A User's Guide to C-SIDE (Software for Intake Distribution Estimation). Version 1.0.

Dietary Assessment Research Series Report 8

Sarah M. Nusser, Alicia L. Carriquiry, Kevin W. Dodd,
Wayne A. Fuller, and Helen H. Jensen

Technical Report 96-TR 31
September 1996

A User's Guide to

C-SIDE

Software for Intake Distribution Estimation

Version 1.0

Technical Report 96-TR 31

Dietary Assessment Research Series Report 8

September 1996

Department of Statistics and
Center for Agricultural and Rural Development
(CARD)
Iowa State University

Table of Contents

Foreword

C-SIDE implements the method for estimating usual intake distributions developed at Iowa State University. Distributions can be estimated for dietary components consumed daily, such as nutrients, or for components consumed less frequently, such as foods.

The program is distributed with three example analyses. Step-by-step instructions on how to execute the examples are included at the end of the manual. These examples illustrate the use of the software and its many options.

We gratefully acknowledge input from early users of SIDE, including: George Beaton of the University of Toronto; Adeline Wilcox and Phillip Kott from the Agricultural Research Service (ARS) of the U.S. Department of Agriculture; and Gerry Parise, Mary Jane Novenario and Chun-Fu Chen from the Center for Agricultural and Rural Development (CARD) at Iowa State University. Paul Sisler from CARD provided invaluable support with technical editing and formatting of previous editions of the manual. Karleen Gillen contributed her usual superb typing expertise.

Sarah Nusser, Alicia Carriquiry, Kevin Dodd, and Wayne Fuller developed the methodology. Todd Krueger wrote the C language code, including development of the X Windows interface. Helen Jensen participated in developing the method and the manual.

Iowa State University
September 1996

Preface

Welcome to C-SIDE (Software for Intake Distribution Estimation)! C-SIDE can help you obtain estimates of usual nutrient and food intake distributions, their moments, and their percentiles. The software uses methods developed at Iowa State University and described in Nusser, Carriquiry, Dodd, and Fuller (1996) and Nusser, Fuller, and Guenther (1996).

If you have nutrient intake data (consumed almost daily), with C-SIDE you can
- Read ASCII text files containing intake data
- Adjust dietary intake data for discrete or continuous factors of your choice, such as day of the week or age of the individual, selecting any sample day or the grand mean as the standard
- Incorporate sampling weights associated with each individual
- Analyze data collected on consecutive days using an estimate of the correlation among intake days for each different dietary component
- Estimate the distribution of usual intake for any number of nutrients and for subpopulations, such as subpopulations defined by gender
- Compute the proportion of all individuals with intake above or below any number of given thresholds for each nutrient, including unequally spaced thresholds
- Estimate the standard error of each estimated percentile of the distributions using linearization variance estimators
- Estimate the standard error for moments and percentiles of the distribution for data from complex survey designs using a jackknife or balanced repeated replication (BRR) method
- Output data necessary for producing plots of estimated densities with another software package.

If you have food intake data (not consumed daily), with C-SIDE you can
- Read ASCII text files containing intake data
- Adjust positive dietary intake data for discrete or continuous factors of your choice, such as day of the week or age of the individual, selecting any sample day or the grand mean as the standard
- Incorporate sampling weights associated with each individual

- Estimate the usual intake distribution for consumption days only for any number of foods and for subpopulations, such as subpopulations defined by gender
- Estimate the distribution of the probability that an individual consumes the food on a day
- Estimate the distribution of usual intake on all days
- Estimate the fraction of nonconsumers and consumers of the food in the population
- Compute the proportion of all individuals with intake above or below any number of given thresholds for each food, including unequally spaced thresholds, using either usual intake distribution
- Output data necessary for producing plots of estimated densities with another software package.

This manual is divided into four main parts:
1. An introduction, which lists equipment and data requirements
2. A description of the use of the software
3. A description of the output that you can produce with C-SIDE
4. An examples section, with analyses of data sets that require the use of default and non-default program options

This manual was designed for the user who is familiar with standard statistical and computer techniques and terminology. While this package was extensively tested, we encourage you to **always check your results.**

Should you find errors, or should you have any comments or suggestions, please contact us.

Todd Krueger
Survey Section
Statistics Laboratory
219 Snedecor Hall
Iowa State University
Ames, IA 50011-1210

Phone: (515) 294-5242
Fax: (515) 294-2456
Email: tkrueger@iastate.edu

I: BEFORE YOU START

What equipment and software will you need?

C-SIDE (Version 1.0) is designed to work on a computer running UNIX and an X Windows System graphical window manager.

To run C-SIDE you will need either a precompiled executable file or the software necessary to compile C-SIDE in your computer's environment.

Executable files are available for
- DEC Alpha workstations operating under OSF/1 Version 3.2
- DEC MIPS workstations operating under Ultrix 4.3
- SUN Sparc workstations operating under SunOS 5.4 (Solaris 2.4)

To recompile C-SIDE for an environment different from those listed above, you will need*

- ANSI compliant C compiler
- X Window System, Version 11 Release 5 (X11R5), including files needed to compile a program, such as libraries and header files for Xlib, the X Toolkit Intrinsics, and the Athena widget set

* Instructions for compiling non-executable C-SIDE packages have not been included in this manual. If you need assistance compiling C-SIDE, contact us at the address listed in the *Preface* to this manual.

What about your data?

The following must be true of your data set:

Multiple days of intake data must be available on at least some of the individuals in the sample. For foods, there must be multiple positive intakes. Multiple intakes are required because the method estimates the within-individual variation of intakes. This estimation is not possible when only one observation is available on each individual in the sample.

For food intake data, the number of observations for an individual must be the same for all individuals in the data set. For nutrient data, the number of observations may vary among individuals.

All recorded intakes must be nonnegative. If you have transformed your intake data in any way, verify that the transformed intakes input into the program are also nonnegative.

Data must be in ASCII text files, with one observation per line. A unique identification variable for each individual must be present. The file must be ordered so multiple observations on one individual are grouped together. Variables may appear in any order in the data file.

The observed distribution of nutrient intakes must be smooth. No one intake value should appear with frequency significantly higher than the rest of the values. Food intake data may contain a large number of observed zero intakes, but the observed distribution of positive intakes must be smooth.

For food intakes, the usual intake for days on which the food is consumed is assumed to be independent of the frequency with which the food is consumed.

C-SIDE is also capable of accommodating data sets with the following characteristics:

> If nutrient intake data were collected on no more than three consecutive days, and if you want to account for the day-to-day correlations, then you must supply an estimate of the correlation. Please refer to Carriquiry et al. (1995) for a discussion of the procedure for estimating day-to-day correlations. Correlation estimates for several dietary components that were estimated from the 1989-91 Continuing Survey of Food Intake by Individuals (CSFII) are listed in Appendix A. This option is not available for food intakes. (See Chapter II, p. 21, "Correlation".)

> Sampling weights for each individual in the sample may be used. C-SIDE will produce a weighted analysis of your data.

> If your intake data were collected on different days of the week, during different seasons, or from individuals of different ages, ethnic groups, or income level, you can partially remove the effect of those variables from your intake data using C-SIDE's adjustment procedure. You may use continuous or qualitative variables for data adjustment.

> Data may be organized into subpopulations, such as those defined by ethnic groups, geographic location, or income level. C-SIDE can provide a separate analysis for each subpopulation.

> Replicate weight sets may be used to estimate variances of estimated usual intake quantiles and CDF values with a jackknife or balanced repeated replication (BRR) method.

What kind of output can you get?

C-SIDE will produce
1. Diagnostics for the different steps in the method for estimating usual intake distributions
2. Moments of the estimated usual intake distributions
3. Estimated percentile values and the proportion of individuals in the sample with observed intake above and below given thresholds

C-SIDE can also produce ASCII text data files containing
1. Adjusted daily intake data in original units
2. Normal scale daily intakes, power transformed daily intakes, and normal scores from adjusted daily intakes
3. A set of representative usual intakes generated from the estimated usual intake distribution

    4. Percentiles and estimated density function values for the distribution of usual intakes

    5. Estimated usual intakes for each individual in the data set

Installing and starting C-SIDE

The following files are included with C-SIDE:

cside	Executable program file*
csfii85.txt	Example data file
om89_91.txt	Example data file
food85.txt	Example data file
example1.cfg	Example configuration file
example2.cfg	Example configuration file
example3.cfg	Example configuration file

    Copy these files into a directory on your computer system. Type "cside &" from your UNIX command prompt to start the program.

* If you have received C-SIDE as non-executable source code files and need help compiling your software, contact us at the address listed in the *Preface* to this manual.

II: USING C-SIDE

X Windows basics

You interact with C-SIDE through the pointer (usually a mouse with three buttons) and the keyboard. The pointer is used to choose items off lists and menus, and to position the cursor inside boxes that can accept information typed from the keyboard. Presuming that your pointer is a standard three button mouse, in the following narrative pointer button 1 is the left mouse button, pointer button 2 is the middle mouse button, and pointer button 3 is the right mouse button.

Many of the items found on C-SIDE windows perform similar functions. The following section describes user interactions with items of particular types.

Sample X Window

Command buttons

Command buttons appear as ovals containing a text label. A command button is activated when you move the pointer so that the cursor is on top of the command button and press pointer button 1 (click on the command button).

Activating a command button causes one of the following actions:
- A menu appears
- A window appears which allows you to enter more information
- The button will switch between 'on' or 'off'
- The window disappears, possibly after performing an operation
- An operation is performed that may change the contents of the window

Most C-SIDE windows have three command buttons at the bottom of the window: Save, Reset, and Dismiss. Save applies any of the changes that you made to this window and closes the window. Reset restores all items on this window to the values they had when the window was last opened. Dismiss causes the window to disappear, discarding all changes you made to the window since it was last opened.

Menus

To select an item from a menu, press and hold down pointer button 1 on the command button that invokes the menu. The menu will remain on the screen as long as the pointer button is held down. Move the cursor to the item you wish to select and release the pointer button.

Scroll bars

Horizontal or vertical scroll bars appear on the left or right, top or bottom of a framed box. A scroll bar is a thin rectangular area containing a sliding bar that allows you to move the window so that the information you wish to see appears in the frame. The size of the slide bar compared to the overall rectangular area reflects how much of the underlying window is currently shown in the frame.

The slide bar is moved with pointer clicks to display different portions of the underlying window. Click on pointer button 1 anywhere in the rectangular area to cause the slide bar to move to the right (for horizontal scroll bars), or downward (for vertical scroll bars). Click on pointer button 3 anywhere in the rectangular area to cause the slide bar to move to the left (for horizontal scroll bars), or upward (for vertical scroll bars). Click on pointer button 2 to move the left (or upper) edge of the slide bar to that position in the rectangular area. Press and hold pointer button 2 to dynamically position the slide bar.

Text entry

A text area is a rectangle with a caret located at the insertion point. The insertion point is the location where new characters typed from the keyboard will appear on the screen. To enter a value in the item, position the cursor inside the box and type. Whatever characters you type appear at the position of the caret. To move the insertion point, click pointer button 1 at the desired location in the text.

