.nlmixed_bivariate.sfat.energy.sas – This program calls macro NLMixed_Bivariate to fit a bivariate nonlinear mixed model for saturated fat and energy intake. The program uses the parameter estimates from programs 1) and 2) as initial parameter estimates.

4) example4.distrib_bivariate.sfat.energy.sas – This program calls macro Distrib_Bivariate to generate a Monte Carlo distribution of usual saturated fat (g/day) and energy (kcal/day) intake, and then calls macro Percentiles_Survey to estimate percentiles of the population distribution of the percent of usual energy from saturated fat. The program uses parameter estimates from program 3) to generate the distribution.

Section 2: Documentation for SAS macro NLMixed_Univariate

```
/*************************************************************************
*                                                                        *
* SAS macro NLMixed_Univariate fits a univariate nonlinear mixed model for  *
* a food or nutrient.  The food/nutrient can be episodically consumed or    *
* consumed every day.                                                    *
*                                                                        *
* The model is described in:                                             *
*    Tooze JA, et al. A new statistical method for estimating the usual     *
*    intake of episodically consumed foods with application to their        *
*    distribution. J Am Diet Assoc. 2006;106:1575-1587.                     *
*                                                                        *
* Model for episodically consumed foods/nutrients (two-part model):          *
* ------------------------------------------------------------           *
* For episodically consumed foods/nutrients, the macro fits a two-part       *
* nonlinear mixed model, where the first part is the probability to          *
* consume and the second part is the amount consumed on a consumption day.   *
* The model allows for covariates in each part, includes a random effect     *
* for each part, and allows the random effects to be correlated.             *
*                                                                        *
* Model for foods/nutrients consumed every day (one-part model):             *
* ------------------------------------------------------------           *
* For foods/nutrients consumed every day, the macro fits a one-part          *
* nonlinear mixed model of the amount consumed (the probability to consume   *
* is assumed to be 1). The model allows for covariates and includes a        *
* random effect.                                                         *
*                                                                        *
* For a food/nutrient that is consumed nearly every day by nearly everyone,  *
* so that the number of zero values is small, it may be preferable to use    *
* the one-part (consumed every day) model, since the two-part model may      *
* have trouble modeling the probability to consume in such a situation.      *
*                                                                        *
* Note, however, that the one-part model requires all responses to be        *
* greater than zero (zero values are treated as missing values).             *
* Before fitting the one-part model to a food/nutrient that has some zero    *
* values, replace the zero values with a small positive value, such as       *
* half the smallest observed nonzero value.                              *
*                                                                        *
* The macro calls the SAS NLMixed procedure to fit the model.                *
*                                                                        *
*************************************************************************
*                                                                        *
* Macro Parameters:                                                      *
*                                                                        *
*    Required Parameters:                                                *
*       data          = name of SAS data set containing the data to be   *
*                       analyzed. The data set has multiple observations *
*                       for each subject, one for each reptition of the  *
*                       24-hour recall (or other dietary instrument).    *
*       subject       = name of the variable that uniquely identifies each *
*                       subject (i.e., ID variable).                     *
*       repeat        = name of the variable that indexes repeated       *
*                       observations for each subject.                   *
*       response      = name of the food/nutrient variable to be modeled *
*                       (24-hour recall variable for the food/nutrient). *
*       modeltype     = model for food/nutrient:                         *
*                       to fit the two-part (epsisodic) model, specify   *
*                          modeltype = TWOPART                           *
*                       to fit the one-part (every day) model, specify   *
*                          modeltype = ONEPART                           *
*                                                                        *
*    Optional Parameters:                                                *
```

```
*       covars_prob   = list of variables that are covariates in the        *
*                       probability part of the two-part model.             *
*                       if modeltype=ONEPART, then covars_prob is ignored.  *
*       covars_amt    = list of variables that are covariates in the        *
*                       one-part model or the amount part of the            *
*                       two-part model.                                     *
*       link          = link function for the probability part of the two-  *
*                       part model. to fit a logistic model, specify        *
*                          link = LOGIT                                     *
