DataSwap User's Guide

Version 3.4*

December 2020

Prepared for:

National Center for
Education Statistics

Prepared by:

WESTAT
Rockville, Maryland

* Note: This user's guide is for *DataSwap* Macro v3.3.2.

TABLE OF CONTENTS

Section Page

CONTENTS CONTINUED

CONTENTS CONTINUED

Section Page

LIST OF APPENDIXES

Appendixes Page

LIST OF TABLES

Tables Page

LIST OF EXHIBITS

Exhibits Page

1. INTRODUCTION

1.1 Purpose of *DataSwap*

The question of whether the release of statistical data for public use may lead to the disclosure of the identity of individual units is a long-standing concern. Federal and state agencies are struggling with the need to release study data while protecting the confidentiality of the individuals or institutions included in these data. There are several laws to ensure that information provided by individuals is kept private. These include the Privacy Act of 1974,[1] the Education Sciences Reform Act of 2002, the USA Patriot Act of 2001, and the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA).[2] Failure to protect the confidentiality of individuals in accordance with these laws can result in a fine and/or a prison term, as well as doing irreparable damage to the study and reputation of the Federal agency.

The goal for the Federal agency is to ensure confidentiality while limiting data perturbations in a manner that would minimize the impact on data quality. Some Federal agencies provide guidelines and standardized procedures for ensuring data confidentiality and quality. The National Center for Education Statistics (NCES) as part of the Institute for Education Sciences (IES) has implemented confidentiality standards for all NCES data releases.[3]

The IES confidentiality standards generally require two separate procedures for public-use microdata files: (1) identify any sensitive variables and records through external matching and mask the variables and records that match using directed data swapping, and (2) introduce an additional measure of uncertainty into the data using random swapping. The *DataSwap* software was developed and enhanced in an effort to ease the implementation of both procedures. When public release of the data is only available through a Data Analysis System (DAS), the first procedure is no longer required, though the second procedure (random data swap) is still required for fulfillment of IES confidentiality standards.

The basic idea of data swapping is to protect a database by interchanging, or "swapping" values of one or more variables between records. The benefit of using swapping as a statistical perturbation technique is that it maintains the unweighted univariate distribution of each variable while still introducing uncertainty about the identity of records. A data intruder does not know which variables or records contain

[1] Section 552a protects records maintained on individuals: http://www.justice.gov/opcl/privstat.htm.

[2] Title V of the E-Government Act of 2002, Public Law 107-347 (December 17, 2002). These standards affect all Executive Branch agencies: http://www.eia.doe.gov/oss/CIPSEA.pdf

[3] For the detailed standards, see http://nces.ed.gov/statprog/2002/std4_2.asp.

swapped information, so randomized swapping helps, to make reidentification of a record uncertain for the intruder.

1.2 Software

IES uses a controlled random swapping approach on all restricted-use and public-use microdata files. Former NCES staff developed the methodology for data swapping and designed a series of SAS macros automating this methodology. The underlying procedures are described by Kaufman, Seastrom, and Roey (2005). The software, now collectively packaged in a software application called *DataSwap*, is the only IES Disclosure Review Board (DRB) approved swapping software package. It has been reviewed and further commended by the NCES/NISS Data Confidentiality Task Force as an appropriate data protection software tool.[4]

Having one standardized software package for data swapping is advantageous for several reasons. Since the DRB must approve not only the approach used for data confidentiality but the software as well, their review of disclosure analyses is expedient once they are familiar with *DataSwap* and its capabilities. Familiarity with one software package for both IES staff and contractors facilitates the understanding of the process and the results. Staff, time, and review effort required for ad-hoc swapping software would be detrimental to the timeliness and cost of the study.

DataSwap software enables testing different swapping scenarios to facilitate review and revision of the process, and ultimately assists in limiting the impact on data quality. The capability of linking variables for swapping and analyzing the impact on outcome variables provides assurance that data consistency is maintained.

Having standardized software for data swapping allows for consistency in swapping methodology across studies and provides systematic swapping outputs. The parameter input sheet option makes the software easy to use and share with other contractors.

1.3 Methods

DataSwap employs a controlled random swapping approach by selecting records for swapping using different random sampling approaches (simple random sampling, stratified, probability proportionate to size). It selects swapping partners based on a distance (or bias) measure and can be used on data with

[4] NCES/NISS Data Confidentiality Task Force final report dated February 5, 2008.

complex samples. IES standard practice is to process the swapping five to seven times and review the global utility measures to recommend one of the runs.

For the controlled random swapping approach, “controlled” means two things: First, it means that the user is responsible for identifying the data swapping variables and parameters. The user/data analyst should be familiar with which data would be the most identifiable and sensitive. The user should also understand the content of the data file as well as the purpose and focus of the study so s/he can therefore guide the data swapping procedures accordingly.

The second meaning of “controlled” is that, in *DataSwap*, once the target records are selected, the file is partitioned into swapping cells. The swapping methodology is designed to find a swapping partner that limits data distortion. The methodology includes the use of swapping cells to identify swapping partners in adjacent cells with similar (or identical) weights and close (with at least one or some different) variable values. The pair with the smallest swapping bias is selected as the swapping partner. For more details and illustrations of swapping cells and donor selection, please see Section 4.

Before implementing *DataSwap*, the user must consider several facets of the swapping process, including how the targets are to be selected, if particular records will be targeted for swapping, how the swapping cells are to be formed, and which variables will be swapped. Output reports help the user evaluate the swapping impact, which can also be controlled through the software parameters.

1.3.1 Targets

The first step in the software is the selection of the records whose values will be swapped (i.e., target records). Target records are selected systematically with probabilities proportionate to a measure of size. The user may specify a stratified design and/or use variables to sort the data before selection. The sampling rate is predetermined by the user and entered as a parameter. Note that cases with a high risk of disclosure may be given a higher selection probability.

1.3.2 Forming Cells

The second step in the algorithm is the selection of swapping partners (i.e., donors). The user must select variables to define the swapping cells. The swapping cells are formed by cross-classifying key categorical variables (i.e., identifiers such as age and education attainment categories). These variables are also the swapping variables – those variables whose values will be swapped. The search for swapping partners proceeds as follows. Consider a selected target record in a given cell. Two potential swapping partners for the target record are initially selected, one from each neighboring (adjacent) cell. That is, within each neighboring cell, the record with the closest sampling weight to the target record is selected as a potential swapping partner. The search process continues by comparing the swapping bias and the potential swapping partner. The record that results in the smallest swapping bias is chosen as the swapping partner. An example follows, and more details are provided in Section 3.

For example, if a user wants to swap values of occupation among cases with the same age, sex, and race, then the swapping cells will be formed by the cross-classification of those three variables with occupation as the last (right-most) variable. For most cases, a swapping partner will be chosen such that the values of the first three variables are as similar as possible, but with values of occupation that are different (to ensure that swapping takes place) and with a minimum calculated swapping bias. It is also possible that a record resulting in the lowest bias is one that has the same value of occupation, but differing levels of one or more of the other variables.

As an alternative, the user may vary the order in which the variables are cross-classified to form the swapping cells across all the cases. When specified, the records in the input file are randomly allocated to groups such that each swapping variable is used as the right-most variable an equal number of times. In the example above, swapping cells using this alternative method will still be formed as a cross-classification of the three variables; however, each of the three variables will be the right-most variable in the cross-classification an equal number of times (with the other variables ordered in a random fashion). Using this method will result in a more balanced distribution of swapping across the swapping variables since more race and sex values will be swapped rather than having values of occupation swapped most often as in the former example.

The former (original) method tends to cause disproportionate swaps to the right-most variable. This is useful when the file contains only one or two highly identifying variables. The alternative (balanced) method provides a more balanced distribution of swapping among the variables. If the variables used for swapping are all important and useful for analysis, it is preferable to lessen the impact of change on an individual variable. The balanced method provides data changes that are equitably distributed across the set of swapping variables on the data file so that no individual variable is adversely affected by the swapping.

If there are key variables that absolutely should not change value, but are critical in controlling the swapping partner search, then one or more variables may be specified as hard boundary variables. If hard boundaries are specified, only potential swapping partners within the same hard boundaries will be considered. If a swapping partner cannot be found within the hard boundaries, the algorithm will not proceed, and some adjustment to the hard boundary variables will be required. There may only be one neighboring cell if the target's cell comes at the beginning or end of the sort order within a hard boundary, or at the beginning or end of the file if no hard boundary is specified.

After the selection of the swapping partners, a check is made to determine that a swapping partner is used only once. If a partner is used more than once, then the partner is assigned to the target resulting in the smallest absolute bias. Ties are handled with a random selection. Next, a check is made to see if all target cases have final swapping partners. If any are found without final swapping partners, the partner search is repeated until all swaps have unique final swapping partners.

1.3.3 Swapping Values

In the third step, the data are swapped. By definition, the variables that are allowed to be swapped are referred to as "swapping variables." Therefore, hard boundary variables are not swapping variables. The swapping of data occurs as the values of the swapping variables are switched between each target and their respective partners (i.e., the values of the target's variables identified for swapping are assigned to the respective partner and the partner's values are assigned to the respective target case). The user also has the option of specifying other variables to swap, apart from those used to form the swapping partner cells. These variables can be linked to the swapping variables so that, as the value of a particular swapping variable changes, the linked variable(s) will also be changed. For example, if a user wants to use age category as a swapping variable, the detailed age variable should be specified as a linked variable so that the two variables remain consistent on the final file.

1.3.4 Reports

The last step is the generation of output reports for the user, the DRB chair, and the DRB members. The *DataSwap* software produces output that can be used to document and evaluate the swapping results. The output compares unweighted and weighted frequencies, means, correlations, and regression models before and after swapping. The output for the DRB can be pasted into a memorandum for the DRB.

The original swapping algorithm written by Steve Kaufman is preserved in this version of the software. Westat has been responsible for making the software more user-friendly and for enhancing its features. Section 2 describes the parameters and syntax and provide some user guidelines for the *DataSwap* application, respectively. Section 3 provides further technical descriptions. Section 4 includes examples using these data.

2. DATA SWAPPING

This section describes the user specifications, provides some notes on processing, and gives an overview of the output reports from *DataSwap*. A description of the *DataSwap* user specification is provided in Section 2.1. Some processing notes and helpful hints are given and a list of error messages is provided in Section 2.2. Section 2.3 provides an overview of the various output reports intended for users of the software, the Disclosure Review Board (DRB) chair, and its members.

2.1 Specifications

User specifications for the SAS macro can be submitted on a parameter sheet (see Exhibit 2-1). Once the parameters are entered into the program, then the program is invoked through a SAS macro call. The call is detailed in Section 2.1.1. Each parameter, including controllers for input, sampling, swapping data, and output, is described in Section 2.1.2. An example of a SAS macro call for *DataSwap* is provided in Section 2.1.3.