The Main window

This is the main window of C-SIDE. It consists of three sections. The top section contains command buttons, the middle section displays program output, and the lower section displays diagnostic messages.

menu buttons

output window

diagnostic window

```
cside
(Quit) (Configure...) (Analyze...) (Output...) (Execute)

C-SIDE
Software for Intake Distribution Estimation
Version 1.00, September 1996

Department of Statistics and
Center for Agricultural and Rural Development
Iowa State University

Copyright (c) 1996 Iowa State University
Statistical Laboratory.  All Rights Reserved.
```
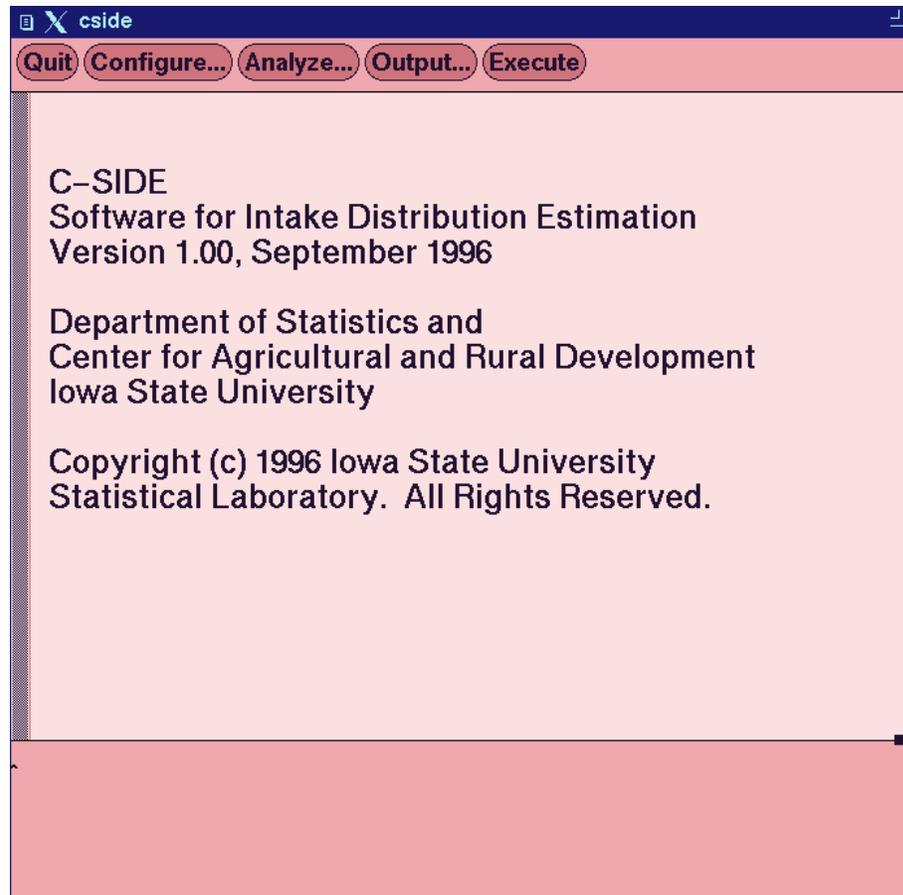
Menu buttons

There are five command buttons on the main window. They are, from left to right, Quit, Configure, Analyze, Output and Execute.

Quit

The Quit button quits C-SIDE and closes the main window.

Configure

The Configure button displays the configuration menu. This menu contains items that allow you to save and restore information that you entered in other windows.

Analyze

The Analyze menu contains items that enable you to enter all the information needed for C-SIDE to perform an analysis. This includes data file format and contents, methodology parameters, and the variables to be used in each run of the analysis.

The **Output** menu contains items that allow you to store data produced by the program. This includes data written to the output window of the main menu or data sets created by the program.

`Output`

The **Execute** button begins an analysis. This item should not be chosen until the proper information has been entered in the **Analyze** windows.

`Execute`

Output window

The output window displays results of an analysis. Upon program startup, an informational message listing program name, version number, and date is displayed in the window. This information is replaced by program output when the first analysis is performed. The output window has a scroll bar along the left side that allows you to view different portions of the program output in the window.

Diagnostic window

The diagnostic window displays information and error messages related to program operation. This window automatically scrolls, showing you only the last few lines of diagnostic messages. A scroll bar will appear when necessary to allow you to view messages that have scrolled off the window. The diagnostic window can be made larger and the output window smaller by moving the grip (small rectangular box) located on the line separating these two windows.

Menu items

The **Analyze**, **Output**, and **Configure** menus on the Main window contain the following items:

Analyze	Output	Configure
Define file format	Output data files	Open
Define variables	Save output window	Reset
Define replication weight sets		
Assign variables	Clear output window	Save
Parameters		

For a new data file that has never been analyzed before, you are likely to choose menu items in the following order:

Define file format...
Define variables...
Assign variables...
Parameters... (if necessary)
Output data files... (if necessary)
Save...
Execute
Save output window...

If you wish to run an analysis on a data file that has already been defined, then the following might be a typical session:

Open...
Assign variables...
Parameters... (if necessary)
Output data files... (if necessary)
Save...
Execute
Save output window...

The following pages describe each menu item, in the order it appears on the menu and in the order it is expected to be used. All items on the Analyze menu are described first, followed by the Output menu, and then the Configuration menu. The File selection window, which is accessed from several of the other windows, is described last.

The Analyze menu

The Analyze menu: Define file format

▤ X cside – Define file format	⏌

Data file:

Input fields are (choose one):

(Fixed–length)   (Variable–length (delimited))

This window is used to describe an input file's basic characteristics.

Delimiting character(s):

Compress consecutive delimiters: (Yes) (No)

Trim leading spaces from strings: (Yes) (No)

Trim trailing spaces from strings: (Yes) (No)

Missing values are indicated by: .

Initial number of lines to skip: 0

Maximum number of records to read: unlimited

(Save)   (Reset)   (Dismiss)

Data file:
Click on the large command button under the label to open a **File selection** window. Use the window that appears to enter the name of the file containing data to be analyzed by C-SIDE. (The **File selection** window is described on page 30.)

Input fields are (choose one):
Choose one of the two items by clicking on the item. Your choice will affect some of the other data entry items on this window, either enabling or disabling them depending on your choice. Fixed-length fields have data elements in pre-defined columns in the file. In the following example file, name is always in columns 1-5, weight is in columns 6-8, height is in 10-11, and score is in column 13.

```
Bill 171 68 3
Bob  194 73 5
Linda142 67 4
Mary  98 62 8
```

Variable-length fields are separated by a delimiting character. In the following file, each field is separated by a space.

```
Bill 171 68 3
Bob 194 73 5
Linda 142 67 4
Mary 98 62 8
```

Delimiting character(s):
(Only applicable if Variable-length (delimited) is chosen.) Enter one or more characters that separate data fields in the input file. A space, tab, comma, or the | symbol are common delimiting characters. Care must be taken to not use a character as a delimiter that otherwise appears in the data fields. By default, a single space character appears in this field. If your delimiter is not a space, be sure to backspace over it before entering your delimiting character.

Compress consecutive delimiters:
(Only applicable if Variable-length (delimited) is chosen.) If 'Yes' is chosen, then more than one consecutive occurrence of a delimiter in a data file will be treated as one delimiter. If 'No' is selected, then two or more consecutive delimiters implies that there is a missing value in the data file. If the input data line was (note the two spaces between green and 189.8):

```
green  189.8 high
```

and compress consecutive delimiters was set to 'Yes', then 'green' would be the value of the first field, '189.8' would be the second, and 'high' would be the third. If compress consecutive delimiters was set to 'No', then 'green' would be the first field, a missing value would be the second, '189.9' would be the third, and 'high' would be the fourth.

Trim leading spaces from strings:
(Only applicable if Fixed-length is chosen.) For string (non-numerical) fields, setting this option to 'Yes' causes spaces occurring at the beginning of a field to be discarded. In the following data file the month field is defined as character positions 1-9, and color as character positions 10-15.

```
October   orange
November  brown
December    red
January   white
```

If leading spaces are trimmed, then the value of color for each of the observations would be 'orange', 'brown', 'red', and 'white'. If leading spaces were not trimmed the values would be 'orange', ' brown', ' red', and ' white'.

Trim trailing spaces from strings:
(Only applicable if Fixed-length is chosen.) For character (non-numerical) fields, setting this option to 'Yes' causes spaces occurring at the end of a field to be discarded. In the preceding example, if trailing spaces were trimmed then the input values for month would be 'October', 'November', 'December', and 'January'. If trailing spaces were not trimmed, the values would be 'October ', 'November ', 'December ', and 'January '.

Missing values are indicated by:
If this character or string of characters occurs in a data field, then C-SIDE will treat the field as a 'missing' or null value. The default missing value indicator is a period. (If compress consecutive delimiters is set to 'No', then missing values can also be indicated in delimited data files by the occurrence of two consecutive delimiters.)

Class, continuous, and weight variables may not contain missing values.

Initial number of lines to skip:
If a data file has one or more lines of informational material appearing at the beginning of the file, then this option can be used to ignore these lines. Examples include data files received as part of an e-mail message with the delivery information contained at the top of the file, and data base files from some packages that contain variable format and definition information in the first few lines.

Maximum number of records to read:
This item can be used to limit the number of observations read from the input data file. You may wish to use this item to test your C-SIDE configuration on a subset of your data file (especially if your data file is very large) or to tell C-SIDE to disregard non-data lines appearing at the end of your data file. When the **Define file format** window first appears, the word 'unlimited' appears in this item. To change this item, remove the word 'unlimited' by backspacing over it and replace it with a number. If you enter 0, a negative number, or a number greater than 1,000,000, then the number of lines read will not be limited (the word 'unlimited' will reappear the next time this window is opened).

Save:
This command button saves any of the changes that you made to this window and closes the window. Until you click the Save button, none of the entries you have made since the last time the window was opened will be recognized by the rest of the program.

Reset:
Clicking on this button will restore all items on this window to their current 'saved' settings. All changes you made to this window since it was last opened will be discarded.

Dismiss:
Clicking on Dismiss will cause the **Define file format** window to disappear. All changes you made to this window since it was last opened will be discarded.

The Analyze menu: Define variables

This window is used to describe the variables in your input data file. Up to 100 variables may be specified in this window, with 10 being shown at a time. The vertical scroll bar along the right hand side of the window allows you to control which set of 10 entry rows out of the 100 total are displayed. All variables in your data file must be entered here.

Field or Columns:
If Variable-length (delimited) input fields were chosen on the **Define file format** window, then the first column will have the title "Field" and will consist of a list of numbers representing the order of the fields in the data file. If Fixed-length input fields were chosen on the **Define file format** window, then the first column will have the title "Columns" and will allow you to specify the starting column and length of the fields in the data file. Entries of fixed-length field column specifiers can be made in one of three formats (X, Y, and Z represent integer numbers):

     X   A single character field at column X.

Define variables

X-Y   Field spanning columns X through Y (Y > X).
X:Z   Field starting at column X with length Z.

After entering variable definitions on the **Define variables** window, you may find that you need to insert a forgotten variable into or delete a variable from the middle of the list. This can be accomplished by editing the Field entry. To insert a variable, simply fill in the Name and Type information on a free row and then change the Field entry for that row to the necessary value. To place the new (Variable-length) variable between the fields previously defined as 5 and 6, change the Field value of the new variable to be 5.5. The next time the list is saved, the new variable will be sorted into place and the Field numbers will automatically be renumbered to whole numbers.

If you wish to remove a row, change the Field number (for Variable-length files) or start column number (for Fixed-length files) to be less than or equal to zero. The entries on that row will be removed when the window is saved (Variable-length Field numbers will automatically be adjusted).