*                       to fit a probit model, specify                      *
*                          link = PROBIT                                    *
*                       by default, link = LOGIT.                           *
*                       if modeltype = ONEPART, then link is ignored.       *
*       lambda        = Box-Cox transformation parameter for the amount     *
*                       part of the model. If lambda is not specified,      *
*                       then it is estimated as part of the model.          *
*       var_u1        = variance of the random effect in the probability    *
*                       part of the two-part model.                         *
*                       If var_u1 is not specified, then it is estimated    *
*                       as part of the model.                               *
*                       if modeltype = ONEPART, then var_u1 is ignored.     *
*       var_u2        = variance of the random effect in the one-part model *
*                       or the amount part of the two-part model.           *
*                       If var_u2 is not specified, then it is estimated    *
*                       as part of the model.                               *
*       indep_u       = Y if random effects u1 and u2 are independent.      *
*                       = N if random effects u1 and u2 are dependent.      *
*                       by default, indep_u = N.                            *
*                       if modeltype = ONEPART, then indep_u is ignored.    *
*       replicate_var = name of the sampling weight variable if the data    *
*                       is from a complex survey with weights.              *
*                       by default, the macro performs an unweighted        *
*                       analysis (assumes a simple random sample).          *
*       nloptions     = options for the NLMixed procedure that are          *
*                       appended to the PROC NLMIXED statement, e.g.,       *
*                          nloptions = technique=newrap maxiter=200,        *
*       init_parms    = name of SAS data set that contains initial          *
*                       parameter estimates. See the description of output  *
*                       data set parms_u (below) for further information.   *
*                       if init_parms is not specified, then the macro      *
*                       calculates initial parameter estimates.             *
*       print         = Y to print the output from the model.               *
*                       = N to supress printing the output from the model.  *
*                       = V (verbose) to print extra output.                *
*                       The verbose option is useful for debugging.         *
*                       by default, print = Y.                              *
*       ntitle        = number of titles defined by the user.               *
*                       by default, ntitle = 2.                             *
*                                                                           *
*****************************************************************************
*                                                                           *
* Output Data Sets:                                                         *
*                                                                           *
*   parms_u = data set containing parameter estimates for the model.        *
*             parms_u contains the following variables:                     *
*                                                                           *
*                 A_Intercept = intercept in the amount part of the model.  *
*                 A_varname   = regression slope for covariate "varname"    *
*                               in the amount part of the model.            *
*                 A_LogSDe    = Log(Sqrt(Var_e))                            *
*                 LogSDu2     = Log(Sqrt(Var_u2))                           *
*                 Var_e       = variance of the within-person error in the  *
*                               amount part of the model.                   *
```

```
*                 Var_u2      = variance of the random effect in the       *
*                               amount part of the model.                  *
*                                                                          *
*             if fitting the two-part model, then parms_u also contains    *
*             the following variables:                                     *
*                                                                          *
*                 P_Intercept = intercept in the prob. part of the model.  *
*                 P_varname   = regression slope for covariate "varname"   *
*                               in the prob. part of the model.            *
*                 LogSDu1     = Log(Sqrt(Var_u2))                          *
*                 z_u1u2      = Fisher transformation of Corr_u1u2:        *
*                                  z = ln[(1+corr)/(1-corr)] / 2           *
*                 Var_u1      = variance of the random effect in the       *
*                               prob. part of the model.                   *
*                 Cov_u1u2    = covariance of random effects u1 and u2.    *
*                 Corr_u1u2   = correlation of random effects u1 and u2.   *
*                                                                          *
*             note: if specifying initial parameter estimates using the    *
*                   init_parms option, the init_parms data set should have *
*                   the same variables as parms_u, except it should not   *
*                   include var_e, var_u2, var_u1, cov_u1u2 or corr_u1u2   *
*                   (these are derived parameters, i.e., functions of the  *
*                    other parameters).                                    *
*                                                                          *
*   pred_x_u = data set containing predicted values for the model.         *