2.1.1 Syntax

DataSwap is invoked by using a SAS macro call, as follows:

```
%DataSwap ([macro parameter = value],[macro parameter = value], . . .)
```

For PC/VAX/LINUX users, the following statements must precede the macro call:

```
%INCLUDE ‘file name containing the macro statements’;
```

Data Swapping

3.2

Parameter specification form

Parameter		*	Entry	Default	Description
Input	**DATA=**	**R**			input file
Controllers	**SEED=**	**O**		**0**	random seed
	ID=	**R**			case identification - a single variable
	VARSTRAT=	**O**			Variance Stratum
	VARUNIT=	**O**			Variance Unit
Sampling	**RATE=**	**R**			swapping rate - a single variable or a number
Controllers	**MOS=**	**O**		**1**	measure of size - a single variable
	STRATUM=	**O**		**1**	stratum - a single variable
	SORTVARS=	**O**		**BOUNDARY|| SWAPVARS**[1]	sort order for non-certainty selection - a list of variables
	SWAPMETH=	**O**		**2**	swapping method – 1:orginal 2:balanced
	SWAPVARS=	**R**			primary swap variables – a list of variables
	SWAPVARS_T=	**O**		**O for all**	SWAPVARS variable type: Ordinal (O) Nominal (N)
	SWAPVARS_MD	**O**			alternate values to be treated in output statistics as missing in the SWAPVARS variables
	BOUNDARY=	**O**			hard boundary - a list of variables
Swap	**BOUNDARY_T=**	**O**		**O for all**	BOUNDARY variable type: Ordinal (O) Nominal (N)
	BOUNDARY_MD	**O**			alternate values to be treated in output statistics as missing in the BOUNDARY variables
Controllers	**BIASVAR=**	**O**		**the right- most var in SWAPVARS**	variable for calculating bias - a single variable (must be numeric)
	WGT =	**R**			case survey weight - a single variable
	LINKSWAP=	**O**			linked swap variables - a list of variables separated by #
	IMPUTE=	**O**		**y**	impute if swapvars have missing values.
	MISSINGDEF=	**O**		**‘ ‘ or .**	values to be defined in selection of target records as missing – a list of values separated by #
	OUT=	**R**			dataset name for swapped result
	KEYOUT=	**O**			key outcome continuous variable for weighted means and correlations (must be numeric)
	KEYOUT_MD	**O**			alternate values to be treated in output statistics as missing in the KEYOUT variables
Output	**KEYVARS=**	**O**			key variables for output tables and correlations
Controllers	**KEYVARS_T=**	**O**		**O for all**	KEYVARS variable type: Ordinal (O) Nominal (N)
	KEYVARS_MD	**O**			alternate values to be treated in output statistics as missing in the KEYVARS variables
	MODELS=	**O**			a list of models separated by # exp. X:Y Z#W : V1 V2

MODELCLASS=	**O**	______________________		a list of class variables separated by # exp. Y Z#V1
TOLFLAG=	**O**	______________________	**0.10#45#1.96#1.1**	tolerance measure for flagging outliers
MAXCAT=	**O**	______________________	**20**	Maximum # of categories for DRB table variables
LISTPAIR=	**O**	______________________	**S#1.0**	Subset listing for swapping pairs

* O: Optional   R: Required

[1] The double vertical bar denotes the concatenation of the list of BOUNDARY and SWAPVARS variables. Version 3.4, December 2015

Exhibit 2-1 - *DataSwap* Macro Parameter Specification Form

2.1.2 Parameter Description

The *DataSwap* macro uses the following parameters. Most parameters can be either numeric or character variables. However, some parameters are restricted to numeric values only, as described below.

2.1.2.1 Input Controllers

DATA = *SAS dataset* specifies the SAS dataset name of the file used in *DataSwap*. This parameter is required. The user may specify a one- or two-level name (i.e., a temporary or permanent SAS dataset). The dataset must be SAS version 7 or higher (i.e., have a sasb7dat extension) and may contain up to 255 variables.

SEED = *nonnegative integer value or string of values* (separated by spaces) provides input to the RANUNI function in SAS to create one or more random numbers from a univariate distribution. This parameter is optional and each value must be less than 2^{31}-1 (or 2,147,483,647), with no more than seven seeds allowed. This random number is then used to determine the random starting point for target record selection. If only one seed is specified, then the run is treated as a test run and all program and output file names will be given the suffix "_Test" before the file extension. If more than one seed is specified (SEED1 SEED2 … SEED#), all program and output file names will contain the suffix "_Run1," "Run2,"…,"_Run#" respectively.

Specifying the same seed as that shown in the output from a prior run ensures the same set of target records are sampled and that the same set of swapping partners are selected, assuming that all other parameters affecting sample selection and swapping partner selection remain the same.

The default value is 0 in which case the macro randomly selects one seed based on the computer clock. In the event that the default value is used, the resulting seed is printed on the information page so the user may replicate the results generated by a run with the default seed value.

ID = *variable* specifies one variable to uniquely identify each record. This parameter is required.

VARSTRAT = variable specifies the variance stratum to be used to calculate standard errors using Taylor Series Linearization. This parameter is optional; if missing, no standard errors will be calculated.

VARUNIT = *variable* specifies the variance unit to be used to calculate standard errors using Taylor Series Linearization. This parameter is optional; if missing, no standard errors will be calculated.

2.1.2.2 Sampling Controllers

RATE = *numeric variable or value* specifies the desired sampling rate within a stratum for selecting target records. This is the swapping rate. If there is only one stratum or if the rate is the same for every stratum, then the value of the rate can be given directly in the macro call rather than creating a variable on the input dataset. A probability proportionate to size (PPS) sample is selected within strata, or overall, if there are no strata. When a rate is specified, the target record sample size overall, or in the stratum, is expected to be *n* = RATE * (number of records in the file/stratum). The total measure of size (MOS) is computed (TOTMOS) and then the sampling interval of TOTMOS / *n* is calculated—for each stratum if appropriate. The rate must be positive and less than or equal to one. Note that the perturbation rate is equal to 2 * RATE.

MOS = *numeric variable or value,* specifies the measure of size pertaining to a PPS sample selection approach. If entered as a value, the only valid entry is 1. If entered as a variable, the values can be anything greater than 0. This parameter is optional. The default is set equal to 1, which results in an equal probability sample.

STRATUM = *variable* identifies the strata for the sampling of the target records. This parameter is optional. The default is set equal to 1, prompting the macro to create a stratum variable equal to 1 for all records.

SORTVARS = *list of variables*[5] specifies the sort order for the noncertainty selection of target records. This parameter is optional. The default is the concatenation of the list of BOUNDARY and SWAPVARS variables, denoted by BOUNDARY||SWAPVARS, which defines the variable sort order within each stratum.

2.1.2.3 Swap Controllers

SWAPMETH = *1 or 2* indicates whether the swapping is done primarily on the right-most variable in the SWAPVARS specification (*1*) or in a more balanced way across all SWAPVARS (*2*). If

[5] Note that each variable must be specifically listed and delimited by spaces. Shortcuts such as VAR1-VAR5 or VAR* may not be used.

SWAPMETH = *1*, the SWAPVAR variables remain in the order specified for all targeted cases. This parameter is optional. The default is SWAPMETH = *2.*

If SWAPMETH = *2*, random groups of records are formed. The number of random groups is equal to the number of SWAPVARS variables. Within each random group, a random ordering of SWAPVARS variables is assigned such that each SWAPVARS variable occurs in the right-most position once across the random groups. Under this approach, BIASVAR is dynamic (if specified, it is ignored) in that it changes as it is set equal to the right-most SWAPVARS variable for each random group.

For example, If SWAPMETH = *2* and SWAPVARS = A B C, then one-third of the swaps is performed with the order of SWAPVARS being A B C or B A C (C is the BIASVAR, and the order of the other two variables is randomly determined); another third is performed with SWAPVARS ordered as A C B or C A B; the remaining third is performed with SWAPVARS ordered as C B A or B C A. Because of the random ordering, there is no assurance that the same SWAPVARS variable is in each position once; the three groups in this example may be A B C, C B A, and A C B with B being the second variable in two groups, and A being the first variable in two groups. Hard boundary variables may be used with either method, but their order is not randomized in SWAPMETH = *2*.

SWAPVARS = *list of categorical variables* are the variables to be swapped. This parameter is required. The swapping partner cells are created as the concatenation of the list of BOUNDARY and SWAPVARS variables (denoted by BOUNDARY||SWAPVARS). Only the variables in SWAPVARS are swapped. Since these are used to define the swapping cells, these variables should contain categories rather than detailed or continuous values. Variables with the greatest likelihood of change are toward the right-hand side of the ordered list of variables. The donor (i.e., swapping partner) is selected from the adjacent cell. These variables are used to evaluate the swapping results through measures of association and also used as table domains for computing weighted percents/means before and after swapping. If no BIASVAR is specified, the right-most variable in the SWAPVARS list is used as the BIASVAR (see description below). In this case, the right-most variable must contain numeric values only. If SWAPMETH = *2*, all SWAPVARS must be numeric, and contain more than one level, and have no missing values (or imputed by specifying the IMPUTE parameter). No more than 20 swapping variables may be specified at one time.

SWAPVARS_T = *O or N O or N*...indicates whether each swapping variable is ordinal or nominal. This parameter is required. This parameter takes n entries of *O* or *N*, each separated by a blank space, for each of the n variables listed in SWAPVARS. If a SWAPVARS variable is nominal (as specified by SWAPVARS_T=*N*) and has more than two levels, indicator variables will be created for each level of the SWAPVARS variable to be used in correlations. For example, two indicator variables would be created for the variable GENDER (with values 1 and 2): GENDER_1 will have the value of 1 when GENDER=1, and will be 0 otherwise; GENDER_2 will have the value of 1 when GENDER=2, and will be 0 otherwise. In the output, all nominal variables will be used as variables in the regression models. Note that if a variable has more than 16 levels, indicator variables will not be created. To increase the number of nominal levels to be converted, specify NL = <*new value*> in the macro call. The default value for each variable is *O* indicating that all swapping variables are ordinal.

SWAPVARS_MD = *a list of values separated by blanks* is a list of special values to be treated as missing for SWAPVARS (i.e., -8, -9, or other values that are not SAS missing values). This list of special missing values will be removed from the computation of measures of association and regressions when SWAPVARS is used. Separate each value by a blank. Do not use quotes even if the imputation variables are character. The macro will supply quotes if necessary. SAS missing values will automatically be included so they need not be included in this list.

BOUNDARY = *a list of categorical variables* defines hard boundaries for finding the swapping partners. Each variable must have more than one level. The swapping partner cells are created as the concatenation of BOUNDARY and the list of SWAPVARS variables. The swapping partner search is limited to within the boundaries defined by the concatenation of the BOUNDARY variables. This parameter is optional and the default is that there are no boundary variables.

BOUNDARY_T = *O or N O or N*...indicates whether each boundary variable is ordinal or nominal. See the description of SWAPVARS_T above for a description of it use. This parameter is required only if boundary variables are specified.

BOUNDARY_MD = *a list of values separated by blanks* is a list of special values to be treated as missing for BOUNDARY (i.e., -8, -9, or other values that are not SAS missing values). This list of special missing values will be removed from the computation of measures of association and regressions when BOUNDARY is used. Separate each value by a blank. Do not use quotes even if the imputation variables are character. The macro will supply quotes if necessary. SAS missing values will automatically be included so they need not be included in this list.