Name:
Enter the names of the variables in this column. These names will be used for variable identification on the output and in other windows. To avoid confusion (to both the user and the program), each variable name must be unique. The maximum length of each variable name is 15 characters. If you enter variable names that are longer than the space available in the text entry box, you will not be able to see the entire variable name unless you expand the width of the window using the resize widget on the title bar.

Type:
A column of oval command buttons appears under Type. Click and hold pointer button 1 on each command button to activate a menu. Slide the pointer to your selection and release the pointer button. Your selection will then be displayed on the menu button. Your choices are:

Nutrient: Dietary components (having real nonnegative values), for which a distribution is to be estimated using the nutrient method. The observed distribution of nutrient intakes must be smooth, without a large number of observed zero intakes.

Food: Intakes with nonnegative, real values for which distributions are to be estimated using the food method. Food data generally contain a large number of observed zero intakes.

Individual ID: A number or string no more than 15 characters long that uniquely identifies an individual in the data file. There can be only one

| Define variables |

individual ID field in the data file. The data file must be sorted so that all observations for the same individual are grouped together.

Sample day: An integer variable representing the order of observations. You can use the sample day variable in cases where either the observations in the data set do not appear in the same order that they were taken, or to indicate which observations for an individual are missing from the data set. In the latter case, if an individual is expected to have three observations but has only two, the observations may be numbered 1 and 2 to indicate that the third observation is missing, or 1 and 3 to indicated that the second is missing. There can be only one sample day variable defined per input file. If no sample day is defined for a data file, then the observations will be numbered corresponding to the order they appear in the data file. C-SIDE uses sequence of observation information during ratio adjustment, adjustment of the variance, and during calculation of variance components. A sample day variable is considered a **Class, obs** variable (defined below) for program purposes.

Weight: A variable with nonnegative, real values representing sampling or replication weights for an individual. Weight values should not differ among observations on the same individual.

Class, indv: A categorical variable that may be used for ratio adjustment. Values can be a number or string no more than 15 characters long, and are associated with the individual and may NOT change for different observations on the same individual.

Class, obs: A categorical variable that may be used for ratio adjustment. Values can be a number or string no more than 15 characters long, and MAY change with each observation. Examples of common class variables are indicators for weekday versus weekend, month of the year, and interview sequence.

Continuous, indv: A continuous variable that may be used for ratio adjustment. It must be a real number or integer and usually has a large set of possible values in the data file. Values are associated with the individual and may NOT change for different observations on the same individual.

Continuous, obs: A continuous variable that may be used for ratio adjustment. It must be a real number or integer, and usually has a large set of possible values in the data file. Values are associated with the observation and MAY change for different observations on the same individual.

Nonanalysis: A variable that does not fit any of the other definitions and is not actively used in the analysis.

Save:

This command button saves any of the changes that you made to a window and closes it. Until you click the Save button, none of the entries you have made since the last time the window was opened will be recognized by the rest of the program. Always save the **Define variables** window before opening the **Assign variables** window in order to prevent confusion.

Reset:
Clicking on this button will discard any changes you made on this window since it was last opened and will cause the window to be displayed as if it were just opened. (If input field type was changed and applied in the **Define file format** window since the **Define variables** window was last opened, then Reset will cause the Field/Columns column to appear correctly.)

Dismiss:
Clicking on Dismiss will cause the **Define variables** window to disappear. All changes you made to this window since it was last opened will be discarded.

The Analyze menu: Define replication weight sets

The Define replication weight sets window is used to identify a set of replicate weights to be used for jackknife or BRR variance estimation.

<hr>

Define replication
weight sets

The linearization estimates of standard errors reported by C-SIDE for the usual intake percentiles and CDF values are only valid for data sets obtained from a simple random sample of individuals. In cases where the assumptions of simple random sampling are not met, such as when the data are obtained from a complex survey design including stratification and clustering, replication methods can be used to estimate appropriate variances.

C-SIDE supports the use of both the jackknife and the balanced repeated replication (BRR) method for approximating variances of estimates obtained for nutrient data (replication is not supported for food data). Both of these methods compute estimates of the parameter of interest from each of several subsamples of the parent sample. The estimated variance of the parent sample estimator is estimated by the variability between the subsample estimates. See Wolter (1985) for a more thorough discussion of replication methods for variance estimation.

The replicate weights must be created outside of C-SIDE and provided with the input data. See Dodd (1996) for instructions on how to create a set of replicate weights from a parent or reference weight.

Name of weight set:
Enter a name used to uniquely identify this weight set. This name will appear on the list of assignable weight variables on the **Assign variables** window and should be different from the weight variable names assigned using the **Define variables** window. To perform an analysis using replication, you will simply select this weight set from the **Assign variables** window in a manner similar to the selection of a regular sampling weight variable.

Reference weight:
This is the parent sample weight variable. Press and hold pointer button 1 on the command button beside the label. A menu showing the weight variables from the **Define variables** window is displayed. Select the variable you wish to use as the reference weight and release the pointer button.

Type of weight set:
Click on 'BRR' if the replication method to be performed is balanced repeated replication. Click on 'JK' if the replication method is jackknife.

Replicate weights:
The names of the weight variables that you previously defined using the **Define variables** window will be shown on the matrix of command buttons below the label. Select a variable to be used as a replicate weight by clicking pointer button 1 on the appropriate command button. The command button will change color and a check mark will appear in the box to show that the variable has been selected. Click on the command button a second time to deselect the variable.

If your data set contains more weight variables than can be displayed on the available command buttons, then <prev page> and <next page> will appear on the first and last command buttons. Click on one of these buttons to cause the matrix of command buttons to display another set of weight variable names.

Define replication weight sets

```
┌─────────────────────────────────────────────┐
│ ▣  ✕  cside – Define replication weight set  ⌐│
│                                               │
│              Record 3 of 3                    │
│   Name of weight set:  BR15                   │
│    Reference weight:  Weight                  │
│   Type of weight set:  BRR   JK               │
│    Replicate weights:                         │
│  <prev page>   ☐ BRW11     ☐ BRW17            │
│  ☐ BRW6        ☑ BRW12     ☐ BRW18            │
│  ☐ BRW7        ☐ BRW13     ☑ BRW19            │
│  ☑ BRW8        ☐ BRW14     ☐ BRW20            │
│  ☐ BRW9        ☑ BRW15     ☐ BRW21            │
│  ☑ BRW10       ☐ BRW16     <next page>        │
│                                               │
│     (Done)       (Prev)       (Next)          │
└─────────────────────────────────────────────┘
```

Done:
When you have finished defining replicate weight sets, click on the Done command button to cause the window to disappear.

Prev:
Next:
If you have defined more than one replication weight set, clicking on the Prev or Next command button will display a different weight set on the **Define replication weight sets** window. Continually clicking on Prev or Next will cause the display to cycle through the complete list of replication weight sets, including a blank record upon which you may enter a new replication weight set definition.

To delete a weight set definition, either erase the weight set name or select the blank entry on the reference weight menu. The record will be discarded when the Done command button is selected.

The Analyze menu: Assign variables

The Assign variables window is used to indicate which variables you wish to analyze. Once a data file is first described using the Define file format and Define variables windows, many different combinations of analyses may then be performed on that file simply by changing the information on the Assign variables window.

Report title:
At the top of the window is a box where you may optionally enter a customized title for the report that will appear in the output window. Text will automatically wrap to fit.

Runs form:
The rectangular area from **Run number:** to **Cutoffs:** and stretching to the right edge of the window is called the runs form, and is used to identify how many distributions will be estimated and the variables used in each analysis. Each separate estimation of distribution(s) for one analysis variable is called a run. The specific variables and values used for each run are contained in a numbered vertical box (called a run box) inside the runs form. Titles for the entry areas are along the left hand side, outside the form. There is room in this form to enter 10 runs, only 4 of which are displayed at one time. Use the horizontal scroll bar at the top of the form to view the other runs.

Run number:
The run number is displayed at the top of each vertical run box. Run 1 will be analyzed by the program first, run 2 second, and so on. The run number is for informational purposes and cannot be edited.

Analysis:

Press and hold pointer button 1 on the appropriate analysis command button to display a menu of the analysis variables. These are the variables that you defined as **Nutrient** or **Food** variables using the **Define variables** window. Select the name of the variable for which you wish to estimate a distribution and release the pointer button. To erase a variable name appearing in a run box, invoke the menu and release the pointer button on the first menu choice, which is blank. If you select a variable defined as a **Nutrient** variable, a nutrient analysis will be performed. If you select a variable defined as a **Food** variable, a food analysis will be performed.

Assign variables

Weight:

Position the cursor on the weight command button under the appropriate run. Press and hold the pointer button to display a menu of the weight variables. These are the variables that you defined as **Weight** variables using the **Define variables** window or as replicate weight set using the **Define replication weight sets** window. Select the name of the sampling weight variable or the replication weight sets that you wish to use for the analysis and release the pointer button. To erase a variable name appearing in a run box, simply call up the menu and release the pointer button on the first blank choice on the menu.

Subpopulation:

To perform an analysis on a smaller subset of data based on the value of a categorical variable, assign a subpopulation variable. Call up a menu by pressing and holding on the command button beside the subpopulation label and under the appropriate run. Select the desired variable name and release the pointer button. Your choices are limited to the variables defined as **Class, indv** or **Class, obs** on the **Define variables** window. For this feature to function properly, you must also enter a value for the subpopulation variable in the text entry area below.

Subpop value:

Type in the value of the subpopulation variable to be used for this run. You will need to determine the possible values of the subpopulation variable shown above and enter ONE of the values here. For an observation to be used in the analysis, the value in the file must EXACTLY match the value you enter here. (To run an analysis on a different value of a subpopulation variable, enter the appropriate information in another run box.)

Correlation:

If your data set consists of two or three daily observations taken consecutively and if you want the software to account for the correlation between observations, then you can supply the value of the between-day correlation here. Simply type in a correlation value between -1 and 1. (Correlation entries are used in estimating the consumption day usual intake distribution during a food analysis, but the estimated consumption probability distribution does not account for correlation among days. The

Assign variables

resulting estimated usual intake distribution for all days is likely to be approximately unbiased.)

Cutoffs:
If you want the software to estimate the proportion of the population with intake above and below certain cutoff values, you can specify the cutoff values here. There is room to enter a maximum of 10 values, all contained within a box that has a vertical scroll bar along the right side. If you enter more than two cutoff values, you will need to use the scroll bar to move more empty text entry areas into view.

Class:
To adjust the analysis variables for the effects of one or more classification variables, supply the appropriate variable names here. The classification variables entered here apply to all runs specified in the runs form. Click and hold on one of the oval command buttons to display a menu of variables. These are the variables that you defined as **Class, indv**, **Class**, **obs**, **Continuous**, **indv**, and **Continuous**, **obs** using the **Define variables** window. Select the name of a variable that you wish to use in ratio adjustment and release the pointer button. To erase a variable name, simply call up the menu and release the pointer button on the first blank choice on the menu. There is room for ten variables within a box, with a vertical scroll bar along the right side. Use the scroll bar if necessary to bring more unused command buttons into view.

Using too many class variables in an analysis can lead to execution errors caused by attempts to invert large matrices. The combined number of levels for all class variables used in an analysis should not exceed 50.

Save:
This command button saves the changes that you made to this window and closes the window. Until you click on the Save button, none of the entries you have made since the last time the window was opened will be recognized by the rest of the program. You should always either Save or Dismiss the **Assign variables** window before selecting **Execute** from the Analyze menu.