*              pred_x_u contains all the variables in the input data set,  *
*              plus the following variable:                                *
*                                                                          *
*                 pred_x_a = linear predictor for mean amount on           *
*                            consumption day.                              *
*                                                                          *
*              if fitting the two-part model, then pred_x_u also contains  *
*              the following variable:                                     *
*                                                                          *
*                  pred_x_p = linear predictor for probability of          *
*                             consumption.                                 *
*                                                                          *
****************************************************************************/
```

Section 3: Documentation for SAS macro NLMixed_Bivariate:

```
/*************************************************************************
*                                                                        *
* SAS macro NLMixed_Bivariate fits a bivariate nonlinear mixed model for two*
* foods/nutrients.  The first food/nutrient can be episodically consumed  *
* or consumed every day, while the second food/nutrient is assumed to be  *
* consumed every day.                                                    *
*                                                                        *
* The model for two foods/nutrients that are consumed every day is       *
* described in:                                                          *
*    Freedman LS, et al. The population distribution of ratios of usual  *
*    intake of dietary components that are consume every day can be      *
*    estimated from repeated 24-hour recalls. J Nutr. 2010;140:111-116.  *
*                                                                        *
* Model for episodically consumed foods/nutrients (two-part model):      *
* ---------------------------------------------------------------        *
* For episodically consumed foods/nutrients, the macro fits a two-part   *
* nonlinear mixed model, where the first part is the probability to      *
* consume and the second part is the amount consumed on a consumption day.*
* The model allows for covariates in each part, includes a random effect *
* for each part, and allows the random effects to be correlated.         *
*                                                                        *
* Model for foods/nutrients consumed every day (one-part model):         *
* ------------------------------------------------------------           *
* For foods/nutrients consumed every day, the macro fits a one-part      *
* nonlinear mixed model of the amount consumed (the probability to consume*
* is assumed to be 1). The model allows for covariates and includes a    *
* random effect.                                                         *
*                                                                        *
* For a food/nutrient that is consumed nearly every day by nearly everyone,*
* so that the number of zero values is small, it may be preferable to use*
* the one-part (consumed every day) model, since the two-part model may  *
* have trouble modeling the probability to consume in such a situation.  *
*                                                                        *
* Note, however, that the one-part model requires all responses to be    *
* greater than zero (zero values are treated as missing values).         *
* Before fitting the one-part model to a food/nutrient that has some zero*
* values, replace the zero values with a small positive value, such as   *
* half the smallest observed nonzero value.                              *
*                                                                        *
* Note: Initial parameter estimates must be supplied by the user.        *
* They can be estimated using SAS macro NLMixed_Univariate.              *
*                                                                        *
* The macro calls the NLMixed procedure to fit the model.                *
*                                                                        *
*************************************************************************
*                                                                        *
* Macro Parameters:                                                      *
*                                                                        *
*    Required Parameters:                                                *
*       data          = name of SAS data set containing the data to be   *
*                       analyzed. The data set has multiple observations *
*                       for each subject, one for each reptition of the  *
*                       24-hour recall (or other dietary instrument).    *
*       subject       = name of the variable that uniquely identifies each*
*                       subject (i.e., ID variable).                     *
*       repeat        = name of the variable that indexes repeated       *
*                       observations for each subject.                   *
*       response1     = name of first food/nutrient variable to be modeled*
*                       (24-hour recall variable for first food/nutrient).*
*       response2     = name of second food/nutrient variable to be modeled*
*                       (24-hour recall variable for second food/nutrient).*
```

```
*       modeltype    = model for first food/nutrient:                          *
*                      to fit the two-part (epsisodic) model, specify         *
*                         modeltype = TWOPART                                 *
*                      to fit the one-part (every day) model, specify         *
*                         modeltype = ONEPART                                 *
*       init_parms   = name of SAS data set that contains initial             *