BIASVAR = *a numeric SWAPVARS variable* is used in computing the bias. The bias is used in selecting a swapping partner. Hence, the BIASVAR variable is important in selecting the swapping partner for each target case. The BIASVAR variable must be one of the SWAPVARS variables. This parameter is optional and the default BIASVAR is the right-most variable in SWAPVARS. If a BIASVAR that is not a SWAPVARS variable is specified, the program aborts. The BIASVAR variable must contain numeric values only. Under SWAPMETH=*2* (balanced swap approach), the BIASVAR is dynamic; since the cell definition changes within the file, it consequently changes the right-most SWAPVARS variable, which becomes the BIASVAR variable. Therefore, under SWAPMETH=*2*, all SWAPVARS variables need to be non-missing (or imputed by specifying the IMPUTE parameter) and all SWAPVARS variables need to contain numeric values (see the description of SWAPMETH=*2* above). If BIASVAR is specified by the user when SWAPMETH=*2*, the BIASVAR specification is disregarded.

WGT = *a variable or a constant* is the full sample weight for analysis. This parameter is required. The weight is used in selecting the swapping partner. It is important to use a base weight or final weight since it is a key factor used in computing the bias. Bias is used in selecting the swapping partner for each target case to be swapped.

LINKSWAP = *list of variables# list of variables#* indicates variables to swap which are linked to those specified in the SWAPVARS parameter to maintain logical consistency between variables. This parameter is optional. This parameter takes *n* lists of variables separated by a "#"for each of the *n* variables listed in SWAPVARS. In the case of SWAPVARS=*Race Educ Income* and LINKSWAP=*Hisp Asian # EduHigh EduElem # Income1 Income2*, any time a value of Race changes, Hisp and Asian change as well; any time Educ changes, EduHigh and EduElem change; and any time Income changes, Income1 and Income2 change. If a variable in SWAPVARS has no linked variables, use the key word NULL (e.g., *NULL # NULL # Income1 Income2*). These variables are used to evaluate the swapping results as table domains for weighted percents/means before and after swapping. They will not be included in the measures of association or regression models unless also specified as KEYVARS variables.

IMPUTE = *Y or N* specifies whether or not to impute for at least one of the SWAPVARS variables when forming swapping partner cells. This parameter is optional. If IMPUTE = *Y* and there are no BOUNDARY variables, then each SWAPVARS variable is imputed using a random hotdeck method (random draws from its empirical probability density function). If there are BOUNDARY variables, then each SWAPVARS variable is imputed using a random within cells hotdeck method, where cells are defined as unique combinations of the BOUNDARY variables. If IMPUTE = *N*, missing values are always swapped with adjacent cells. For example, if variable A has the values 1,2,3,4,9 (with 9 indicating missing values), records selected as targets with values of 9 are almost always swapped with values of 4; this may not be an appealing solution, especially if those with missing codes of 9 are very different from those with values of 4. The imputed variables are dropped at the end of processing and only values of the original variables are swapped and retained. The default is IMPUTE = *Y*, but the user may turn off this option by specifying IMPUTE=*N*.

MISSINGDEF = *xvalue1 xvalue2 … xvaluen # yvalue1 yvalue2 … yvaluen # … # zvalue1 zvalue2 … zvaluen* indicates values of SWAPVARS to be imputed when IMPUTE=*Y*. This parameter is optional. The "#" entry provides the separator for each SWAPVARS variable, and the list is delimited by spaces. If MISSINGDEF is left blank, then the default is used, which is **.** for numeric variables and " " for character variables. Character values must be surrounded by quotation marks. If MISSINGDEF is not blank, and you want to impute for values of **.** or " ", they must be specified. SAS missing values such as .M, .D, etc, are interpreted as **.** by the macro, so specifying **.** as a MISSINGDEF results in any .M, .D, etc, values also being imputed. If you do not want to impute for a particular SWAPVARS variable, use the keyword NULL. If there are no missing values in the SWAPVARS data and IMPUTE = *Y* is specified (or used as default) the program sets MISSINGDEF to NULL for each SWAPVARS variable.

2.1.2.4 Output Controllers

OUT = *SAS dataset* specifies the SAS dataset file name containing the swapping results. This parameter is required. The file has the same content as the file specified in DATA parameter, except that some values within the SWAPVARS variables and LINKSWAP variables (if specified) are switched between pairs of records. The file can have a one- or two-level name (i.e., may refer to a temporary or permanent SAS dataset).

KEYOUT = *list of numeric variables* is used to assess the data quality of important survey outcomes before and after swapping. This parameter is optional. These analysis variables are generally continuous but can be categorical. The weighted means of the KEYOUT variables before and after swapping are presented in the output tables to help the user evaluate the swapping results. They are also included in the list of variables for which measures of association are computed and are used as dependent variables in regressions with all SWAPVARS as independent variables. An example of a KEYOUT variable is an assessment score. Any changes noted in the output tables for these variables will help evaluate the effect that swapping has on outcome measures. If not specified, the weighted means and regressions are omitted from the output reports.

KEYOUT_MD = *a list of values separated by blanks* is a list of special values to be treated as missing for KEYOUT (i.e., -8, -9, or other values that are not SAS missing values). This list of special missing values will be removed from the computation of measures of association and regressions when KEYOUT is used. Separate each value by a blank. Do not use quotes even if the imputation variables are character. The macro will supply quotes if necessary. SAS missing values will automatically be included so they need not be included in this list.

KEYVARS = *list of variables* denotes important key domains for assessing data quality before and after swapping. This parameter is optional. The KEYVARS variables are used to evaluate the swapping results through measures of association. For example, if gender is specified as a KEYVARS, then measures of association between the SWAPVARS and gender are produced to see if the relationship between gender and the swapped variables changed as a result of swapping values of the SWAPVARS. If any variables are specified in the LINKSWAP parameter, include them as KEYVARS if monitoring their swapping impact on measures of association and regression coefficients. The list of variables can include categorical and continuous variables. Any variables with "alpha" characters are recoded into a sequential numbering beginning with a value of 1 and continuing with 2, 3, ….

KEYVARS_T = *O or N O or N*...indicates whether each KEYVARS variable is ordinal or nominal. See the description of SWAPVARS_T above. This parameter is required only if KEYVARS variables are specified.

KEYVARS_MD = *a list of values separated by blanks* is a list of special values to be treated as missing for KEYVARS (i.e., -8, -9, or other values that are not SAS missing values). This list of special missing values will be removed from the computation of measures of association and regressions when KEYVARS is used. Separate each value by a blank. Do not use quotes even if the imputation variables are character. The macro will supply quotes if necessary. SAS missing values will automatically be included so they need not be included in this list.

MODELS = *X: Y Z # A: B C D #...* specifies main effects linear regression models to be fit for assessing multivariate data quality before and after swapping. This parameter is optional. Each model specification is separated by a "#", with the dependent variable listed first followed by a colon (:) and the list of independent variables. All variables used in the models must be included either in the KEYOUT, SWAPVARS, or KEYVARS parameters. The dependent variables must all be numeric. By default, regressions of each KEYOUT on all SWAPVARS are provided.

MODELCLASS = *Y # B D #*...indicates which variables in the MODELS parameter are considered class variables (nominal or ordinal) for the regression estimation. If specified, *m*-1 "dummy" variables indicating individual levels of the *m*-level class variables will be created and used in the regression analysis. This parameter is optional. By default the macro will construct this parameter internally based on the SWAPVARS_T and KEYVARS_T parameters. This is a reserved *DataSwap* parameter which may be expanded upon in future versions of the software. If specified, and inconsistent with the SWAPVARS_T or KEYVARS_T parameters, the macro will use the information provided in the SWAPVARS_T or KEYVARS_T parameters.

TOLFLAG = *constant_x # constant_y # constant_z # constant_v* specifies the tolerance level for the relative difference before and after swapping (*constant_x*), the minimum sample size (*constant_y*) for a table domain, the tolerance level on the number of standard errors that the after-swapping correlation or regression coefficient may deviate from the before-swapping values (*constant_z*), and the maximum ratio of before- and after-swapping *standard errors* (*constant_v*). The four constants are delimited by a "#".

If the absolute value of the relative difference is greater than the value of *constant_x*, and if the sample size is greater than *constant_y*, then an asterisk[6] is placed next to the value of the relative difference indicating that the difference between the original and swapped estimate (be it a weighted mean or percentage) for the variable is exceeds the user's tolerance level.

For correlations and regression coefficients, if the difference between the before- and after-swapping estimates is more than *constant_z* multiples of the before-swapping standard error and the sample size used in the estimate is greater than *constant_y*, then an asterisk is placed next to the value of the relative difference. These parameters are also considered in calculating the global data utility measures for tables. Please see the discussion in Section 3.3.3.3 for more information.

If VARSTRAT and VARUNIT are specified, ratios of before- and after-swapping standard errors are provided. If the ratio is more than *constant_v* and the sample size used in the estimate is greater than *constant_y*, then an "@" is placed next to the value of the ratio.

The default tolerance levels are *0.1 # 45 # 1.96 # 1.1*. The default sample size is 45 and is derived from a rough approximation of potential minimum cell sizes under the assumptions of a complex sample design for surveys.

MAXCAT = *integer* specifies the maximum number of categories of SWAPVARS and LINKSWAP variables to use as table domains (or "by" variables) in output tables designated for the DRB. This parameter is optional. If a SWAPVARS or LINKSWAP variable has more levels than the value of MAXCAT, then it is not included in tables showing weighted percents/means before and after swapping or in tables showing the before and after values of any imputed SWAPVAR variables. For user-only tables, variables in the SWAPVARS and LINKSWAP lists are excluded from the output reports if their number

[6] Note that the presence of an asterisk or "@" does not necessarily indicate statistical significance regardless of the values of these parameters. *DataSwap* is not designed to calculate the appropriate measures for significance testing.

of categories is greater than 300. This parameter is present to reduce the volume of output in the DRB tables and does not affect the swapping procedure. The default is 20.

LISTPAIR = *alpha # constant* where the value of *alpha* can be *S* or *B*, controls the number of swapping pairs (the selected target record and its partner record) are printed. The value of *constant* must be between 0 and 1, inclusive. This parameter is optional. If *alpha* = *S*, then the constant specifies the sampling rate. This requests a systematic dump of a subset of swapping pairs. If *constant* = *0*, no pairs are printed. If *constant*=*1*, then all pairs are printed. If *constant* = *0.10*, then one out of every 10 pairs is listed, starting with the first pair. This parameter is mainly used for reducing the volume of output. The default is S # 1.0 indicating that all pairs be printed.

If *alpha* = B, then any pair with an absolute value of relative bias greater than *constant* is printed. The absolute value of relative bias is computed as:

$$\left|\frac{(WGT1\times BIASVAR2+WGT2\times BIASVAR1)-(WGT1\times BIASVAR1+WGT2\times BIASVAR2)}{(WGT1\times BIASVAR1+WGT2\times BIASVAR2)}\right|$$

where,

WGT1 = the weight of the target record;

BIASVAR1 = the *BIASVAR* value for the target record; and,

WGT2 = the weight of the partner record.