Reset:
Clicking on this button will discard any changes you made on this window since it was last opened and will cause the window to be displayed as if it were just opened.

Dismiss:
Clicking on Dismiss will cause the **Assign variables** window to disappear. All changes you made to this window since it was last opened will be discarded.

If you change variable names or types in the **Define variables** window, the **Assign variables** window is NOT automatically updated to include the new

names or types, and you will have invalid variables appearing in the **Assign variables** window. You should use the menus to reselect the variables that have been modified to insure that they are correct. The program will ignore variables appearing on the **Assign variables** window that are no longer valid. This is likely to cause your analysis to be performed in a manner different from that which you intended.

The Analyze menu: Parameters

This window allows you to modify some of the parameters used in the C-SIDE methodology. In general, you will not need to modify the default values.

Fraction of mean added:
The C-SIDE method requires all input data to be non-zero. The value entered here specifies the fraction of the mean that is added to each observed value to shift the entire distribution away from zero. The default value of 0.0001 will cause one ten-thousandth of the mean to be added to each observation (which should be adequate for most data sets).

Power transformation limit:
A procedure used in the C-SIDE method involves transforming a data set in an attempt to make it as normal as possible. A series of exponents and the natural log is applied to the data, and the best of those is chosen as the transformation. The range of exponents used in the power transformation is: 1, 1/1.5, 1/2, 1/2.5, ..., 1/maximum. The maximum is specified here. For most data the default value of 10 is more than sufficient; only highly skewed, heavy-tailed distributions may require a larger value.

Maximum join points/parameters:
Enter the maximum number of join points allowed when fitting a grafted polynomial function to a normal probability plot. A grafted

Parameters

polynomial defined over $p + 1$ segments has p join points and p coefficients. If the normal probability plot of your data has sharp curves near the middle of your data, or if you specify a very large value for the Type I error rate for Anderson-Darling tests, then you may have to use many join points to get an acceptable semiparametric normality transformation.

Observations in end segments:
This controls the size of the first and last segments in the grafted polynomial fitting. The default value of 2 specifies that the fitted functions should be linear for the two smallest and the two largest observations. The remaining observations are allocated to cubic segments. A minimum of two observations will always be used in each linear segment.

Type I error rate for A-D tests:
This is the Type I error rate (alpha in most statistics textbooks) used to test significance with an Anderson-Darling test for normality. Changing this value may cause a different number of parameters to be used in the grafted polynomial fit. Press and hold down on the command button to bring up a list of acceptable values. Point to the desired value and release the pointer button.

Values larger than 0.15 correspond to approximate critical values for the Anderson-Darling statistic, but other values correspond to the tabled critical values given by Stephens (1974).

Correct for heterogeneous variances?
Click on the 'Yes' value to have the program adjust for the effect of heterogeneous within-individual variances when estimating the usual intake distribution. Click on 'No' to not do these adjustments. You should first perform an analysis with this option set to 'Yes'. If the t-test for the presence of heterogeneous variances is not significant, then you might choose to run the analysis a second time with this option set to 'No'.

Adjust to the first sample day?
Click on 'Yes' to use the first sample day as the standard for the ratio and variance adjustments performed as part of the preliminary data adjustments. Click on 'No' to use the grand mean as the standard. Other sample days besides the first can be used as the standard by selecting 'Yes' and then resorting the data set or changing values of the sample day variable so that the desired day will be identified as day one.

Save:
This command button saves any of the changes that you made to this window and closes the window. Until you click the Save button, none of the entries you have made since the last time the window was opened will be recognized by the program.

Reset:

Clicking on this button will discard any changes you made on this window since it was last opened and will cause the window to be displayed as if it were just opened.

Dismiss:
Clicking on Dismiss will cause the **Parameters** window to disappear. All changes you made to this window since it was last opened will be discarded.

The Output menu: Output data files

This window can be used to create files containing data generated at certain intermediate steps in the C-SIDE methodology.

Files are written with ASCII characters, one record per line, with a single space between fields. The first five entry areas on this screen are command buttons which cause the **File selection** window to appear. Click on the appropriate button and enter the name of the output data file in the window that appears. (The **File selection** window is described in another part of the documentation.) The selected file name will then appear on the appropriate command button.

*If you execute a series of runs and then wish to alter entries in an Analyze window and execute another series of runs, then you must change output file names to prevent the files created with the first execute from being overwritten. Each time the **Execute** menu option is selected, any old files with the selected file names will be destroyed and new output files will be created in their place.*

Output data files

If you have previously entered an output file name and later decide you do not wish a file to be created, you must erase the file name. Call up

the **File selection** window and erase the directory name appearing in the Selection box. Selecting **Ok** will cause the file name appearing on the command button to be erased.

Adjusted daily intakes
This data set contains results of the preliminary adjustments applied to the observed intakes before the usual intake estimation is performed.

Output variables include:
    Run number
    Analysis variable name
    Subpopulation variable (if applicable)
    Individual ID variable
    Any class variables used in the analysis
    Ratio-adjusted dietary components
    Equal-weight dietary components

    Variables appear only if they were specified on the **Assign variables** window for that run and, in the case of dietary component variables, only if that part of the analysis was actually performed. Exact ordering of output variables will be printed under the DESCRIPTIVE STATISTICS portion of the output.

Daily intake prob plot data:
These data are generated when estimating the grafted polynomial function in the semiparametric transformation into normality.

The order of the variables in the output data set is
    Run number
    Analysis variable name
    Subpopulation variable (if applicable)
    Power-Transformed scale adjusted daily intakes
    Normal scores
    Normal scale daily intakes

    Plotting the normal scores against the power-transformed scale adjusted daily intakes (which are shifted, power transformed, equal-weight, ratio- and variance-adjusted dietary components) produces a normal probability plot of those intakes. Plotting the normal scale daily intakes against the power-transformed scale adjusted daily intakes produces a plot of the fitted grafted polynomial. If you overlay the plots just described, you can see how well the grafted polynomial fits the data. If the fit looks poor, you may wish to repeat the analysis, adjusting some of the values on the **Parameters** window.

Representative usual intakes: Output data files
This data set consists of a set of 400 values that approximate the

distribution of usual intakes (for consumption days only for foods). They do not correspond to estimates of usual intakes for the individuals in the input data set.

The order of the variables in the output data set is
      Run number
      Analysis variable name
      Subpopulation variable (if applicable)
      Observation

Usual intake percentiles:
The order of the variables in the output data set is
      Run number
      Analysis variable name
      Subpopulation variable (if applicable)
      Probability value for which the percentile is computed
For nutrient analysis there are three more variables:
      Percentile for the usual intake distribution
      Density function for the usual intake distribution
      Estimated standard error of the percentile
For food analysis there are five more variables:
      Percentile for the usual intake distribution for consumption days
      Density function for the usual intake distribution for consumption days
      Percentile for the usual intake distribution on all days for consumers
        only
      Density function for the usual intake distribution on all days for
        consumers only
      Percentile for the usual intake distribution on all days for the entire
        population
      Density function for the usual intake distribution on all days for the
        entire population

The linearization standard errors included for a nutrient analysis are computed under the assumptions of simple random sampling and nonconsecutive intake days. A properly prepared replicate weight set may be used to compute appropriate standard errors when these assumptions are not met.

The program will produce 41 pre-selected percentiles (0.01, 0.025, ..., 0.975, 0.99) plus the number specified by Additional percentiles as described below. Plotting the probability value against the percentile produces a plot of the CDF for the distribution of usual intakes. Plotting the density function value against the percentile produces a plot of the probability density function (pdf) for the distribution of usual intakes.

Individual usual intakes:
This data sets consists of estimated usual intakes for each of the individuals in the input data file. This data set is only applicable for a nutrient analysis.

The order of the variables in the output data set is
    Run number
    Analysis variable name
    Individual ID variable
    Estimated individual usual intake

These individual usual intakes are NOT used in the calculation of the usual intake distribution. They are obtained from auxiliary calculations, and by their nature are much less precise than estimates of overall characteristics of the usual intake distribution (e.g., moments and percentiles). Keep this fact in mind when interpreting the individual usual intakes C-SIDE produces.

Additional percentiles:
Enter the number of additional percentiles you will generate for the **Usual intake percentiles** data file. The additional percentiles are chosen to be equally spaced in probability from $1/(n+1)$ to $n/(n+1)$ in steps of $1/(n+1)$. If you want the 1st through 99th percentiles, enter a value of 99.

Save:
This command button saves any of the changes that you made to this window and closes the window. Until you click the Save button, none of the entries you have made since the last time the window was opened will be recognized by the program.

Reset:
Clicking on this button will discard any changes you made on this window since it was last opened and will cause the window to be displayed as if it were just opened.

Dismiss:
Clicking on Dismiss will cause the **Output data files** window to disappear. All changes you made to this window since it was last opened will be discarded.

The Output menu: Save output window

This item allows you to save the contents of the output window (the middle section of the main window) to a file. Click and hold on the Output menu button and draw the pointer to **Save output window**. Release the pointer button and a **File selection** window appears. Use this window to enter a file name.

The Output menu: Clear output window

Use this item to erase the output window. Upon selection of this item, a small dialog window will appear:

Clear output window – are you sure?

(Yes) (Cancel)

The Configure menu

The Configure menu: Open

Use this menu item to read in the contents of a configuration file. A configuration file contains problem setup information that was previously entered in C-SIDE Analyze and Output windows and then saved with the **Save** menu item. Configuration files allow you to save time and effort when

- Rerunning an analysis
- Modifying a previously run analysis
- Performing a different analysis on a data file already described
- Describing a new data file that is similar to one already used

After choosing the **Open** menu item, a **File selection** window will appear. Use the window to choose the desired configuration file. A default mask of '*.cfg' is provided for you (it is assumed all configuration file names end in '.cfg'; see **Save** below). C-SIDE will read configuration information from the file and will place the configuration file's name in the upper right hand corner of the main window.

The Configure menu: Reset

This item will clear out all entry areas and command buttons on all windows. All items on all windows will be set back to their default state, erasing any configuration information previously entered. Upon selection of this item, a small dialog window will appear:

Reset configuration data – are you sure?

(Yes) (Cancel)

The Configure menu: Save

This menu item will allow you to store all the information entered in the Analyze and Output menu windows into a file. The program's current state can be restored by using the **Open** menu item to read in the saved configuration file. Choose the **Save** item from the Configure menu and enter a file name in the **File selection** window provided. You should end all configuration file names with the characters '.cfg' to take advantage of the default mask of '*.cfg' used when opening files, and to make it easier to identify your configuration files later.

Execute

After the other items on the Analyze menu have been selected and their corresponding windows filled in, then you may use this item to begin the analysis. The edits you make to any of the windows in the program are not recognized until you click on the Save button for that window. Therefore, it is a good idea to close all other windows before selecting the **Execute** item. Analysis will begin and output should start to appear in the output window. You will not be able to interact with the program further until the analysis has been completed. The status of the output window and messages displayed in the diagnostic window will allow you to keep track of the program's progress in performing the analysis.