*                      parameter estimates. See the description of output     *
*                      data set parms_b (below) for further information.      *
*                                                                             *
*    Optional Parameters:                                                     *
*       covars_prob1  = list of variables that are covariates in the          *
*                      probability part of the two-part model for the         *
*                      first food/nutrient.                                   *
*                      if modeltype = ONEPART, then covars_prob is ignored.*
*       covars_amt1   = list of variables that are covariates in the          *
*                      one-part model or the amount part of the               *
*                      two-part model for the first food/nutrient.            *
*       covars_amt2   = list of variables that are covariates in the          *
*                      one-part model for the second food/nutrient            *
*       link          = link function for the probability part of the two-    *
*                      part model for the first food/nutrient.                *
*                      to fit a logistic model, specify                       *
*                         link = LOGIT                                        *
*                      to fit a probit model, specify                         *
*                         link = PROBIT                                       *
*                      by default, link = LOGIT.                              *
*                      if modeltype = ONEPART, then link is ignored.          *
*       lambda1       = Box-Cox transformation parameter for the first        *
*                      food/nutrient. If lambda1 is not specified, then       *
*                      it is estimated as part of the model.                  *
*       lambda2       = Box-Cox transformation parameter for the second       *
*                      food/nutrient. If lambda2 is not specified, then       *
*                      it is estimated as part of the model.                  *
*       var_u1        = variance of the random effect in the probability      *
*                      part of the model for the first food/nutrient.         *
*                      If var_u1 is not specified, then it is estimated       *
*                      as part of the model.                                  *
*                      if modeltype = ONEPART, then var_u1 is ignored.        *
*       var_u2        = variance of the random effect in the amount           *
*                      part of the model for the first food/nutrient.         *
*                      If var_u2 is not specified, then it is estimated       *
*                      as part of the model.                                  *
*       var_u3        = variance of the random effect in the amount           *
*                      part of the model for the second food/nutrient.        *
*                      If var_u3 is not specified, then it is estimated       *
*                      as part of the model.                                  *
*       indep_u1      = Y if random effect u1 is independent of u2 and u3.    *
*                     = N otherwise. by default, indep_u1 = N.                *
*                      if modeltype = ONEPART, then indep_u1 is ignored.      *
*       indep_u2      = Y if random effect u2 is independent of u1 and u3.    *
*                     = N otherwise. by default, indep_u2 = N.                *
*       indep_u3      = Y if random effect u3 is independent of u1 and u2.    *
*                     = N otherwise. by default, indep_u3 = N.                *
*       replicate_var = name of the sampling weight variable if the data      *
*                      is from a complex survey with weights.                 *
*                      by default, the macro performs an unweighted           *
*                      analysis (assumes a simple random sample).             *
*       nloptions     = options for the NLMixed procedure that are            *
*                      appended to the PROC NLMIXED statement, e.g.,          *
*                       nloptions = technique=newrap maxiter=200,             *
*       print         = Y to print the output from the model.                 *
*                     = N to supress printing the output from the model.      *
*                     = V (verbose) to print extra output.                    *
```

```
*                       The verbose option is useful for debugging.         *
*                       by default, print = Y.                              *
*       ntitle        = number of titles defined by the user.               *
*                       by default, ntitle = 2.                             *
*                                                                           *
*****************************************************************************
*                                                                           *
* Output Data Sets:                                                         *
*                                                                           *
*   parms_b = data set containing parameter estimates for the model.        *
*             parms_b contains the following variables:                     *
*                                                                           *
*                 A1_Intercept = intercept in the amount part of the model  *
*                                for the first food/nutrient.               *
*                 A1_varname   = regression slope for covariate "varname"   *
*                                in the amount part of the model for the    *
*                                first food/nutrient.                       *
*                 A2_Intercept = intercept for the second food/nutrient.    *
*                 A2_varname   = regression slope for covariate "varname"   *
*                                for the second food/nutrient.              *