GraphType = *PDF* or *RTF* specifies the format of the graphs produced by the program, either portable document format (*PDF*) or rich text file format (*RTF*). If the user specifies *PDF* the graphs may only be accessed by Adobe Reader; if the user specifies *RTF* the output may accessed by various programs including Microsoft Word.

USEDPI = *constant* specifies the dots per inch for portable document format (*PDF*) graphics files. The default is 600.

2.1.3 An Example Macro Call

The following is an example of a *DataSwap* macro call with example values for each parameter:

```
%DataSwap(
  DATA=lib.INFILE,
  RATE=0.1,
  SEED=9234 54 973 98 234,
  ID=ID,
  MOS=MOS,
  BOUNDARY=RACE,
  BOUNDARY_T=N,
  BOUNDARY_MD=7 8 9,
  SWAPVARS=EDUC INCOME,
  SWAPVARS_T=O  O,
  SWAPVARS_MD=7 8 9,
  BIASVAR=INCOME,
  WGT=BWGT,
  LINKSWAPS=EDUC1 EDUC2 # INCOME1 INCOME2,
  VARSTRAT=VARSTRAT,
  VARUNIT=VARUNIT,
  IMPUTE=Y,
  MISSINGDEF= 'RF' 'DK' # . 77 99,
  SWAPMETH=1,
  OUT=lib.OUTFILE
  KEYOUT=SCORE,
  KEYOUT_MD=7 8 9,
  KEYVARS=VAR1 VAR2 VAR3 VAR4,
  KEYVARS_T=N N N N,
  KEYVARS=7 8 9,
  MODELS=SCORE : DAGE3 VAR1 EDUC3,
  MODELCLASS=VAR1,
  TOLFLAG=0.1#45#1.96#1.1,
  LISTPAIR=S#1,
  MAXCAT=20,
  GraphType=PDF,
  USEDPI=600
  )
```

The parameter values in this program are interpreted as follows. Since five seeds are specified, the macro will run five times. With each run, the program selects 10 percent (i.e., RATE) of the records in the dataset as target records, through a PPS selection approach, where a variable named MOS contains the measure of size for each record. The software then finds a swapping partner for each target record using the variables BWGT and INCOME to calculate bias. The swapping cells are defined by RACE EDUC and INCOME. Before swapping, EDUC and INCOME are imputed. Values of EDUC equal to 'RF' and 'DK' are imputed using a random within cells hotdeck method within categories of RACE. Values of INCOME equal to **.**, 77, 99 (including any missing values indicated by .D, .M, etc.) are also imputed. The swapping partner is selected within RACE categories as defined by BOUNDARY. The software swaps values of INCOME and EDUC. If a value of INCOME changes, then the values of INCOME1 and INCOME2 change as well. If a value of EDUC changes, the values of EDUC1 and EDUC2 change.

Unweighted and weighted distributions of the levels of EDUC and INCOME will be produced as well as for the linked variables INCOME1, INCOME2, EDUC1, and EDUC2. If the weighted absolute value of the relative difference before and after swapping is more than 0.10 and the sample size for the level is more than 45, then an asterisk (*) will appear to the right of the after-swapping value. Additionally, before- and after-swapping standard errors will be produced. If the ratio of those estimated standard errors exceeds 1.10 and the sample size for the level is more than 45, an “@” will appear to the right of the ratio.

The key output variable is SCORE, other key variable specified are the variables VAR1 – VAR4. Weighted means and standard errors of SCORE will be produced by EDUC and INCOME as well as the linked variables INCOME1, INCOME2, EDUC1, and EDUC2. Large differences will be flagged as described above.

Additionally, all possible measures of association between EDUC and INCOME and SCORE and VAR1-VAR4 will be calculated. If the difference between the before- and after-swapping estimates is more than 1.96 multiples of the before-swapping standard error and the sample size used in the estimate is greater than 45, then an asterisk is placed next to the value of the relative difference.

In addition to the default regression model with SCORE as the dependent variable and EDUC and INCOME as the independent variables, a second model will be fit using SCORE as the dependent variable and DAGE3, VAR1, and EDUC3 as the independent variables. Any differences between the before- and after-swapping regression coefficients will be displayed and flagged with an asterisk as described above for the measures of association.

The .log and .lst files will be named *<file name>_Run1.log* and *<file name>_Run1.lst…<file name>_Run5.log* and *<file name>_Run5.lst;* the output datasets and charts will be named similarly. The utility measures will be summarized in *<file name>_summary.lst* and graphed in *<file name>_SummaryGraph.pdf*

2.2 Processing

Some system requirements and notes for using *DataSwap* are described in Section 2.2.1. Section 2.2.2 provides some helpful hints, and Section 2.2.3 includes a list of errors that may occur during processing.

2.2.1 System Requirements and Processing Notes

DataSwap is a SAS macro. SAS version 8 or 9 is required to run the macro. The following are some helpful programming notes for using *DataSwap*.

1. Parameters are separated by commas. For example, in %DATASWAP (BOUNDARY=A, SWAPVARS=B,…).

2. Parameters do not have to be in order. They can be specified as %DATASWAP(BOUNDARY=A,SWAPVARS=B) or
   %DATASWAP(SWAPVARS=B,BOUNDARY=A).

3. SAS shortcut techniques for variable lists do not work. For example, VAR1-VAR3 should be written as VAR1 VAR2 VAR3.

4. All intermediate SAS names created by the program start with a __ (two underscores).

5. The messages generated by *DataSwap* start with "*DataSwap* Error" so they can be distinguished from SAS messages.

6. TITLE3, TITLE4, TITLE5, FOOTNOTE1, FOOTNOTE2, and FOOTNOTE3 are reserved for use by *DataSwap*. Users should avoid these titles/footnotes since they will be overwritten.

7. When IMPUTE = *Y*, *DataSwap* creates new variables with the imputed values and names them <*imputed variable name*>_I. For this reason, it is strongly recommended that the user exclude or rename any existing variables in the input dataset that uses this suffix. Failure to do so may cause unexpected results.

2.2.2 Helpful Hints

There are certain special situations that arise for which *DataSwap* can be used, or for which pre-*DataSwap* processing must occur. Other approaches to data swapping within *DataSwap* can be found in Krenzke, et al. (2006). These approaches address reducing disclosure risk for high risk variables and domains, and for hierarchical data structures. Other recommendations in the paper address maintaining data consistency and reducing swapping impact.

2.2.2.1 Using IMPUTE=Y

If MISSINGDEF values are specified and they do not include SAS missing values, then the SAS missing values are not imputed. Also, NULL can be specified for any SWAPVARS variables for which imputation is not desired.

If the same two missing values of a swapping variable are imputed to two different values and then swapped with each other, the result is a missing value being swapped with the same missing value. In this event, the change flag for the swapping variable is not set.

The row with “selected” in the sampling flag column and “not swapped” in the swapping flag column indicates that the number of selected records without any changes in values. This could further be seen by reviewing the listing of swapping pairs.

2.2.2.2 Swapping Rate and Perturbation Rate

The swap rate is defined as the percentage of records that are selected for the swapping procedure and is specified by the user in the parameter RATE. This should not be confused with the perturbation rate. The perturbation rate is defined as the percentage of records with one or more values changed as the result of swapping. Thus, it should be approximately twice the swapping rate.

It may be that more than one value is swapped on one or more records. As a result, the sum of the perturbation percentages for each of the swapping variables will be greater than or equal to twice the value specified in the RATE parameter. In the event of a missing-to-missing swap described above, the actual perturbation rate will be less than twice the specified RATE.

2.2.2.3 Negative Values and Special Character Values

If SWAPVARS variables contain negative values, when the concatenation for defining the swapping cells occurs, a value of “-1” is not the same as a sort on positive numeric values. So a -1 is next to a 9 (if 9 is the maximum value), instead of next to a 0. The user can make a temporary recode of the variable and use it in the SWAPVARS parameter. If so, then put the original variable should be specified as a LINKSWAP variable.

It is important that all variables be in a form consistent with the parameters. For example, if a variable is specified as the swapping rate (RATE), it must be constant within STRATUM since differential target sampling rates may not be implemented within a STRATUM. The parameter BIASVAR must also be a numeric variable since it is used in calculations for finding swapping partners.

It is also beneficial to pre-process cross-tabulations, which use the BOUNDARY and SWAPVARS variables to depict the formation of the swapping partner cells. With BOUNDARY variables, cross-tabulations may identify cells where a swapping partner will not be found. Under this condition, collapsing of categories is recommended. The cross-tabulations will show any problematic cells where there could be undesirable swapping of certain values between records. For instance, swapping a "Refused" code of 9 with a 2 may be undesirable. Cross-tabulations may also reveal singletons. For this situation, the record with the singleton could be given a higher probability of swapping by increasing the MOS for this record.

2.2.2.4 Handling Skip Patterns

A designated SWAPVARS variable could be involved in a skip pattern. Suppose a questionnaire item (A) had two levels: 1=Yes, 2=No. People responding "yes" to A are asked an additional question (B); all others are asked another question (C). If values of B and C are to be swapped, then a new variable needs to be created to determine swapping partners. To do this, let item A be assigned as the BOUNDARY variable. Then create a hybrid variable (BC) from variables B and C, where BC = B, if A = 1; and BC = C, if A = 2. The variable BC becomes the BIASVAR variable and the right-most variable in SWAPVARS. Suppose B and C are coarsened variables, and they are associated with more detailed variables called B_DETAIL and C_DETAIL, respectively. In order to swap the original variables B and C appropriately and also B_DETAIL and C_DETAIL, the parameters would be:

```
BOUNDARY = A
SWAPVARS = B C BC
BIASVAR = BC
LINKSWAP = B_DETAIL # C_DETAIL # NULL
```

2.2.2.5 One Variable Linked to two SWAPVARS Variables

To maintain consistency between SWAPVARS variables and other items when swapping occurs, the LINKSWAP parameter can be used. The program will not allow a variable to be listed more than once among SWAPVARS and LINKSWAP variables. So, when one variable needs to change when either of two SWAPVARS variables change then the following can be done.

Suppose SWAPVARS=A B C and the variable D needs to change whenever A or B change. The specification of LINKSWAP=D#D#....... is not allowed. It might be helpful to create one SWAPVARS variable from A and B, such as AB=10×A+B or AB=A||B (A concatenated with B so that A=1, B=2, AB=12). Then specify SWAPVARS=AB C and LINKSWAP=A B D#NULL, which will effectively change the values of A, B, and D whenever AB is swapped. The temporary variable AB can be dropped after processing is complete.

2.2.2.6 Key Outcome Variables

The swapping effect on key outcome variables such as assessment score can be controlled by creating a categorical variable and using it as a hard boundary variable. Doing so will force all swapping to occur within the same level of assessment score. Therefore, the swapping impact on assessment score is limited; increasing the number of categories of assessment score will limit the swapping impact even further. The continuous key outcome variable can also be specified in the KEYOUT parameter in order to monitor data quality due to the swapping effect.