File selection

The **File selection** window consists of 5 major parts. The first is the Filter text entry box at the top of the window. The next is the Directories list below the Filter and on the left side of the window. The Files list is to the right of the Directories list. The Selection text entry box is below the two lists, and the command buttons are at the bottom. To enter a file name in the Selection box, place the pointer inside the Selection box and type a file name. Press enter in the Selection box or click on the Ok command button to cause the window to accept your file name and close. As an alternative to just typing the name of the file, you may use the Filter box, Directories list, and Files list to help you choose an existing file name.

cside – Open config file

Filter

/local/home/username/*.cfg

Directories   **Files**

.   ccdata.cfg
..   csf2cm.cfg
Data   csfnut2.cfg
Mail   example1.cfg
Out   example2.cfg
bin   example3.cfg
temp   om89_91.cfg
                         om89_91f.cfg
                         yw85.cfg

Selection

/local/home/username/

(Ok)   (Filter)   (Cancel)
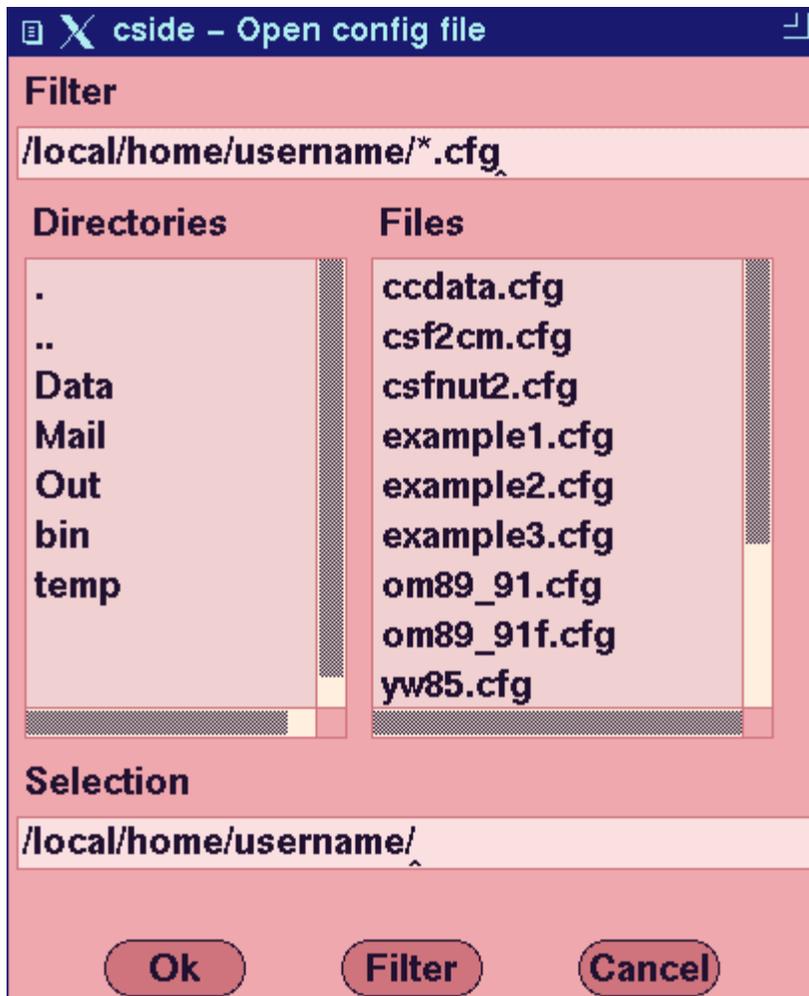
The File selection window is accessed from several C-SIDE windows and is used to specify a file name. It is used by the Open and Save menu items to name configuration files, by the Define file format window to provide the input data file name, and by the Output data files window and Save output window menu option to name new output files. The window functions in the same way, independent of where it is accessed.

The Filter controls what is displayed in the Directories and Files lists. These lists show all of the directories and some or all of the files contained in a particular directory on your file system. The name of the directory, and a mask showing which files are to be displayed, appear in the Filter text area. You can change the directory name by typing a new name in the Filter box and then either pressing enter or clicking on the Filter command button. You must end all directory names with a slash. You may also choose a directory that appears in the Directories list by double clicking on it. The Filter box and the two lists will change to reflect the new current directory. In conformance with standard file system syntax, you can move up one directory level by double clicking on the double dot directory entry.

The mask in the Filter box always follows the last slash marking the end of the directory name. The Filter for C-SIDE consists of one or more characters and an asterisk. The asterisk can be at the beginning of the mask, at the end, or both at the beginning and at the end. If the asterisk is at the beginning of the mask, then only file names ending with the exact characters following the asterisk will be displayed in the Files list. If the asterisk is at the end, then only files beginning with the exact characters preceding the asterisk will be displayed. A set of characters between two asterisks will

cause only files having that sequence of characters somewhere in their name to be displayed. The default mask of a single asterisk will display all files in the current directory.

Once you have located the file you wish to choose in the Files list, double click on it. You may also single click on it and then click on the Ok command button. The file name will be selected and the window will disappear.

If you are using the **File selection** window to name a new output file, and if a file with the name you choose already exists, then the following dialog box will appear:

Choose **'Yes'** to destroy the existing file and replace it with the new file created by C-SIDE. Choose **'Cancel'** to choose a different file name.

File exists. Do you wish to overwrite it?
Yes **Cancel**
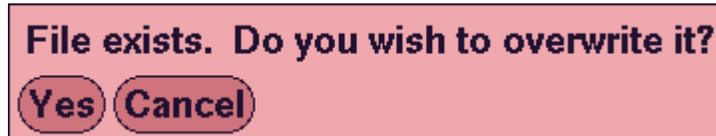
The Cancel command button at the bottom of the window allows you to close the window without selecting a file name.

If the full pathname for a file becomes too long to fit in the selection text entry box, simply expand the width of the box using the resize widget on the title bar.

III: OUTPUT REPORT

Output produced by the program during an analysis is displayed in the Output window. Information in this window can be saved to a file using the **Save output window** option that appears on the Output menu. The following is a description of each section of output. The output for a nutrient analysis differs from that for a food analysis. Example output is presented for nutrients first, and then for foods.

Nutrients

```
C-SIDE
Software for Intake Distribution Estimation
Version 1.0, September 1, 1996
Department of Statistics and
Center for Agricultural and Rural Development
Iowa State University
Copyright (c) 1996 Iowa State University,
Statistical Laboratory, All Rights Reserved.


User supplied title goes here.
```

The first section contains the date and time the analysis was begun, the software's title and version number, and a copyright notice.

Nutrients Output

```
PROBLEM IDENTIFICATION
Data file is /local/home/username/csfii85.txt

  Variables used in the analysis:
      Field        Type                Name
      1st          Individual ID       INDIV
      2nd          Class, obs          SEQUENCE
      3rd          Nutrient            KCAL
      10th         Class, obs          DOW
      11th         Weight              INDWT

Variable used to determine sample day:              SEQ
Variable used to tie observations to individuals:   INDIV
Fraction of mean used to adjust for zero values:    0.0001
Minimum power tested for best transformation:       1 / 10
Maximum number of join points allowed:              12
Number of observations in linear end segments:      2
Type I error rate used in Anderson-Darling tests:   0.15
```

The problem identification section contains information regarding data file, variables, and parameters used in the analysis. Some information is only printed if it differs from the program default. This section can be used to verify that the problem was identified correctly.

The following sections will be printed once for each individual run shown on the **Assign variables** window. Results are printed in the same order that computations were performed.

```
Nutrient KCAL (weighted by INDWT).

PRELIMINARY ADJUSTMENTS
Transformation is power of 1 / 2, scale factor of 10e1.
The maximum number of observations per individual is 4.

Ratio adjustment variables are:
       DOW, and SEQUENCE.

Data are adjusted to Day 1 variance after ratio adjustment
and in transformed scale.  Original scale standard
deviations for adjusted data each sample day are:
          Day 1        729.191
          Day 2        782.004
          Day 3        756.602
          Day 4        854.61
```

```
DESCRIPTIVE STATISTICS     Original Data   Adjusted Data
  Total number of observations      2948            2948
  Total number of individuals        737             737
  Mean                            1496.3          1689.3
  Minimum                            2.4          9.6793
  Maximum                         5847.7          6302.5
  Variable order in adjusted daily intakes output file
is: (run) (label) INDIV DOW SEQUENCE INDWT (ratio-
adjusted) (variance-adjusted) (equal-weighted)
```

Nutrients Output

      C-SIDE displays the analysis variable name along with weight variable name and subpopulation variable name and value if applicable.

      The initial power applied to the observed intakes to transform them to approximate normality is printed along with the power of 10 used to scale the data. The maximum number of observations per individual, as determined by the program from the actual data, is printed.

      If ratio adjustment was performed, then the actual variables used in this process are shown next.

      The standard deviation of each day's data is printed, and the variance of each day's data is adjusted to match the variance of the first day.

      Descriptive statistics are calculated for the original data and for the data resulting from ratio adjustment and variance adjustment. In this example an adjusted daily intakes data set was created, so the order of the variables in that data set is shown directly below the descriptive statistics.

```
NORMALITY TRANSFORMATION
Transformation is power of 1 / 2, scale factor of 10e1.
Normality test hypothesis not rejected at 4 join points.
The Anderson-Darling test statistic of 0.551 was less
than the 15% critical value of 0.576
```

      Results from the transformation that carries the adjusted daily intakes into normality come next. A grafted polynomial function is fit to the data, with the transformed adjusted data as the dependent variable and corresponding normal scores as the independent variable. The number of join points used is started at 3, and then increased as needed (up to a maximum) until an Anderson-Darling test for normality does not fail. The power applied, scale factor, number of join points in the resulting model, and the Anderson-Darling test statistic are printed. If no model was able to achieve the critical value of the test statistic, then an appropriate message will be printed, and there will be no further analysis for this run.

```
NORMALITY TESTS FOR TRANSFORMED DATA
   ADS 5% Critical Value                            0.787
   ADS for Observed Intakes                         0.551
   ADS for Day  1 Intakes    (737 observations)     0.318
   ADS for Day  2 Intakes    (737 observations)     0.331
   ADS for Day  3 Intakes    (737 observations)     0.394
   ADS for Day  4 Intakes    (737 observations)     0.435
   ADS for Individual Mean Intakes                  0.360
```

Anderson-Darling tests for normality are applied to the predicted normal scores from the grafted polynomial fit. Included are a reference value, the Anderson-Darling test statistic for the complete set of transformed observations, the Anderson-Darling test statistic (ADS) for each day's transformed observations, and the Anderson-Darling test statistic for the individual means of the transformed observations.

```
VARIANCE COMPONENTS FOR TRANSFORMED DATA
Number of individuals                                  737
Number of individuals with more than 1 day            737
Mean of individual means                           0.00060
Variance of transformed observed intakes            1.0021
Variance of usual intakes                           0.37758
Variance of measurement error                       0.62555
Measurement error degrees of freedom                  2208

Diagnostics for heterogeneous individual variances
T-test for presence of heterogeneous variances 2.3900
 Test for linear relationship between
    individual mean and standard deviation
   Slope of std dev regressed on mean        0.00029019
   T-test for slope = 0                        0.016875
   Pr > |t|    (on    735 d.f.)                0.98654
```

This section displays statistics calculated as part of Step 3 in the methodology.