*                 A1_Lambda    = Box-Cox transformation parameter for the   *
*                                first food/nutrient.                       *
*                 A2_Lambda    = Box-Cox transformation parameter for the   *
*                                second food/nutrient.                      *
*                 A1_LogSDe    = Log(Sqrt(Var_e2))                          *
*                 A2_LogSDe    = Log(Sqrt(Var_e3))                          *
*                 z_e2e3       = Fisher transformation of Corr_e2e3:        *
*                                  z = ln[(1+corr)/(1-corr)] / 2            *
*                 Var_e2       = variance of within-person error e2 (amount *
*                                part of model for first food/nutrient).    *
*                 Var_e3       = variance of within-person error e3 (second *
*                                food/nutrient).                            *
*                 Var_u2       = variance of random effect u2 (amount part  *
*                                of model for first food/nutrient).         *
*                 Var_u3       = variance of random effect u3 (second       *
*                                food/nutrient).                            *
*                                of the model for the second food).         *
*                 Cov_e2e3     = covariance of random errors e2 and e3.     *
*                 Corr_e2e3    = correlation of random errors e2 and e3.    *
*                 Cov_u2u3     = covariance of random effects u2 and u3.    *
*                 Corr_u2u3    = correlation of random effects u2 and u3.   *
*                                                                           *
*             if fitting the two-part model for the first food/nutrient,    *
*             then parms_b also contains the following variables:           *
*                                                                           *
*                 P1_Intercept = intercept in the prob. part of the model   *
*                                for the first food/nutrient.               *
*                 P1_varname   = regression slope for covariate "varname"   *
*                                in the prob. part of the model for the     *
*                                first food/nutrient.                       *
*                 Var_u1       = variance of random effect u1 (prob. part   *
*                                of model for first food/nutrient).         *
*                 Cov_u1u2     = covariance of random effects u1 and u2.    *
*                 Cov_u1u3     = covariance of random effects u1 and u3.    *
*                 Corr_u1u2    = correlation of random effects u1 and u2.   *
*                 Corr_u1u3    = correlation of random effects u1 and u3.   *
*                                                                           *
*             note: initial parameter estimates must be supplied by the     *
*                   user using the init_parms option.                       *
*                   the user-supplied data set will have the same variables *
*                   as data set parms_b, except it should not include the   *
*                   following variables (intital estimates for these        *
*                   parameters will be set to zero):                        *
```

```
*                        z_e2e3                                              *
*                        Cov_e2e3                                            *
*                        Corr_e2e3                                           *
*                        Cov_u1u2 Cov_u1u3 Cov_u2u3                          *
*                        Corr_u1u2 Corr_u1u3 Corr_u2u3                       *
*                    All the necessary initial parameter estimates can be    *
*                    estimated using the SAS macro NLMixed_Univariate.       *
*                                                                            *
*   pred_x_b = data set containing predicted values for the model.           *
*              pred_x_b contains all the variables in the input data set,    *
*              plus the following variables:                                 *
*                                                                            *
*                 pred_x_a1 = predicted mean amount on consumption day for   *
*                             the first food/nutrient.                       *
*                 pred_x_a2 = predicted mean amount for the second           *
*                             food/nutrient.                                 *
*                                                                            *
*             if fitting the two-part model for the first food/nutrient,     *
*             then pred_x_b also contains the following variable:            *
*                                                                            *
*                 pred_x_p1 = predicted probability of consumption for       *
*                             the first food/nutrient.                       *
*                                                                            *
******************************************************************************/
```

Section 4: Documentation for SAS macro Distrib_Bivariate:

```
/*************************************************************************
*                                                                        *
* SAS Macro Distrib_Bivariate estimates the bivariate distribution of true  *
* usual intake of two foods/nutrients using Monte Carlo simulation.      *
* The first food/nutrient can be episodically consumed or consumed every    *
* day, while the second food/nutrient is assumed to be consumed every day.  *
*                                                                        *
* Macro Distrib_Bivariate reads data sets parms_b and pred_x_b output by    *
* SAS macro NLMixed_Bivariate, and uses Monte Carlo simulation of the       *
* random effects to generate the distribution of true usual intake.      *
*                                                                        *