2.2.2.7 Calibration of Survey Weights

As may be seen from examples presented in this guide, variables selected for swapping are frequently demographic variables such as age, gender, and race. These types of variables are also frequently used in the calibration of survey weights to known population totals. However, if swapping occurs after the calibration process, and common variables are used in both processes, the weighted totals may no longer match population totals within categories after swapping. For this reason, users should consider swapping as an intermediate step in the weighting process if the same – or related – variables are used in both calibration and swapping.

2.2.2.8 Using Data Utility Measures to Determine the Final Swapped Dataset

The DRB standards for *DataSwap* processing and analyses, involve (1) the selection, testing and finalization of the parameters; and (2) the running and rerunning of *DataSwap* while maintaining static parameters except for the SEED. Thus, the sampling methodology is maintained while different target samples are selected. There is enough variation generated between runs such that the swapping results can be improved by replicating the *DataSwap* run several times since there is a random component involved with selecting target records by using the SEED parameter. The SEED that affords the least data distortion would be selected for the final delivery.

2.2.2.9 Comparing Replicate Runs Using the Same Swapping Scenario

As a guideline, select two or three of the best replicated runs with the smallest multivariate data utility measures (those based on pairwise and regression associations). Then select the replicate run resulting in the best data utility among the HD measures among the initially chosen two or three best replicated runs.

2.2.2.10 Exploring Alternative Swapping Scenarios

It may be necessary to adjust other parameters in order to achieve the highest utility. At the time of running *DataSwap* the DRB may have already approved the Disclosure Analysis Plan, which includes the *DataSwap* parameters. Therefore, it is important to notify the DRB of any alternative scenarios being considered since parameters may not be changed without further approval.

One situation that may occur is that some records selected as targets may not have acceptable partners. To address this issue, the MOS parameter may be adjusted so that such problematic cases have a lower chance of selection as swapping targets.

In some cases it may be advantageous to use a more detailed or less detailed version of an approved swapping variable (for example, a three-level age category variable as opposed to six-level, or vice versa). Less detail in swapping variables results in larger swapping cells and more potential swapping partners for a selected target, which could in turn improve the data utility.

In contrast, using more detailed swapping variables may also be advantageous, provided that the resulting swapping cells are of acceptable size. Doing so results in more swapping cells, better swapping partners in adjacent cells, and may result in a lower bias. For example, if detailed age is swapped along with a six-level collapsed age (instead of a three-level variable), this will result in detailed age values closer together being swapped.

The order of the variables forming the swapping cells will also impact the data utility. If a *DataSwap* run results in unacceptable data utility, a logical reordering of the variables may have a positive impact.

Lastly, if none of the replicate runs have acceptable data utility, the reason may be the result of parameter mis-specification. The list of KEYOUT and KEYVARS should be reviewed since they can impact the data utility measures. In addition, the data utility results are impacted if any of the variable types (ordinal or nominal) of BOUNDARY, SWAPVARS, or KEYVARS are mis-specified.

2.2.2.11 Comparing Swapping Scenarios

It is imperative that with each change in parameter specification, utility measures from several replicate *DataSwap* runs (each with different target samples) be compared with each other, and where possible, against other sets of replicate runs to take into account not only the replicate variation but the effect of parameter changes. It is recommended that the swapping scenario (that is, the set of parameters) be selected based on the multivariate data utility measures (related to pairwise and regression coefficients). That is, select the swapping scenario with the best data utility on average – or the smallest values of the utility measures on average. Then select the replicate run (SEED parameter) resulting in the best data utility among the HD measures for the chosen swapping scenario.

For more information about the characteristics and comparability of the data utility measures across swapping scenarios, see Section 2.3.3.3.

Other helpful hints regarding these measures are:

- Careful KEYVAR and KEYOUT specification is required since both parameters are used to calculate the utility measure from the measures of association and the latter is used to calculate the utility measure from the regression coefficients.

- The multivariate measures of data utility using the correlations and using the multiple regression coefficients are sensitive to the proper specification of variable types: BOUNDARY_T (correlations only), SWAPVARS_T (correlations and regression), and KEYVARS_T (correlations only).

- If LINKSWAP variables are specified, these should also be included as KEYVARS. Doing so ensures they will be included in the correlation calculations and the data utility measure based on the correlations. If not included as KEYVARS, any loss of data quality as a result of swapping these variables will not be shown to the user.

- Since the HD measures depend on the TOLFLAG parameter to determine cells that are "too small," care should be taken when making this specification. The example in Appendix C results in the following HD measures:

```
                                                    For         Utility
               Application                       Variable(s)     value


Hellinger's Distance, all cells                  Across All     0.485251!
Hellinger's Distance, excluding small cells      Across All     0.000000
Hellinger's Distance, all cells                  DRACE3         0.142679!
```

```
Hellinger's Distance, all cells                EDUC3          0.136437!
Hellinger's Distance, all cells                DAGE3          0.061723!
Hellinger's Distance, excluding small cells    DRACE3         0.073021
Hellinger's Distance, excluding small cells    EDUC3          0.071382
Hellinger's Distance, excluding small cells    DAGE3          0.015712
```

The tolerance for cell sizes was set to the default of 45. However, the dataset has 182 observations and 23 swapping cells, resulting in an average cell size of 8. Hence, when all “small” cells were excluded, the HD measure across all the variables is zero. It may be that all or nearly all of the cells were excluded from this calculation and the remaining cells had no values swapped. Hence, the utility is measured on no data or only a small portion of the data. This utility value should not be used to conclude that the full swapped data file maintains the same utility as the original data file.

2.2.2.12 Using risk assessment results to inform the swapping process

The results of a risk assessment process can help inform the data swapping process. Suppose the values of the output variable, RiskStratum, are set to 0 for low risk, and up to 4 for the highest risk category. The values can be used to form the measure of size for the probability proportionate to size selection of target records for swapping. Let RS denote the values of the risk stratum. Then the measure of size can be set to be a function of RS, such as RS + 1, or $RS^2 + 1$. A value of 1 is added to ensure all records have a chance of selection.

2.2.3 Errors

If the program aborts due to errors, they will be printed in *<file name>_Test.log* or *<file name>_Run#1.log.* The program aborts with an error message under the following conditions:

1) If a record’s MOS, RATE, WGT, STRATUM, ID or BIASVAR has a missing value, for example:

   DataSwap Error: Variable MOS=“mos” has missing value(s). You will need to impute for missing data, or recode the variables to have a nonmissing code.

2) If ID, WGT, and, if specified, MOS, STRATUM, RATE, BIASVAR is not a single variable/value or if its value is out of range, for example:

   DataSwap Error: Parameter MOS must be a single variable or a value. You have specified x y.

DataSwap Error: Parameter MOS must be 1 when entered as a number. You have specified 3.

DataSwap Error: Parameter MOS must have values greater than 0. Please check your variable.

3) If the RATE is not constant within a stratum or it is not a value in (0,1], for example:

DataSwap Error: Parameter RATE must be constant within a stratum.

DataSwap Error: Parameter RATE must be between '(0, 1]'. You have specified 2.0.

DataSwap Error: Parameter RATE must values between '(0, 1]'. Please check your variable.

4) If the number of potential swap partners is less than the number of swaps, for example:

DataSwap Error: NOT ENOUGH CASES TO DO THE SWAP. The sampling rate (RATE) must be reduced.

5) If one or more cases have no swap partner, for example:

DataSwap Error: At least one selected case has no swapping partner. Run a cross-tabulation on the BOUNDARY*SWAPVARS variables to determine how to collapse cells so that swapping partners can be found.

6) If there is no change between the swap and its partner, for example:

DataSwap Error: Failed to change the value of SWAPVARS="A B" while swapping id=7 and id=9.

7) If the number of SWAPVARS does match that of the sets in MISSINGDEF when IMPUTE=y, for example:

DataSwap Error: The number of variables in SWAPVARS does not match the number of sets in MISSINGDEF.

8) If SWAPVARS variable X's type does not match the syntax in MISSINGDEF, for example:

DataSwap Error: Variable X is numeric but the missing values in MISSINGDEF are in quotes or

DataSwap Error: Variable X is character but the missing values in MISSINGDEF are not in quotes.

9) If a variable has only missing values in a boundary cell when IMPUTE=y, for example:

DataSwap Error: Variable X is all missing in at least one boundary group and cannot be fully imputed.

10) When SWAPMETH = 2, all variables in SWAPVARS must be numeric, for example:

DataSwap Error: Variable X needs to be numeric type for SWAPMETH=2 to work.

11) When SWAPMETH = 2, all variables in SWAPVARS must contain non-missing values, for example:

DataSwap Error: Variable X cannot have missing value for SWAPMETH=2 to work.

12) Parameter SWAPMETH can only take values of 1 or 2, for example:

DataSwap Error: Parameter SWAPMETH can only be 1 or 2.

13) Within the lists of SWAPVARS and LINKSWAP variables, a variable can only be listed once.

DataSwap Error: One of your LINKSWAP - <variable name> has been used twice. Please check your parameters. Program has aborted.

14) The number of variables in the BOUNDARY_T, SWAPVARS_T and KEYVARS_T parameters must be the same as the number of BOUNDARY, SWAPVARS and KEYVARS respectively:

DataSwap Error: Parameter KEYVARS_T should have equal number of items as in KEYVARS. You have specified KEYVARS = X and KEYVARS_T= O # N.

15) If the parameter(s) BOUNDARY or KEYVARS is not specified, then the respective type parameter(s) must be left blank, for example:

DataSwap Error: Parameter BOUNDARY_T must be left blank when BOUNDARY is not used. You have specified BOUNDARY_T=N.

16) A swapping variable cannot be included as a KEYOUT variable:

DataSwap Error: Variable used in SWAPVARS cannot be used in KEYOUT. You have specified SWAPVARS=X. and KEYOUT=X.

17) The variable specified as the bias variable must also be included in the list of swapping variables:

DataSwap Error: Parameter BIASVAR must be a variable in SWAPVARS. You have specified SWAPVARS= X Y and BIASVAR=Z.

18) The number of SWAPVARS and parameters separated by “#” signs must match, for example:

DataSwap Error: The number of variables in SWAPVARS does not match the number of sets in LINKSWAP. You have specified SWAPVARS=X Y and LINKSWAP=X1 X2 # Y1 # Y2.

DataSwap Error: The number of variables in SWAPVARS does not match the number of sets in LINKSWAP. You have specified SWAPVARS=X Y and LINKSWAP=X1 X2.

19) All variables specified as being ordinal in the BOUNDARY_T, SWAPVARS_T, or KEYVARS_T must be numeric:

DataSwap Error: Variable X must be a numeric variable if its type is O. You have defined X as a character variable.

20) There can be no more than 20 SWAPVARS specified for any particular *DataSwap* run:

DataSwap Error: The maximum number of variables allowed in SWAPVARS is 20. You have specified 24.

21) In some extreme condition, no target record is selected:

DataSwap Error: There were no cases selected during the sampling process. Consider making a change to the strata or rate parameter.

22) The parameters SWAPVARS_MD, BOUNDARY_MD, KEYOUT_MD, or KEYVARS_MD should be left blank when no SWAPVARS, BOUNDARY, KEYOUT, or KEYVARS was specified.