In the Diagnostics for heterogeneous individual variances subsection, a t-statistic for testing the null hypothesis of homogeneous within-individual variances is given. Values larger than 2 indicate that some evidence of heterogeneous within-individual variances is present. C-SIDE computes the sample standard deviation and sample mean of the transformed observations for each individual with more than one observed intake. A regression model containing constant and linear terms is fit to the resulting paired data, with the standard deviations as the dependent variable and the means as the independent variable. The results of the regression along with the associated t-test are printed.

```
ESTIMATED MOMENTS FOR USUAL INTAKES IN ORIGINAL SCALE
  Mean            1690.3
  Variance        197691
  Std. Dev.       444.62
  Skewness        0.54299

ESTIMATED PERCENTILE VALUES FOR USUAL INTAKES
  Probability     Percentile     Std. Error
     0.010           801.           38.3
     0.025           921.           36.0
     0.050          1028.           33.6
     0.075          1100.           31.9
     0.100          1156.           30.6
     0.125          1204.           29.5
     0.150          1245.           28.6
     0.175          1283.           27.8
     0.200          1318.           27.1
     0.225          1351.           26.5
     0.250          1382.           26.0
     0.275          1411.           25.6
     0.300          1440.           25.3
     0.325          1468.           25.0
     0.350          1495.           24.9
     0.375          1522.           24.8
     0.400          1548.           24.7
     0.425          1575.           24.8
     0.450          1601.           24.9
     0.475          1627.           25.0
     0.500          1654.           25.2
     0.525          1681.           25.5
     0.550          1708.           25.9
     0.575          1735.           26.4
     0.600          1764.           26.9
     0.625          1793.           27.5
     0.650          1823.           28.3
     0.675          1854.           29.1
     0.700          1887.           30.1
     0.725          1921.           31.3
     0.750          1958.           32.6
     0.775          1997.           34.1
     0.800          2040.           35.9
     0.825          2086.           38.0
     0.850          2139.           40.4
     0.875          2198.           43.3
     0.900          2269.           46.9
     0.925          2357.           51.6
     0.950          2477.           58.1
     0.975          2672.           68.9
     0.990          2918.           82.4
```

Estimated percentiles and standard errors for 41 probability values are printed. Unless replication is used, these standard errors are approximations of the standard errors of the appropriate quantities computed under the assumption of simple random sampling of individuals on nonconsecutive days. If standard errors cannot be calculated, then a row of asterisks will appear in their place. If replication is used, then standard errors will also be shown for the moments.

```
CDF Values for Usual Intakes
    Value  Prob. Below  Prob. Above  Std. Error
     2000     0.7767       0.2233       0.0247
```

The proportion of the population with intake below, and above, the specified threshold value are also printed.

Foods

```
C-SIDE
Software for Intake Distribution Estimation
Version 0.99, August 29, 1996
Department of Statistics and
Center for Agricultural and Rural Development
Iowa State University
Copyright (c) 1996 Iowa State University
Statistical Laboratory.  All Rights Reserved.
```

```
PROBLEM IDENTIFICATION
  Data file is /local/home/username/food85.txt

  Variables used in the analysis:
      Field      Type              Name
      1st        Individual ID     INDIV
      2nd        Class, obs        DOW
      3rd        Class, obs        SEQ
      4th        Food              DVEG

Variable used to determine sample day:          SEQ
Variable used to tie observations to individuals: INDIV
Fraction of mean used to adjust for zero values: 0.0001
Minimum power tested for best transformation:    1 / 10
Maximum number of join points allowed:             18
Number of observations in linear end segments:      2
Type I error rate used in Anderson-Darling tests:  0.1
```

The first two sections of Foods output contains the same information as
Nutrients output.

Foods Output

```
Food      DVEG (not weighted).

SUMMARY STATISTICS
                              All        Positive
                            Intakes     Intakes
  Number of observations     2972         287
  Number of individuals       743         235

  Proportion of respondents who consumed food on:
          0 days       0.68371
          1 day        0.25168
          2 days       0.05922
          3 days       0.00538
          4 days       0.00000

T-test for correlation between intake and probability
of consumption        0.068974
```

The first table displays counts for the number of observations and
individuals in the data set, along with the number of positive
observations and the number of individuals with at least one positive
observation. The number of positive observations is counted for each
individual. The relative frequency of individuals who consume the food
on 0, 1, 2, ... sample days is shown. This data is weighted if a sampling
weight was provided.

A test for correlation between intake and the probability of
consumption is performed, with the resulting t-statistic shown.

Foods Output

```
ESTIMATED CONSUMPTION PROBABILITY DISTRIBUTION
  Value     Mass
   0.00    0.14432    ***************
   0.02    0.05903    ******
   0.04    0.05111    ******
   0.06    0.05834    ******
   0.08    0.07886    ********
   0.10    0.11254    ************
   0.12    0.14915    ***************
   0.14    0.16005    *****************
   0.16    0.11995    ************
   0.18    0.05351    ******
   0.20    0.01199    **
   0.22    0.00113    *
   0.24    0.00004    *
   0.26    0.00000
   0.28    0.00000
   0.30    0.00000
   0.32    0.00000
   0.34    0.00000
   0.36    0.00000
   0.38    0.00000
   0.40    0.00000
   0.42    0.00000
   0.44    0.00000
   0.46    0.00000
   0.48    0.00000
   0.50    0.00000
   0.52    0.00000
   0.54    0.00000
   0.56    0.00000
   0.58    0.00000
   0.60    0.00000
   0.62    0.00000
   0.64    0.00000
   0.66    0.00000
   0.68    0.00000
   0.70    0.00000
   0.72    0.00000
   0.74    0.00000
   0.76    0.00000
   0.78    0.00000
   0.80    0.00000
   0.82    0.00000
   0.84    0.00000
   0.86    0.00000
   0.88    0.00000
   0.90    0.00000
   0.92    0.00000
   0.94    0.00000
   0.96    0.00000
   0.98    0.00000
   1.00    0.00000
```

A distribution of 51 consumption probabilities is estimated from
the consumption proportions using methods described in Nusser, Fuller,
and Guenther (1996). The estimated distribution, along with a simple
graphical representation, is displayed. Note that the estimated mass for a
zero consumption probability is the estimated fraction of the population
that never consumes the food (fraction of nonconsumers in the
population).

Foods Output

The program then applies the same method for estimating usual
nutrient intake distributions to the positive food intake from the data set
to obtain a distribution for consumption days only.

```
ESTIMATING USUAL INTAKE DISTRIBUTION FOR CONSUMPTION DAYS

PRELIMINARY ADJUSTMENTS
  Transformation is power of 1 / 5, scale factor of 10e0.
  The maximum number of observations per individual is 4.

  Ratio adjustment variables are:
       SEQ, and DOW.

  Data are adjusted to Day 1 variance after ratio
adjustment and in transformed scale.  Original scale
standard deviations of adjusted data for each sample day
are:
           Day 1      113.622
           Day 2      116.63
           Day 3      80.921
           Day 4      99.8928

DESCRIPTIVE STATISTICS         Original Data  Adjusted Data
  Total number of observations    287               287
  Total number of individuals     235               235
  Mean                           3.97            81.698
  Minimum                        0.91           0.19034
  Maximum                         474            460.58

NORMALITY TRANSFORMATION
  Transformation is power of 1 / 4.5, scale factor of
10e0.  Normality test hypothesis not rejected at 4 join
points.  The Anderson-Darling test statistic of 0.124 was
less than the 10% critical value of 0.656


NORMALITY TESTS FOR TRANSFORMED DATA
  ADS 5% Critical Value                          0.787
  ADS for Observed Intakes                       0.124
  ADS for Day 1 Intakes (64 observations)        0.495
  ADS for Day 2 Intakes (63 observations)        0.329
  ADS for Day 3 Intakes (76 observations)        1.11
  ADS for Day 4 Intakes (84 observations)        0.326
  ADS for Individual Mean Intakes                0.283


VARIANCE COMPONENTS FOR TRANSFORMED DATA
  Number of individuals                            235
  Number of individuals with more than 1 day        48
  Mean of individual means                    -0.00386
```

```
Variance of transformed observed intakes        1.0007
Variance of usual intakes                       0.33850
Variance of measurement error                   0.66967
Measurement error degrees of freedom                 49

Diagnostics for heterogeneous individual variances
  T-test for presence of
    heterogeneous variances                     0.33789
  Test for linear relationship between
    individual mean and standard deviation
    Slope of std dev regressed on mean      0.030556
    T-test for slope = 0                        0.36056
    Pr > |t|    (on    46 d.f.)                 0.72008
```

Output for the above sections is identical to that for a nutrient analysis. The estimated consumption day usual intake distribution is presented in the next section, along with summaries of the estimated usual intake distribution for all days.

```
ESTIMATED MOMENTS FOR USUAL INTAKES IN ORIGINAL SCALE
          CONSUMPTION DAYS        ALL DAYS          ALL DAYS
                                  CONSUMERS          ENTIRE
                                    ONLY           POPULATION
  Mean          81.506            9.1387            7.8198
  Variance      2186.8            45.278            49.057
  Std. Dev.     46.764            6.7289            7.0041

  Percent nonconsumers      14.4
  Percent consumers         85.6
```

ESTIMATED PERCENTILE VALUES FOR USUAL INTAKES

Probability	Consumption Days Percentile	All Days Consumers Percentile	All Days Entire Pop Percentile
0.010	14.	0.7	0.0
0.025	18.	1.0	0.0
0.050	23.	1.4	0.0
0.075	27.	1.8	0.0
0.100	31.	2.1	0.0
0.125	34.	2.5	0.0
0.150	36.	2.8	0.6
0.175	39.	3.1	1.2
0.200	42.	3.5	1.6
0.225	44.	3.8	2.0
0.250	47.	4.1	2.5
0.275	49.	4.5	2.9
0.300	52.	4.8	3.2
0.325	54.	5.1	3.6
0.350	57.	5.4	4.0
0.375	59.	5.8	4.4
0.400	62.	6.1	4.8
0.425	64.	6.5	5.2
0.450	67.	6.8	5.5
0.475	69.	7.2	5.9
0.500	72.	7.6	6.3
0.525	75.	8.0	6.7
0.550	78.	8.4	7.2
0.575	81.	8.8	7.6
0.600	84.	9.2	8.1

```
    0.625              87.             9.7            8.6
    0.650              91.            10.2            9.1
    0.675              94.            10.7            9.6
    0.700              98.            11.2           10.2
    0.725             102.            11.8           10.8
    0.750             107.            12.5           11.4
    0.775             111.            13.2           12.1
    0.800             117.            13.9           12.9
    0.825             122.            14.8           13.8
    0.850             129.            15.8           14.8
    0.875             136.            16.9           15.9
    0.900             145.            18.2           17.3
    0.925             156.            20.0           19.0
    0.950             172.            22.3           21.4
    0.975             196.            26.3           25.4
    0.990             226.            31.5           30.6


CDF Values for Usual Intakes
Consumption Days
   Value   Prob. Below   Prob. Above
      10    0.003722         0.9963
     100      0.7114         0.2886
All Days - Consumers Only
   Value   Prob. Below   Prob. Above
      10       0.6415        0.3585
     100        1.000         0.000
All Days - Entire Population
   Value   Prob. Below   Prob. Above
      10       0.6932        0.3068
     100        1.000         0.000
```

Information from the estimated consumption probability distribution is combined with information from the estimated usual intake distribution for consumption days to obtain three distributions: consumption days (days with positive intake), all days (consumption and nonconsumption) for those who actually consume the food, and all days for the entire population (consumers and nonconsumers). The estimation procedure for the usual intake for all days assumes that the consumption day usual intakes are independent of the frequency with the food is consumed.