* The method for two foods/nutrients that are consumed every day is         *
* described in:                                                          *
*    Freedman LS, et al. The population distribution of ratios of usual     *
*    intake of dietary components that are consume every day can be         *
*    estimated from repeated 24-hour recalls. J Nutr. 2010;140:111-116.     *
*                                                                        *
*************************************************************************
*                                                                        *
* Macro Parameters:                                                      *
*                                                                        *
*    Required Parameters:                                                *
*       param      = name of SAS data set containing the estimated model    *
*                    parameters from macro NLMixed_Bivariate.              *
*       predicted  = name of SAS data set containing the predicted values   *
*                    calculated by macro NLMixed_Bivariate.                *
*                    typically, each subject has only one predicted value,  *
*                    so has only one observation in the data set.          *
*                    some models, however, allow a subject to have         *
*                    different predicted values on different days (e.g.,    *
*                    subjects may eat differently on weekends than on      *
*                    weekdays). for such models, each subject should have   *
*                    one observation for each unique predicted value.      *
*                    macro Distrib_Univariate will then calculate true     *
*                    usual intake as a weighted average of usual intake on  *
*                    the different days (see day_wht, below).              *
*       subject    = name of the variable that uniquely identifies each     *
*                    subject in the predicted data set (i.e., ID variable). *
*       modeltype  = model for first food/nutrient:                         *
*                       to fit the two-part (epsisodic) model, specify      *
*                          modeltype = TWOPART                              *
*                       to fit the one-part (every day) model, specify      *
*                          modeltype = ONEPART                              *
*                                                                        *
*    Optional Parameters:                                                *
*       link       = link function for the probability part of the two-     *
*                    part model. to fit a logistic model, specify          *
*                       link = logit                                       *
*                    to fit a probit model, specify                        *
*                       link = probit                                      *
*                    by default, link = logit.                             *
*                    if modeltype = ONEPART, then link is ignored.         *
*       nsim_mc    = number of pseudo-individuals to simulate for each      *
*                    real individual in the data set. if the data set has   *
*                    n subjects, then the monte carlo distribution will    *
*                    have nsim_mc * n pesudo-individuals.                  *
*                    by default, nsim_mc = 1.                              *
*       day_wgt    = name of the "day weight" variable in the predicted     *
*                    data set. if the model allows subjects to have        *
*                    different predicted values on different days (see      *
*                    predicted, above), then macro distrib will calculate   *
```

```
 *                    each subject's true usual intake as a weighted average *
 *                    of usual intake on the different days (using the       *
 *                    weights in the "day weight" variable) . for example,   *
 *                    if each subject has one predicted value for weekends   *
 *                    (Friday-Sunday) and one predicted value for weekdays   *
 *                    (Monday-Thursday), then the day weight variable should *
 *                    equal 3/7 for the weekend predicted value and 4/7 for  *
 *                    the weekday predicted value.                           *
 *                    by default, if subjects have multiple predicted        *
 *                    values, then true usual intake is calculated as an     *
 *                    unweighted average of usual intake on different days.  *
 *       min_a1     = minimum true usual intake for a one-part model, or     *
 *                    mininum true usual amount on consumption day for a     *
 *                    two-part model, for the first food/nutrient.            *
 *                    the monte carlo method generates amount consumed on    *
 *                    a transformed scale, then back-transforms it to the    *
 *                    original scale. occasionally, the generated amount is  *
 *                    too small to be back-transformed. when this happens,   *
 *                    amount on the original scale is set to min_a1.         *
 *                    by default, min_a1 = 0.                                *
 *       min_a2     = minimum true usual intake for the second food/nutrient.*
 *                    by default, min_a2 = 0.                                *
 *       backtran   = 1 to use a numerical integration method to integrate   *
 *                    back-transformed reported intake over the distribution *
 *                    of within-person error.                                *
 *                  = 2 to use a Taylor linearization approximation to the   *
 *                    integral (not recommended).                            *
 *                  = 3 to back-transform without integrating over the       *
 *                    distribution of within-person error. (not recommended).*