DataSwap Error: Parameter KEYVARS_MD should be left blank when no KEYVARS was specified. You have specified KEYVARS_MD=7 8##7 8 9;

23) The parameters SWAPVARS_MD, BOUNDARY_MD, KEYOUT_MD, or KEYVARS_MD should have equal number of items as in SWAPVARS, BOUNDARY, KEYOUT, or KEYVARS.

DataSwap Error: Parameter KEYVARS_MD should have equal number of items as in KEYVARS; You have specified KEYVARS=BIC0501 BID0101 and KEYVARS_MD=7 8##7 8 9.

2.3 Output Reports

The amount of output generated from *DataSwap* depends on the number of seeds specified in the macro call. For each seed specified, the output described in this section will be provided in the respective .lst and graphics (.pdf or .rtf) files. If more than one seed is specified, additional output will be created; the utility measures from all runs will be summarized in <*file name*>*.lst* and graphed in <*file name*>*_SummaryGraph.pdf* (or <*file name*>*_SummaryGraph.rtf* depending on the graphics specification). See Section 2.3.5 for details on the additional reports provided with multiple seeds.

Certain information included in the output is considered confidential. Thus, the *DataSwap* output contains information for the user, the DRB chair, and the DRB members. The user will need to cut and paste the appropriate output into a disclosure analysis plan (DAP) for the DRB.

The format of the output found in the .lst file generated from each seed provided for an individual *DataSwap* run is presented in Table 2-1 below.

Table 2-1. Individual *DataSwap* Output Report Structure

Information Page
User-Only Output
Frequency of Change Flags
Sampling Result and Imputation
DataSwap Results by Pairs
Aggregated Level -- Changes to Swapped Variable
Unweighted Percents
Weighted Percents and Standard Errors
Weighted Means and Standard Errors
Weighted Percents Sorted by Relative Difference
Weighted Means Sorted by Relative Difference
Supplemental Tables for DRB Chair Only
Percent of Records Changed Among Swapped Variables in SWAPVARS
Percent of Records Changed Among Swapped Variables in LINKSWAP
Supplemental Tables for DRB Members
Aggregated Level -- Changes to Swapped Variable
Unweighted Percents
Weighted Percents and Standard Errors
Weighted Means and Standard Errors
Weighted Percents Sorted by Relative Difference
Weighted Means Sorted by Relative Difference
Check creation of indicator variables
Unweighted Measures of Association in Swapping Variables and Key Output Variables
Weighted Measures of Association in Swapping Variables and Key Output Variables
Unweighted Multiple Regression Coefficients for Default Model
Weighted Multiple Regression Coefficients for Default Model
Unweighted Multiple Regression Coefficients for User-specified Models
Weighted Multiple Regression Coefficients for User-specified Models
Global Data Utility Measures

Additionally, three graphs are produced in separate files. These may be shared with the DRB members, so they can be considered as part of the last category in Table 2-1.

- Plot of weighted percents before and after swapping - *<file name>_PLOT_PCT_Run#.pdf* (or *<file name>_PLOT_PCT_Run#.rtf* depending on the graphics specification).

- Plot of weighted means before and after swapping - *<file name>_PLOT_MEAN_Run#.pdf* (or *<file name>_PLOT_MEAN_Run#.rtf* depending on the graphics specification).

- Plot of weighted measures of association before and after swapping - *<file name>_PLOT_CORR_Run#.pdf* (or *<file name>_PLOT_CORR_Run#.rtf* depending on the graphics specification).

First, an information page is provided as a check to ensure parameters have been specified correctly. The information sheet is, in general, a reflection of the parameter sheet. The user should use this sheet to check the specifications of the parameters. The information sheet also provides the following:

- ***DataSwap* macro version #:** Current version of the software.
- **Observations in input data set:** Number of observations.
- **Swap cell definition:** Shows the concatenation of the BOUNDARY||SWAPVARS variables used to define the swapping partner cells.
- **Total number of cells:** Number of cells defined by the concatenation of the BOUNDARY||SWAPVARS variables.
- **Total number of iterations:** The total number of iterations that are needed to find swapping partners for each target record. Since a record can be the best swapping partner for more than one target selection record, and since it can only be used once, then the process goes through another cycle in order to find the next best swapping partner that is available. The cycles continue until all records have swapping partners, or until the unfavorable result occurs where no swapping partner could be found.

After the information page, the output reports proceed with user-only output (Section 2.3.1), DRB chair only output, (Section 2.3.2), and DRB member output (Section 2.3.3).

2.3.1 User-Only Output

The user-only output (i.e., only the user should see this output) consists of detailed output reports. First, two-way frequencies of any imputed SWAPVARS against their original values are produced. Second, summary counts are produced for each SWAPVARS and LINKSWAP variable showing counts and rates of all records that changed values (i.e., the perturbation rate for each variable).

Third, a printout shows counts of targets, swapping partners (donors), and records not selected for the swapping process. This is useful for checking the sampling rate of target records (i.e., swapping rate). A similar additional printout is shown by STRATUM, to check the sampling rates by stratum.

Fourth, a listing of SWAPVARS and LINKSWAP variables before and after swapping gives the user the opportunity to do a thorough check of the swapping results. The listing consists of all swapping pairs (i.e., target cases with their corresponding swapping partner cases) or some subset depending on the parameter specification. The printout shows the partner case directly underneath the target record so that the pairs are shown together. The listing will show all the variables related to the SWAPVARS and LINKSWAP parameters.

Fifth, the following tables and measures of association will be produced:

- Unweighted percentages for each SWAPVARS and LINKSWAP variable before and after swapping.
- Weighted percentages for each SWAPVARS and LINKSWAP variable before and after swapping, using WGT as the weight variable. If VARSTRAT and VARUNIT are specified, the standard errors of these percentages before and after swapping are presented along with their ratio.
- Weighted means of the KEYOUT variables specified for each level of the variables specified in the SWAPVARS and LINKSWAP parameters before and after swapping, using WGT as the weight variable. If VARSTRAT and VARUNIT are specified, the standard errors of these means before and after swapping are presented along with their ratio.

The weighted results are also reported in descending order of absolute relative difference, separately by weighted percentages and weighted means.

Large differences between weighted percentages and means before and after swapping are indicated with an asterisk. If the difference between the before- and after-swapping estimates is more than *constant_x* and the sample size used in the estimate is greater than *constant_y*, then an asterisk is placed next to the value of the relative difference. *Constant_x* and *constant_y* are the first and second values specified in the TOLFLAG parameter.

Large differences in the before- and after-swapping standard errors are indicated with an "@". If the ratio of the after-swapping to the before-swapping standard errors is larger than *constant_x* and the sample size used in the estimate is more than *constant_y* then an "@" is placed to the right of the ratio. Note that the presence of "@" does not necessarily indicate statistical significance regardless of the values of these parameters. *DataSwap* is not designed to calculate the appropriate measures for significance testing.

In the event that the before-swapping value of a weighted mean is zero and the after-swapping value is non-zero, a "~" will appear next to the after-swapping value and the following message will appear at the bottom of the page: "~ denotes value was zero before swapping and non-zero after swapping."

For user-only tables, a SWAPVARS and LINKSWAP variable is excluded from the output reports if the number of categories is greater than 300.

2.3.2 Tables for DRB Chair Only

The following data are provided in output intended for the DRB chair only.

- The percentage of records with changes for each SWAPVARS variable.
- The overall percentage of all records that had at least one value changed.
- The percentage of records with changes in each LINKSWAP variable.

2.3.3 Tables for the DRB Members

The following summary data are produced for the DRB members for their review of the swapping results (some of this information can be cut and pasted into the DAP). These are described in detail in Section 2.3.1. If any variable has more categories than MAXCAT, tables involving that variable are omitted from these tables (but will be present in the user-only output).

- Unweighted percentages for each SWAPVARS and LINKSWAP variable before and after swapping.
- Weighted percentages for each SWAPVARS and LINKSWAP variable before and after swapping, using WGT as the weight variable. If VARSTRAT and VARUNIT are specified, the standard errors of these percentages before and after swapping are presented along with their ratio.
- Weighted means of the KEYOUT variables specified for each level of the variables specified in the SWAPVARS and LINKSWAP parameters before and after swapping, using WGT as the weight variable. If VARSTRAT and VARUNIT are specified, the standard errors of these means before and after swapping are presented along with their ratio.

The weighted results are also reported in descending order of absolute relative difference, separately by weighted percentages, and weighted means.

Further measures of data quality using the BOUNDARY, SWAPVARS, KEYVARS, and KEYOUT variables are also printed for the DRB members.

- Crosstabulations illustrating the creation of dummy variables for any nominal variables specified for use in the measures of association.

- Unweighted and weighted measures of association. These are described further in Section 2.3.3.1.
- Unweighted and weighted regression output. These are discussed further in Section 2.3.3.2.
- Data utility measures for tables, pairwise associations, and regression coefficients. These are discussed further in Section 2.3.3.3.

2.3.3.1 Measures of Association

Part of the output contains unweighted and weighted Pearson product correlations (expressed as *r*), the formulae for which are shown in Section 3.6.2. Since the Pearson product correlations are not statistically appropriate for certain types of variables, the user is encouraged to specify any nominal variables using the BOUNDARY_T, SWAPVARS_T or the KEYVARS_T parameters as appropriate. Nominal variables, sometimes referred to as qualitative variables, have categories that you cannot sort in any way that makes sense. Nominal variables therefore are not directly used in correlations or regression models; however, as is done in *DataSwap*, nominal variables are transformed into a series of indicator variables so that statistical relationships with other variables can be quantified in correlations and regression models. Examples of nominal variables are race, gender, and geographic regions. Ordinal variables, sometimes referred to as quantitative variables, have categories that can be sorted in a meaningful way, for instance, from largest to smallest. Examples of ordinal variables are age groups, income categories, and height. Ordinal variables can also be continuous measures, for example, age in years, income, and person weight.

Any nominal variable specified as a BOUNDARY, SWAPVARS, or KEYVARS with more than two levels is converted into a series of indicator variables, with one indicator variable created for each value level. These new variables are named *<nominal variable name>_1*, *<nominal variable name>_2*, … *<nominal variable name>_m,* where *m* is the number of value levels of the nominal variable. These created indicator variables are used in the measures of association rather than the parent variable. Nominal variables with only two levels, other than binary variables with 0/1 values, will also be converted to indicator variables in the same manner, but only the variable indicating the first level will be used in the correlation to prevent duplicative results presented in the output. Binary 0/1 variables will be used as-is.

If indicator variables are created and are useful in the correlations, a message is displayed at the bottom of the page to alert the user as follows: "Variable(s) <nominal variable 1>_*, < nominal variable name 2>_* … are indicator variables."

If any variables have alpha characters, then they are recoded by sorting and then sequentially numbering the categories 1, 2, 3, … A message is displayed at the bottom of the page to alert the user as follows, "Variable(s) <*variable or list of variable*>' in these calculations have been recoded to sequential numbers."

Two sets of tables are created:

- **Table set 1**. "Unweighted Measures of Association Among Swapping Variables and Key Output Variables." This table shows *r* among all unweighted pairwise combinations of BOUNDARY, SWAPVARS, KEYVARS and KEYOUT variables computed before and after swapping, where $r_{before} - r_{after} \neq 0$. (If it is desired to have LINKSWAP variables also included, as in previous versions of the software, these must be additionally specified as KEYVARS variables.)