The mean, variance, standard deviation, a list of 41 pre-selected percentiles, and the proportion of individuals above and below and equal to the given cutoff thresholds are shown for the three distributions. The requested cutoff thesholds are shown for all three distributions, although usually applicable to either the Consumption Days distribution only or the All Days distribution only. The estimated percentage of consumers and nonconsumers from the estimated consumption probability distribution is also displayed.

IV: EXAMPLES

<div style="background: black; color: white;">Nutrients</div>

Basic example

This example performs a simple analysis on the csfii85.txt data file. These data are a subset of the data from the 1985 Continuing Survey of Food Intakes by Individuals (CSFII).

The following section provides step-by-step instructions for running the example program.

1. Copy the program and example files into a directory on your computer system.
2. Start the CSIDE program by typing "cside" from your UNIX command prompt.
3. Depress and hold down pointer button 1 on the Configure command button.
4. Move the pointer to highlight **Open** and release the pointer button. A **File selection** window will appear.
5. Double click pointer button 1 on the "example1.cfg" that appears in the Files box. The **File selection** window will disappear and "example1.cfg" will appear in the upper right hand corner of the main window.

(As a shortcut to starting the program and then using **Open** to specify the configuration file, type "cside example1.cfg" in step 2 and then go immediately to step 6.)

6. Click and release the pointer button on the **Execute** command button. Output should start appearing in the middle section of the main window.
7. Analysis is completed when "Done with KCAL." appears in the lower portion of the main window. An output file "ex1adi.out" with Adjusted Daily Intake data is created.

A copy of output produced by this example appears in the OUTPUT REPORT section of this manual.

Example using correlation and subpopulations

This example performs a more complex analysis than was performed in the basic example. The data set om89_91.txt is a subset of the 1989-91 CSFII and contains observations for male respondents age 60 years old and older. The individuals sampled were selected according to a complex survey design.

The variables included in the data set are described in the following table.

Variable	Description
YEAR	Year of survey: 1=1989, 2=1990, 3=1991
ID	Individual ID
SEQ	Interview sequence
MONTH	Month of first interview
WEEKDAY	Day of week
WEIGHT	Sampling weight
CALCIUM	Calcium intake (milligrams)
CALORIES	Energy intake (kilocalories)
IRON	Iron intake (milligrams)
PROTEIN	Protein intake (grams)

Table: Variable listing for om89_91.txt

These data were collected on three consecutive days, so it is reasonable to assume that the observations for each individual are correlated. This example demonstrates how to account for this correlation. Only one dietary component, energy intake, is analyzed. The analysis incorporates sampling weights, and demonstrates how to use a subpopulation variable and how to create output data sets. Ratio-adjustment for the effects of the survey-related class variables SEQ, MONTH, WEEKDAY, and YEAR is also performed.

This example can be executed in a manner similar to the basic example, except substitute "example2.cfg" for "example1.cfg" in the instructions.

From the C-SIDE main window, select the **Assign variables** window and the **Output data files** window to view the correlation, subpopulation, and output data set information entered as part of this example.

▣ X cside – Assign variables				⌐
Report title:	Analysis of a subset of the 1989–1991 CSFII containing observations for male respondents age 60 years and			

Run number:	1	2	3	4
Analysis:	CALORIES	CALORIES	CALORIES	
Weight:	WEIGHT	WEIGHT	WEIGHT	
Subpopulation:	YEAR	YEAR	YEAR	
Subpop value:	1	2	3	
Correlation:	0.17	0.16	0.13	
Cutoffs:				

Class: SEQ
        WEEKDAY
        MONTH

(Save)   (Reset)   (Dismiss)

Because three separate distributions are estimated, output from this example is approximately three times larger than the output from the basic example. Four output files are created by this example: adi.out, nppd.out, rui.out, and pct.out.

Foods

Basic example

This example performs a food analysis on 4-day data from the 1985 CSFII. The configuration file for this example is example3.cfg, and the data file is food85.txt.

The variables included in the data set are described in the following table.

Variable	Description
INDIV	Individual ID
DOW	Day of week of intake
SEQ	Sequence
DVEG	Dark green vegetables
APPL	Apples
ALCH	Alcoholic beverage
LPOP	Low-cal carbon beverage
EGGS	Egg
BEEF	Beef
FRUT	Fruits
MILK	Milk

Table: Variable listing for food85.txt

This example can be executed in a manner similar to the basic nutrients example. A copy of the output produced by this example appears in the OUTPUT REPORT section of this manual.

APPENDIX A:
Correlation Estimates

Estimated day-to-day correlations based upon measurement error model

Component	Males ≥ 20	Females ≥ 20	Persons 0-19
Calcium	0.119	0.131	0.119
Carbohydrates	0.138	0.178	0.164
Carotene	0.097	0.078	0.045
Cholesterol	0.069	0.005	0.027
Copper	0.126	0.150	0.106
Energy	0.126	0.150	0.120
Fiber	0.133	0.167	0.092
Folate	0.111	0.110	0.100
Iron	0.108	0.105	0.020
Magnesium	0.146	0.199	0.129
Monounsaturated fat	0.092	0.064	0.066
Niacin	0.106	0.100	-0.011
Polyunsaturated fat	0.093	0.065	0.045
Phosphorus	0.127	0.153	0.068
Potassium	0.137	0.178	0.122
Protein	0.106	0.100	-0.004
Riboflavin	0.119	0.132	0.100
Saturated fat	0.107	0.100	0.103
Sodium	0.098	0.078	0.026
Total fat	0.097	0.076	0.076
Thiamin	0.117	0.130	0.001

Component	Males ≥ 20	Females ≥ 20	Persons 0-19
Vitamin B6	0.124	0.144	0.077
Vitamin B12	0.118	0.127	0.104
Vitamin A	0.090	0.058	0.062
Vitamin C	0.109	0.108	0.098
Vitamin E	0.106	0.099	0.070
Water	0.141	0.187	0.150
Zinc	0.103	0.093	-0.006
Percent tfat	0.081	0.036	0.035
Percent mfat	0.072	0.014	0.007
Percent sfat	0.105	0.096	0.131
Percent pfat	0.087	0.052	0.034
Percent carb	0.086	0.048	0.068
Percent prot	0.096	0.076	0.017

tfat = total fat
mfat = monounsaturated fat
sfat = saturated fat
pfat = polyunsaturated fat
carb = calories from carbohydrates
prot = calories from protein

APPENDIX B:
Technical Overview
of Methodology

C-SIDE was developed to carry out the statistical method for estimating usual nutrient or food intake distributions. The nutrient method is described in Nusser, Carriquiry, Dodd, and Fuller (1996). The food method is described in Nusser, Fuller and Guenther (1996). Dodd (1996) provides documentation of the procedures implemented by the software. This appendix contains a technical overview of the different steps in the program. The nutrient procedure is described first.

Nutrients

The method for estimating usual nutrient intake distributions consists of several steps. The steps can be grouped according to the major tasks to be accomplished: preliminary data adjustments, semiparametric transformation to normality, estimation of within and between individual variances for intakes, and back transformation into the original scale.

1. **Preliminary data adjustments:**

Preliminary data adjustments include shifting observed intake data by a small amount away from zero, incorporating survey weights by creating an equal weights sample, and correcting for the effect of sample day (first versus all the rest) on the mean and the variance of the distribution of observed intakes.

To shift observed intake data away from zero, the program computes the weighted overall mean intake and adds one ten thousandth of the weighted mean to each observation in the sample. The magnitude of the value added to each observation is small, but it eliminates all zero observations, and avoids potential problems should a log transformation be used.

A regression-based adjustment to account for differences among sample days is applied to shifted observations. Since it is well known that regression estimates have better properties when data are normal, the shifted intake data are power transformed prior to estimating the regression adjustment. The best power transformation for each nutrient is found by using a grid search, and choosing the power that minimizes the departure of the transformed data from normality. A regression model is fit to the shifted power transformed data, where the independent variables in the model include adjustment variables such as day of the week, month of the year, interview sequence, and any other survey related variable that you feel should be accounted for prior to analysis. The regression adjustment is a ratio, rather than the usual linear adjustment. This method guarantees that all adjusted intake values are positive. Usually, the first interview day is the standard, in the sense that observed means for later days are adjusted to be equal to the mean of the first day. Proper reorganization of your data set will allow any other day to be used as the standard, should you so desire. The grand mean can also be used as the standard.

The regression based ratio adjustment described above creates a set of dietary intake data with equal means across survey days. To make the variances across days equal as well, a second adjustment that scales the data such that the variance on each sample day is equal to that of the first sample day (or to a pooled variance if the grand mean is the reference standard) is applied.

The ratio-adjusted, variance-adjusted intake data have common mean and variance across survey days. After applying the inverse power transformation to the adjusted data, we create an "equal weights" sample, as described in Nusser, Carriquiry, Dodd, and Fuller (1996). The data set is now ready for the estimation of the usual intake distribution.

2.   Semiparametric transformation to normality:

Observed intake data (whether adjusted or not) generally have nonnormal distributions. For nutrients such as vitamins and some micro nutrients, skewness is quite extreme.

Most statistical procedures rely on the assumption of normality. For example, the method that was proposed by the National Academy of Sciences (NRC, 1986) for estimating usual intake distributions, assumes that intake data are normal, and proposes that dietary intake data be transformed prior to analysis so that the normality assumption is satisfied. The C-SIDE procedure transforms adjusted dietary intake data into normality as a part of obtaining estimates of usual intake distributions.

The transformation into normality in C-SIDE is done in two steps. In the first step, data are transformed so that their distribution is as close to normal as possible, by using a power transformation similar to the one described earlier. However, power transformed data are not necessarily normal. Thus, a second transformation which takes the power transformed intakes into the normal scale is employed. This second step in the transformation is nonparametric and is based on a grafted polynomial model. The power transformation plus the grafted polynomial function make up the semiparametric transformation into normality.

3.   Estimation of within and between individual variances in intakes:

C-SIDE uses a measurement error model for observed daily intakes, similar to the model proposed by the NAS (1986). The model states that the observed intake for an individual on any day is equal to that individual's usual intake plus a measurement error. The variance of the usual intakes is the between individual variance. The variance of the measurement errors is the within individual variance, and reflects the day-to-day variation in intakes for an individual.

Estimates for both the within and the between individual variances are obtained under the measurement error model, under the assumption of normality. The variances are used to estimate the distribution of usual intakes in the normal scale.

4.   Back-transformation into the original scale:

The final step in the methodology is to transform the estimated usual intake distribution from the normal scale into the original scale. More than "undoing" the transformation into normality by applying the inverse transformation is required because the original transformation is nonlinear. The inverse transformation makes an adjustment for the fact that the mean of a nonlinear function of a random variable is not the nonlinear function of the mean. The inverse transformation is based on an approximation to the mean function. This back transformation is

called mean transformation, since it brings the distribution of usual intakes (true individual means) back into the original scale.

The program estimates statistics of interest from the estimated usual intake distribution. For example, estimates for the mean and the variance of the usual intake distribution for a nutrient, for a set of percentiles, or for the proportion of the population below (or above) a given threshold are available.

Foods

The food methodology involves three main steps. The first is to obtain an estimate of the distribution of food consumption probabilities (the probability that an individual consumes the food on any given day).