 *                    by default, backtran = 1.                              *
*       print      = Y to print summary of the monte carlo distribution.    *
 *                  = N to supress printing summary of the distribution.     *
 *                    by default, print = Y.                                 *
 *       ntitle     = number of titles defined by the user.                  *
 *                    by default, ntitle = 2.                                *
 *                                                                           *
 *****************************************************************************
 *                                                                           *
 * Output Data Set:                                                          *
 *                                                                           *
 *   _mcsim = data set containing the monte carlo distribution of true usual *
 *            intake of the food/nutrient.  if the the predicted data set    *
 *            has n subjects, _mcsim will have nsim_mc * n pseudo-subjects.  *
 *            _mcsim contains all the numeric variables in the predicted     *
 *            data set, plus the following variables:                        *
 *                                                                           *
 *               t1    = true usual intake of first food/nutrient.          *
 *               t2    = true usual intake of second food/nutrient.         *
 *               a1    = true usual amount on consumption day for first     *
 *                       food/nutrient.                                      *
 *               bc_a1 = Box-Cox transformed true usual amount a1.          *
 *               bc_t2 = Box-Cox transformed true usual intake t2.          *
 *                                                                           *
 *            if fitting the two-part model, then _mcsim also contains       *
 *            the following variables:                                       *
 *                                                                           *
 *               p1             = true probability to consume.              *
 *               linear_pred_p1 = transformed true probability to consume:  *
 *                                  = log(p1/(1-p1))  for logistic model     *
 *                                  = probit(p1)      for probit model       *
 *                                                                           *
 *****************************************************************************/
```

Section 5: Documentation for SAS macro Percentiles_Survey:

```
/************************************************************************
 *                                                                      *
 * SAS Macro Percentiles_Survey estimates means, standard deviations and *
 * percentiles for survey data.                                         *
 *                                                                      *
 * if sampling weights are specified, estimates are based on a weighted *
 * empirical distribution. Otherwise, they are based on an unweighted   *
 * empirical distribution.                                              *
 *                                                                      *
 ************************************************************************
 *                                                                      *
 * Macro Parameters:                                                    *
 *                                                                      *
 *    Required Parameters:                                              *
 *       data          = name of SAS data set containing the data to be *
 *                       analyzed.                                      *
 *       var           = name of variable for which means and percentiles *
 *                       are to be estimated.                           *
 *                                                                      *
 *    Optional Parameters:                                              *
 *       byvar         = list of variables which define by-groups in the *
 *                       data. means and percentles will be estimated   *
 *                       separately for each by-group.                  *
 *       weight        = name of the sampling weight variable, if the data *
 *                       is from a complex survey with weights.         *
 *                       by default, the macro assumes equal weights.   *
 *       cutpoints     = list of values for which cut-point probabilities *
 *                       are to be calculated (Prob(X <= cut-point)).   *
 *                       by default no cut-point probabilites are calculted. *
 *       print         = Y to print means and percentiles.               *
 *                       N to supress printing means and percentiles.   *
 *                       by default, print = Y.                         *
 *       ntitle        = number of titles defined by the user.          *
 *                       by default, ntitle = 2.                        *
 *                                                                      *
 ************************************************************************
 *                                                                      *
 * Output Data Set:                                                     *
 *                                                                      *
 *   _percentiles = data set containing estimated means and percentiles. *
 *                  _percentiles contains the following variables:      *
 *                                                                      *
 *                     mean     = estimated mean.                       *
 *                     variance = estimated variance.                   *
 *                     stddev   = estimated standard distribution.      *
 *                     min      = minimum value.                        *
 *                     max      = maximum value.                        *