- **Table set 2**. "Weighted Measures of Association Among Swapping Variables and Key Outcome Variables." This table shows *r* among all pairwise combinations of SWAPVARS, LINKSWAP, KEYVARS and KEYOUT variables computed before and after swapping, where $r_{w,before} - r_{w,after} \neq 0$.

Only pairs for which there is a difference before and after swapping are printed. The total number of association measures computed will be printed at the bottom of the page as follows: "Number of Pairwise Associations computed: *X*." If a table set is empty, then no output is displayed.

Large differences between pairwise correlations before and after swapping may be indicated with an asterisk. If the difference between the before- and after-swapping estimates is more than *constant_z* multiples of the before-swapping standard error and the sample size used in the estimate is greater than *constant_y*, then an asterisk is placed next to the value of the relative difference and a message appears at the bottom of the page as follows: "* denotes the after swapping correlation changed by more than *constant_z* standard errors." *Constant_z* and *constant_y* are the third and second values specified in the TOLFLAG parameter.

Note that the presence of an asterisk does not necessarily indicate statistical significance regardless of the value of *constant_z*. *DataSwap* is not designed to calculate the appropriate measures for significance testing.

2.3.3.2 Regression Results

By default, unweighted and weighted regression models are produced with each KEYOUT variable as a dependent variable and all the SWAPVARS as the independent variables. Additionally, users may specify other regressions using any variables as the dependent or independent variables that are specified as BOUNDARY, SWAPVARS, KEYOUT or KEYVARS variables. Care should be taken to only specify models that are of analytic value for the results of the models will be incorporated into some of the data utility measures. Specification of models should have some intrinsic meaning or face validity.

The user is encouraged to specify any nominal variables in the regressions models using the BOUNDARY_T, SWAPVARS_T or the KEYVARS_T parameters as appropriate. (See the discussion in Section 2.3.3.1 for more information on the definitions of nominal and ordinal variables.) If nominal variables are included as part of the default (including all SWAPVARS as independent variables) or user-specified regression models, the model is first passed through the GLMMOD procedure in SAS to create a data set that contains the design matrix for a model as specified using the effects modeling facilities of the GLM procedure. Then the regression model is fit using the REG procedure. For information about the GLMMOD and REG procedures in SAS, visit http://support.sas.com.

As many as four sets of regression results are created.

The first and second sets of results are from the unweighted and weighted regressions of the KEYOUT variables on the group of swapping variables. The error degrees of freedom, estimates of parameter coefficients, and R^2 values are printed for each model computed on the data before and after swapping.

Any nominal variable is converted to *m-1* indicator variables in the GLMMOD procedure, where *m* is the number of value levels of the nominal variable. These variables will appear in the output named *<nominal variable name> <level 1>, <nominal variable name> <level 2>, …<nominal variable name> <level m-1>.*

If any other models are specified by the user, the third and fourth sets of results are unweighted followed by weighted regression results. Nominal variables will be treated as in the user-specified models, regardless of the number of levels.

For all models, large differences between estimates of regression parameter coefficients before and after swapping may be indicated with an asterisk. If the difference between the before and after swapping estimates is more than *constant_z* multiples of the before-swapping standard error and the sample size used in the estimate is greater than *constant_y*, then an asterisk is placed next to the value of the after-swapping value and a message appears at the bottom of the page as follows: "* denotes the after swapping beta coefficient changed by more than *constant_z* standard errors." *Constant_z* and *constant_y* are the third and second values specified in the TOLFLAG parameter, respectively. The standard errors of the parameters are calculated by the PROC REG procedure in SAS. Note that the presence of an asterisk does not necessarily indicate statistical significance regardless of the values of these parameters. *DataSwap* is not designed to calculate the appropriate measures for significance testing.

2.3.3.3 Data Utility Measures: Assessing Data Quality

The objective of including global data utility measures into *DataSwap* is to have some means to compare multiple swapped datasets that were generated with different random seeds (with all other parameters kept constant). The user is encouraged to use, as the final swapped dataset, the run with the best values of the data utility measures among at least five *DataSwap* runs. For all data utility measures, the lower the measure the higher the data utility. For further demonstration of the variability of the measures among replicated runs and how the measures may be used to evaluate various swapping scenarios and see Dohrmann, et al (2009).

Three sets of global data utility measures are discussed below. The measures aggregate the following computations conducted for the swapping impact analysis: weighted 1) tables cells, 2) pairwise correlations, and 3) regression parameters. Formulae for all measures are provided in Section 3.6.

2.3.3.4 Global Data Utility Measures for Tables

DataSwap computes the Hellinger's Distance (HD) as a global measure of data utility for tabular results. As discussed in Gomatam, et al (2005), Hellinger's Distance is impacted by small cells. The impact of small cells on the HD measure motivates four possible applications, all of which have been implemented into *DataSwap* as follows:

Hellinger's Distance, all cells, across all variables. The weighted original data are tabulated by crossing all SWAPVARS variables. Then the weighted swapped data are tabulated in the same manner. HD is then computed according to the equation above to provide a single value of global data utility. If any cells used in the computation are less than or equal to *constant_y*, an exclamation point is printed to the right of the measure and the following note appears at the bottom of the page: "! denotes small cell(s)." The column of “Cell Count/ Small Cell Count” provides the total number of cells as well as total number of small cells. *Constant_y* is the second value specified in the TOLFLAG parameter.

Hellinger's Distance, excluding small cells, across all variables. The weighted original data are tabulated by crossing all SWAPVARS variables. Then the weighted swapped data are tabulated in the same manner. HD is then computed according to the equation above using only those cells with more than *constant_y* observations. The column of “Cell Count/ Small Cell Count” provides the total number of cells as well as total number of small cells. C*onstant_y* is the second value specified in the TOLFLAG parameter.

Hellinger's Distance, all cells, for SWAPVARS. The weighted original data are tabulated for each SWAPVARS variable separately. Then the weighted swapped data are tabulated in the same manner. HD is then computed for each SWAPVARS according to the equation above to provide one value of data utility for each SWAPVARS variable. If any cells used in the computation are less than or equal to *constant_y*, an exclamation point is printed to the right of the measure and the following note appears at the bottom of the page: “! denotes small cell(s).” The column of “Cell Count/ Small Cell Count” provides the total number of cells as well as total number of small cells. C*onstant_y* is the second value specified in the TOLFLAG parameter.

Hellinger's Distance, excluding small cells, for SWAPVARS. The weighted original data are tabulated for each SWAPVARS variable separately. Then the weighted swapped data are tabulated in the same manner. HD is then computed for each SWAPVARS using only those cells with more than *constant_y* observations to provide one value of data utility for each SWAPVARS variable. The column of “Cell Count/ Small Cell Count” provides the total number of cells as well as total number of small cells. C*onstant_y* is the second value specified in the TOLFLAG parameter.

2.3.3.5 Global Data Utility Measures for Pairwise Associations

Comparing correlations before and after swapping has been proposed in the literature, including the relative difference in Cramer’s V and the Contingency Coefficient (as given in Gomatam et al (2005)). *DataSwap* calculates the following three measures:

1) Pearson Product Correlation Methodology

2) Pearson’s Contingency Coefficient

3) Cramer’s V

2.3.3.6 Global Data Utility Measures for Multivariate Associations

DataSwap compares multivariate relationships before and after swapping using regression coefficients with swapping impact measured in terms of the number of standard error deviations.

2.3.3.7 Graphs

In addition to the printed output, *DataSwap* also provides graphical representations of the impact of swapping on weighted percents, means, and pairwise associations. These are in the form of scatterplots of the before- and after-swapping estimates and are saved as separate graphics files in the same directory as the other *DataSwap* output.

2.3.3.8 Graph of Weighted Percents

To further illustrate the effect of swapping on the weighted percents of the SWAPVARS and LINKSWAP, *DataSwap* generates a scatterplot of weighted cell percents with X-axis showing the estimates before swapping and the Y-axis showing the estimates after swapping. One plot is used to show all the cell percents before and after swapping for each SWAPVARS and LINKSWAP variable that have sample sizes larger than *constant_y* specified in the TOLFLAG parameter. Each variable is represented by a different symbol and color. The graph also includes a 45 degree reference line. The further a percent marker is away from this line, the greater the impact of swapping on the cell percents.

The graph is saved as *<file name>_PLOT_PCT_Run#.pdf* (or *<file name>_PLOT_PCT_Run#.rtf* depending on the graphics specification).

2.3.3.9 Graph of Weighted Means

To further illustrate the effect of swapping on the weighted means of the KEYOUT variables, *DataSwap* generates a scatterplot of weighted means of each KEYOUT variable before and after swapping in the cells specified for each level of SWAPVARS and LINKSWAP. The X-axis shows the estimates before swapping and the Y-axis shows the estimates after swapping. One plot is used to show all the weighted means for each KEYOUT variable before and after swapping in each SWAPVARS and LINKSWAP cells. Only those weighted means that are based on sample sizes larger than *constant_y* specified in the TOLFLAG parameter will be graphed. Each SWAPVARS and LINKSWAP variable is represented by a different symbol and color. The graph also includes a 45 degree reference line. The further an estimate's marker is away from this line, the greater the impact of swapping on the weighted means.

The graph is saved as *<file name>_PLOT_MEAN_Run#.pdf* (or *<file name>_PLOT_MEAN_Run#.rtf* depending on the graphics specification).

2.3.3.10 Graph of Correlation Coefficients

The scatterplot of the weighted correlation coefficients between SWAPVARS, LINKSWAP, KEYVARS, and KEYOUT variables illustrates the effect of swapping on these measures. The X-axis shows the estimates before swapping and the Y-axis shows the estimates after swapping. Each circle on the plot represents a correlation coefficient between a pair of variables. The graph also includes a 45 degree reference line. The further a correlation coefficient is away from this line, the greater the impact of swapping on that measure.

The graph is saved as *<file name>_PLOT_CORR_Run#.pdf* (or *<file name>_PLOT_CORR_Run#.rtf* depending on the graphics specification).

2.3.4 Parameters, Variable Types, and their Inclusion in the Output Reports

Some *DataSwap* parameters allow variables with numeric or character values, while others require numeric variables only. Some *DataSwap* parameters require that a variable type be specified for all variables so that they may be appropriately included in correlations and regressions. If a variable contains non-ordered (or nominal) categories, they will be converted to dummy indicator variables before being included in the correlations and used as class variables in the regressions. While all these details are presented in Section 2.1.2, the tables below provide a more comprehensive look at which types of variables are permitted as parameters, which parameters are included in the various output reports, and when nominal variables are used in their original or dummy-variable form.

Table 2-2 presents the parameters by type of variable: nominal or ordinal, character or numeric, and fewer or more than the limit of 16 categories. Note that the threshold of 16 may be increased by specifying NL = *new value* in the macro parameter sheet.