A modified minimum chi-square estimator is used to obtain an estimated probability mass for 51 equally-spaced consumption probability values (0.00, 0.02, 0.04, ..., 0.98, 1.00). A maximum entropy term in the objective function is used to smooth the mass over the positive consumption probabilities. The estimated probability that an individual's consumption probability is zero is an estimate of the fraction of the population that never consumes the food.

In the second step, the method for estimating usual nutrient intake distributions is applied to the positive intakes to estimate a usual intake distribution for consumption days only. This distribution applies only to days on which the food was consumed.

The third step is to estimate the usual intake distribution for all days from the joint distribution of consumption day usual intakes and individual consumption probabilities. The current method assumes that the usual intakes are independent of the frequency with which the food is consumed.

APPENDIX C:
ERROR MESSAGES

C-SIDE may display some of the following messages to either the diagnostic window or the output window, and in some cases to both. The first set of messages below are informational messages that allow you to determine which analyses are currently being performed. The messages beginning with NOTICE provide information that may or may not be of interest to you. The WARNING messages indicate a problem that could cause some of the results to be inaccurate, but will not cause the program to stop processing.

ERROR messages are displayed when further analysis is not possible. CFG ERROR messages deal with invalid specification of the problem to be analyzed. DATA ERROR messages result from bad or unexpected data in the input data file.

In the following descriptions, *%s* represents some text (such as a variable or file name), *%d* represents an integer number, and *%e*, *%f*, or *%g* represent a floating point number.

Processing variable *%s*, *%d* observations.
> This message indicates that processing has begun for the named variable. All further messages between this one and the "Done with %s." message will pertain to this run.

Done with *%s*.
> This message indicates that processing has been completed for the named variable.

Repeating analysis for *%d* replicate weights.
> This message indicates that processing for the reference weight on the current variable has completed, and that the process will now be repeated the indicated number of times using replicate weights. Nearly all program output will be suspended during processing of data using the replicate weights.

NOTICE: Ratio adjustment NOT done, no values to adjust by.
> One or more ratio adjustment variables were specified, but the values of these variables did vary over the observations being analyzed.

NOTICE: Number of observations in first segment = *%d*.
NOTICE: Number of observations in last segment = *%d*.
> The given value for "Observations in end segments" would have resulted in less than 2 distinct observations in either the first or the last segment in the grafted polynomial fit. Additional observations were added to the indicated segment.

NOTICE: Cannot open file selection dialog. Only one
  file selection dialog may be active at a time.
> Only one file selection window, no matter where in the C-SIDE interface it is called from, may be on the screen at one time.

WARNING: Configuration line not recognized: "*%s* ..."
> A line was found in the configuration file that did not begin with either a keyword or a pound sign (#). The line will be ignored.

WARNING: Could not save output file *%s*
> The program could not perform the "Save output window" request. This is likely due to restrictive file or directory permissions.

WARNING: Could not save configuration file *%s*
> The program could not perform the "Save" operation. This is likely due to restrictive file or directory permissions.

WARNING: Bad class list. ord *%d*, lineno *%d*
> An internal error occurred while reading class variable values.

WARNING: householder pivot *%.1e* for n=*%d*, row=*%d*
> The pivot value obtained while performing a Householder transformation as part of the generalized inverse routine was less than 1e-15. Very small pivot values could cause underflow problems on some computers.

WARNING: Normality test hypothesis rejected for weight *%d*.
  Model with *%d* join points produced an ADS of *%.3f*.

C-SIDE was unable to fit a grafted polynomial to the data during construction of the daily intake transformation using a replicate weight. None of the attempts produced an Anderson-Darling test statistic that was within the specified critical value. The best available fit WILL be used and processing will continue.

WARNING: Negative derivatives detected in the grafted
    polynomial fit used to obtain food density values.
    C-SIDE was unable to fit a grafted polynomial to the data during construction of a transformation that carries unconditional usual intakes for consumers to standard normality. All of the fitting attempts resulted in negative derivatives. Density function data for the food distributions will not be produced.

WARNING: Normality test hypothesis rejected in the grafted
    polynomial fit used to obtain food density values.
    Model with *%d* join points produced an A-D test
    statistic of *%.3f*
    C-SIDE was unable to fit a grafted polynomial to the data during construction of a transformation that carries unconditional usual intakes for consumers to standard normality. None of the attempts produced an Anderson-Darling test statistic that was within the specified critical value. Density function data for the food distributions will not be produced.

WARNING: doest4(), zero determinant.
    An error occurred while trying to test for a linear relationship between individual mean and standard deviation. The results shown in the output for this test should be disregarded.

WARNING: Cannot estimate 4th moment of measurement error.
    There are not enough replicate observations in the data set to estimate this parameter. At least four individuals must have more than one observation in order to obtain this parameter. Without this parameter, C-SIDE may not be able to accurately correct for heterogeneous within-individual variances.

WARNING: Usolve(), matrix is singular.
WARNING: Lsolve(), matrix is singular.
    A singular matrix was encountered during an attempt to find the inverse of a matrix.

CFG ERROR: No input file name.
    Please complete the "Define file format" window.
    An input data file name has not been specified.

CFG ERROR: No analysis to perform.
    Please complete the "Assign variables" window.
    At least one nutrient or food variable run must be specified.

CFG ERROR: Could not open *%s*
>   C-SIDE could not open the input data file. Verify that the file
>   name is spelled correctly and that the file and directory
>   permissions allow the file to be read.

CFG ERROR: Unequal number of days not allowed for foods.
>   C-SIDE foods analysis requires exactly the same number of
>   intakes for each and every individual in the data set.

CFG ERROR: Cannot run replicate weights with foods data.
>   You have assigned a replicate weight set to a food variable.
>   C-SIDE will only perform a jackknife or balanced repeated
>   replication analysis on nutrient data.

CFG ERROR: Two key id variables found, "*%s*" and "*%s*"
>   More than one Individual ID variable was specified on the
>   "Define variables" window. C-SIDE currently can accept only
>   one Individual ID in the data file.

CFG ERROR: Subpopulation variable "*%s*" ignored.
>   A variable was assigned as both a subpopulation variable and as a
>   ratio adjustment variable. The variable will be used for ratio
>   adjustment only.

CFG ERROR: Sample day variable "*%s*" not valid.
>   The defined sample day variable was not a "Class, obs" type
>   variable.

CFG ERROR: Sample day variable "*%s*" is invalid.
>   The defined sample day variable was not found in the list of
>   variable names.

CFG ERROR: Subpopulation variable "*%s*" is invalid.
>   The defined subpopulation variable was not found in the list of
>   variable names.

DATA ERROR: Ran out of records after skipping *%d*.
>   You specified a value for "Initial number of lines to skip" on the
>   "Define file format" window that was larger than the number of
>   lines in the file.

DATA ERROR: Negative observation in line *%d*.
>   A negative observation was detected on the indicated line of the
>   input data file. Negative observations are not allowed.

DATA ERROR: Missing weight value in line *%d*.
DATA ERROR: Missing class variable in line *%d*.
DATA ERROR: Missing continuous variable in line *%d*.
DATA ERROR: Missing sample day variable in line *%d*.

A variable of the indicated type contained a missing value. These variables may not contain missing values. The error occurred on the indicated line of the input data file.

DATA ERROR: Too few fields, record *%d*.

The indicated line of a delimited input data file did not contain enough data variables. Either additional delimiting characters will need to be added to that line of the data file or some of the information on the "Define file format" window will need to be changed.

DATA ERROR: Invalid type (!double), record *%d*.

A variable on the indicated line of the input data file did not contain a real number value as expected.

DATA ERROR: Invalid type (!long), record *%d*.

A variable on the indicated line of the input data file did not contain an integer value as expected.

DATA ERROR: Sample day value out of range, record *%d*, value *%d*.

Values for a sample day variable must be {1,...,N} where N is the maximum number of observations for an individual in the data set. The offending value and the line of the input data file that it occurred on are reported.

ERROR: Negative derivatives in grafted polynomial fit.

C-SIDE was unable to fit a grafted polynomial to the data during construction of the daily intake transformation. All the fitting attempts, each using a different number of join points, resulted in negative derivatives.

ERROR: Normality test hypothesis rejected. Model with *%d* join points produced an Anderson-Darling test statistic of *%.3f*, which was still greater than the *%g*% critical value of *%.3f*.

C-SIDE was unable to fit a grafted polynomial to the data during construction of the daily intake transformation. None of the attempts produced an Anderson-Darling test statistic that was within the specified critical value. The best A-D test statistic obtained is shown.

ERROR: Negative derivatives detected in grafted polynomial
    fit in usual intake transformation.
    C-SIDE was unable to fit a grafted polynomial to the data during
    construction of the usual intake transformation. All the fitting
    attempts, each using a different number of join points, resulted in
    negative derivatives.

ERROR: Normality test hypothesis rejected
    in usual intake transformation.
    C-SIDE was unable to fit a grafted polynomial to the data during
    construction of the usual intake transformation. None of the
    attempts, each using a different number of join points, resulted in
    an Anderson-Darling test statistic that was within the specified
    critical value.

ERROR: Normality test hypothesis rejected in usual intake
    transformation. Model with *%d* join points produced
    an A-D test statistic of *%.3f*, which was still
    greater than the *%g*% critical value of *%.3f*.
    C-SIDE was unable to fit a grafted polynomial to the data during
    construction of the usual intake transformation. None of the
    attempts produced an Anderson-Darling test statistic that was
    within the specified critical value. The model with the best A-D
    test statistic is shown.

ERROR: Negative estimate for usual intake variance.
    Do to the unusual structure of the data, the procedure used to
    estimate the usual intake variance was not successful. This could
    be caused by data that has many individuals with only a single
    intake and a small number of individuals with multiple intakes.

ERROR: Signal *%d* with errno *%d* caught (context *%x*).
    <Additional information on the error may follow.>
    A floating point error that would have normally resulted in
    program termination and possibly a core dump has occurred. If
    running C-SIDE interactively, it is recommended that you quit the
    program (after possibly saving your configuration file and output
    windows) and restart before executing another run.

REFERENCES

Carriquiry, A.L. W.A. Fuller, J.J. Goyeneche, and H.H. Jensen (1995), Estimated Correlations Among Days for the Combined 1989-91 CSFII, Dietary Assessment Research Series 4. CARD Staff Report 95-SR77, Center for Agricultural and Rural Development, Iowa State University, Ames, Iowa.

Dodd, K.W. (1996), *A Technical Guide to C-SIDE*. Department of Statistics, Iowa State University, Ames, Iowa. Dietary Assessment Research Series 9. CARD Technical Report 96-TR32, Center for Agricultural and Rural Development, Iowa State University, Ames, Iowa.

National Research Council (1986), *Nutrient Adequacy, Assessment Using Food Consumption Surveys*. Washington, D.C.: National Academy Press.

Nusser, S. M., A.L. Carriquiry, K.W. Dodd, and W.A. Fuller (1996), "A semiparametric transformation approach to estimating usual daily intake distributions," *Journal of the American Statistical Association* (in press).

Nusser, S.M., W.A. Fuller, and P.G. Guenther (1996), "Estimating usual dietary intake distributions: Adjusting for measurement error and nonnormality in 24-hour food intake data," in L. Lyberg, P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz, and D. Trewin. (eds.) *Survey Measurement and Process Quality*. New York: Wiley (in press).

Stephens, M. A. (1974), "EDF statistics for goodness of fit and some comparisons," *Journal of the American Statistical Association* 69, 730-737.

Wolter, K.M. (1985), *Introduction to Variance Estimation*, New York: Springer-Verlag.