 *                     Pctile1  = estimated first percentile.           *
 *                     Pctile2  = estimated second percentiile.         *
 *                     ...                                              *
 *                     Pctile99 = estimated 99th percentile.            *
 *                                                                      *
 *                  if cut-points are specified, then _percentiles also *
 *                  contains the following variables:                   *
 *                                                                      *
 *                     Prob1    = probability for first cut-point.      *
 *                     Prob2    = probability for second cut-point.     *
 *                     ...                                              *
 *                                                                      *
 *                  if by-group variables are specified, they are also  *
 *                  included in data set _percentiles.                  *
```

```
 *                                                                            *
 ******************************************************************************/
```

Section 6: Documentation for SAS macro BoxCox_Survey:

```
/*************************************************************************
*                                                                        *
* SAS Macro BoxCox_Survey finds the best Box-Cox transformation of survey   *
* data by minimizing the SSE of a normal probability plot.               *
*                                                                        *
* if sampling weights are specified, the normal probability plot is based   *
* on a weighted empirical distribution. Otherwise, it is based on an     *
* unweighted empirical distribution.                                     *
*                                                                        *
*************************************************************************
*                                                                        *
* Macro Parameters:                                                      *
*                                                                        *
*    Required Parameters:                                                *
*       data          = name of SAS data set containing the data to be   *
*                       analyzed.                                        *
*       subject       = name of the variable that uniquely identifies each  *
*                       subject (i.e., ID variable).                     *
*       var           = name of variable for which means and percentiles  *
*                       are to be estimated.                             *
*                                                                        *
*    Optional Parameters:                                                *
*       byvar         = list of variables which define by-groups in the  *
*                       data. means and percentles will be estimated     *
*                       separately for each by-group.                    *
*       weight        = name of the sampling weight variable, if the data   *
*                       is from a complex survey with weights.           *
*                       by default, the macro assumes equal weights.     *
*       print         = Y to print the best Box-Cox transformation.       *
*                     = N to supress printing the the best Box-Cox       *
*                       transformation.                                  *
*                     = V (verbose) to print distributions before and after *
*                       Box-Cox transformation (using proc univariate).  *
*                       by default, print = Y.                           *
*       plot          = Y to print normal probability plots before and    *
*                       after Box-Cox transformation (using proc gplot).    *
*                     = N to supress printing normal probability plots.  *
*                       by default, plot = N.                            *
*       ntitle        = number of titles defined by the user.            *
*                       by default, ntitle = 2.                          *
*                                                                        *
*************************************************************************
*                                                                        *
* Output Data Sets:                                                      *
*                                                                        *
*   _lambda = data set containing the best Box-Cox transformation.       *
*             _lambda contains the following variables:                  *
*                                                                        *
*                lambda_&var  = Box-Cox transformation parameter for     *
*                               variable &var.                           *
*                sse          = sum of squared errors for the normal     *
*                               probability plot.                        *
*                                                                        *
*             if by-group variables are specified, they are also included   *
*             in data set _lambda.                                       *
*                                                                        *
*   &data = input data set containing the data to be analyzed.           *
*           The following variables are added to &data:                  *
*                                                                        *
*                lambda_&var  = Box-Cox transformation parameter for     *
*                               variable &var.                           *
```

```
 *                boxcox_&var  = Box-Cox transformation of variable &var.    *
 *                                                                           *
 *****************************************************************************/
```