Table 2-2. Parameters and Variable Types Allowed in *DataSwap*

	Type of Variable*					
	Ordinal		Nominal**			
Parameter	Numeric	Character (not recommended)	Numeric ≤ 16 Levels	Numeric > 16 Levels	Character ≤ 16 Levels	Character > 16 Levels
BOUNDARY	Allowed	Allowed; excluded from correlations, regressions (if specified), and utility measures	Allowed; dummy variables used in correlations	Allowed, excluded from correlations, regressions (if specified), and utility measures	Allowed; excluded from correlations, regressions (if specified), and utility measures	Allowed; excluded from correlations, regressions (if specified), and utility measures
BIASVAR	Allowed	Not allowed	Not recommended; dummy variables used in correlations, regressions and utility measures	Not recommended; excluded from correlations, regressions, and utility measures	Not allowed	Not allowed
SWAPVARS	Allowed	Allowed; excluded from correlations, regressions, and utility measures	Allowed; dummy variables used in correlations, regressions and utility measures	Not recommended; excluded from correlations, regressions, and utility measures	Allowed; dummy variables used in correlations, regressions and utility measures	Allowed; excluded from correlations, regressions, and utility measures
LINKSWAP	Allowed	Allowed	Allowed	Allowed	Allowed	Allowed
KEYVARS	Allowed	Allowed; excluded from correlations, regressions, and utility measures	Allowed; dummy variables used in correlations and utility measures	Allowed; excluded from correlations, regressions (if specified) and utility measures	Allowed; dummy variables used in correlations and utility measures	Allowed; excluded from correlations, regressions (if specified) and utility measures
KEYOUT	Allowed	Not allowed	Not recommended; will be incorrectly treated as ordinal in correlations, regressions, and utility measures	Not recommended; will be incorrectly treated as ordinal in correlations, regressions, and utility measures	Not allowed	Not allowed

* User is required to specify variable types for each BOUNDARY, BIASVAR, SWAPVARS, and KEYVARS.

** Number of levels for conversion to dummies is set to a default of 16, but can be changed with the NL parameter (see the description of parameter SWAPVARS_T in Section 2.1.2).

Table 2-3 presents which parameter variables are included in the various reports, and whether nominal variables are used in their original state or in the created dummy variable form for the various calculations.

Table 2-3. Variable Lists Included in Printed *DataSwap* Output*

Output	BOUNDARY	SWAPVARS	LINKSWAP***	KEYVARS	KEYOUT
Tables (%)	X	X	X		
Tables (Means)	X	X	X		X
Correlations	$X_{(D)}$	$X_{(D)}$	→	$X_{(D)}$	X
Regressions	$X_{(D)}$**	$X_{(D)}$	→	$X_{(D)}$**	X
Data utility measures based on…					
Hellinger's distance		X			
Pearson correlations	$X_{(D)}$	$X_{(D)}$	→	$X_{(D)}$	X
Pearson contingency coefficient	X	X	→	X	X
Cramer's V	X	X	→	X	X
For regression coefficients		$X_{(D)}$			X

* X indicates all variables in their original form. $X_{(D)}$ indicates the dummy variables created for nominal variables with number of categories less than or equal to the analysis threshold, and all other variables in their original form.

** BOUNDARY and KEYVARS variables are not included in the default regression models, but may be included in additional user-specified models.

*** → indicates that the LINKSWAP variables may be included if also specified as KEYVARS.

As seen in table 2-3, the following four parameters are used in the computation of measures of correlations and regressions: SWAPVARS, BOUNDARY, KEYOUT, and KEYVARS. Special missing values (e.g., 7, 8, and 9) can be designated to remove the records from the computation by specifying the parameters SWAPVARS_MD, BOUNDARY_MD, KEYOUT_MD, and KEYVARS_MD, and all these details are presented in Section 2.1.2. For example, if missing value code is specified as a 7, then the value will be changed to SAS missing and be removed from the computations – that is, a complete case analysis will be conducted.

2.3.5 Additional Output Provided when Multiple Seeds are Specified

If more than one seed is specified, additional output will be created; the utility measures from all runs will be summarized in *<file name>.lst* and graphed in *<file name>_SummaryGraph.pdf* (or *<file name>_SummaryGraph.rtf* depending on the graphics specification) and will appear in the same directory as the other *DataSwap* output.

The summary list of the utility measures for tables, pairwise associations, and multivariate associations are displayed with the values for each run appearing in columns labeled with the run number and including the seed used for each run, respectively. Since the only difference between the swapping runs displayed in this output is the seed specified, the utility measures are directly comparable across runs. For all measures, a lower value indicates more utility; however, care must be taken with the Hellinger’s Distance measures as they are greatly impacted by small cells. (See Section 2.3.3.3.)

The summary graphics file includes three graphs. The first graph contains the values of Hellinger's Distance across all swapping variables with and without the small cells. The second graph contains the three utility measures for pairwise association. The third graph includes the Regression utility measure across all models. A line is drawn between the symbol markers for a measure across the run, not to display any trend across runs, but only as a visual aid. Note that only the overall measures are provided in the graphs. If particular variables or relationships are of concern, the user should refer to the summary listing to determine the run that produces the best utility for those variables or relationships.

3. TECHNICAL DESCRIPTION

This section provides more technical descriptions of the *DataSwap* algorithm provided in this version of the software. This section provides a technical description of the *DataSwap* algorithm including the specifics of target selection (Section 3.1), swapping partner selection (Section 3.2), and variance estimation (Section 3.3). The calculations of the measures of association, regression results, and utility measures are provided in Sections 3.4, 3.5, and 3.6 respectively. Section 3.7 describes the error that may be introduced as a result of data swapping. Section 3.8 provides guidance on the interpretation of utility measures.

3.1 Target Selection

The first step in the swapping algorithm is target selection. Targets that are selected with certainty are identified first and separated from the other cases; all other targets are selected from the remaining data records according to the specified sampling (swapping) rate.

The procedure for identifying certainty selections is to first sort the records (denoted by *i*) by descending measure of size within each stratum *h* (MOS_{hi}), then a target record is selected with certainty if the following is true:

$$n_h \frac{MOS_{hi}}{\sum_i MOS_{hi}} \geq 1,$$

where,

$n_h = round(N_h * RATE_h)$ is the number of targets to be selected in stratum *h*;

N_h = number of records in the file for stratum *h*; and,

$RATE_h$ = sampling rate for target selection for stratum *h*.

After the identification of each certainty case, the sample size n_h is reduced by 1, and its contribution to $\sum_i MOS_{hi}$ is removed. Subsequent cases are tested for certainty status using the updated values in the same manner.

Within a stratum *h*, all other targets are selected systematically with probability:

$$p_{hi} = n_h \frac{MOS_{hi}}{\sum_i MOS_{hi}} .$$

When no strata are defined, calculations defined in the formulae above are performed over the entire dataset (i.e., the "*h*" subscript is removed from the formulae).

3.2 Swapping Partner Selection

The variables to be swapped are typically identifying variables (e.g., individual descriptors, such as sex; or physical location descriptors, such as region). Not all such identifying variables are required to be used, but certainly a reasonable subset is required. If the chosen identifying variables are continuous, it is highly recommended that they be categorized. These categorical variables are then concatenated together to form swapping cells and ordered in such a way so that neighboring cells can be considered similar on potentially identifiable variables.

Each target case is in a swapping cell. Before performing the swap, each target case must be assigned a unique swapping partner. For each target, the program searches for swapping partners in neighboring cells. Within each neighboring cell, the case with the sampling weight closest in absolute value to the target case's sampling weight is identified. If ties occur, they are handled through a random selection process. At this point, each target case has been assigned one[7] or two potential swapping partners. The potential partner with the smallest absolute swapping bias is designated as the swapping partner, with ties handled by a random selection. The swapping bias is $\left[(w_s x_p + w_p x_s) - (w_s x_s + w_p x_p)\right]$, where w_s and w_p are the respective sampling weights for the target case, *s*, and its partner, *p*. The variable *x* represents the last swapping variable in the concatenation described above that defined the swapping cells (default). If a swap occurs with a partner with the same weight as the swapping case, then the swapping bias defined above will be zero.

A case can be used as a swapping partner only once. If one partner from the full sample provides the lowest bias for more than one swapping case, it is assigned to the case with the minimum absolute bias. Thus, cases from the full sample that have been identified as potential swapping partners are processed first. After a swapping partner is assigned to a swapping case, the swapping partner from the full sample is removed from eligibility, and the process is repeated until all swapping cases have been assigned a swapping partner.

[7] A target record can only have one potential swapping partner if hard boundaries are involved or the target record is in the first or last cell in the file.

3.3 Variance Estimation

The Taylor series linearization method (Lohr, 1999; Särndal, Swensson, and Wretman, 1992; Wolter, 1985) is used to estimate standard errors of weighted means or proportions before swapping. When there are stratifications and clusters in the sample design, this method requires the input of first-stage Primary Sampling Unit (PSU) and first-stage stratum identification. However, the method cannot be directly applied to the swapped data for the standard errors after swapping because it does not account for the variance due to swapping. This will result in an underestimation of variance. Following the idea in Li, Krenzke, Brick, Judkins, and Larsen (2011), the formula below is used for the variance estimation after swapping.

$$\operatorname{var}(\hat{\varphi}) = \operatorname{var}(\hat{\theta}) + (\hat{\varphi} - \hat{\theta})^2 ,$$

where $\hat{\theta}$ represents the estimate of a mean or proportion, θ, using the un-swapped data, and $\hat{\varphi}$ is the estimate of θ based on the swapped data. The first term on the right is the estimated variance of weighted means or proportions before swapping. It is computed using the un-swapped data and the Taylor series linearization method. This variance component accounts for the error due to sampling. The second term is the squared difference between the estimates before and after swapping. It accounts for the error due to swapping. As a result, the ratio of the standard errors, $\sqrt{se(\hat{\varphi}) / se(\hat{\theta})}$, is always no less than 1. Li et al. (2011) used a simulation study to show that the above variance estimator can appropriately estimate the variance for perturbed estimates.

3.4 Measures of Association

The unweighted measures of association between variables *x* and *y* are computed as:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

and weighted…

$$r_w = \frac{\sum w_i (x_i - \bar{x}_w)(y_i - \bar{y}_w)}{\sqrt{\sum w_i (x_i - \bar{x}_w)^2 \sum w_i (y_i - \bar{y}_w)^2}}$$

where,

w subscripted represents “weighted,” as in weighted means; and,

w as a variable represents the WGT variable.

The standard error of the correlation coefficient is calculated in *DataSwap* using the formula from normal theory (Wolter, 1985).

$$SE(r) = \frac{(1-r)^2}{\sqrt{n}}$$

and weighted…

$$SE(r_w) = \frac{(1-r_w)^2}{\sqrt{n}}$$

where,

w subscripted represents “weighted,” as in weighted means; and,

n is the number of non-missing values used in the correlation.

3.5 Regression Results

The unweighted and weighted regressions of the *h* KEYOUT variables on the group of swapping variables are calculated as follows:

$$KEYOUT_h = \beta_{h0} + \left(\sum_{j=1}^{k} \beta_{hj} SWAPVARS_j \right) + \varepsilon_h$$

where,

k = the number of swapping variables;

β_{hj} = the unknown parameter coefficient for the *j*th SWAPVARS variable and the *h*th KEYOUT variable; and,

ε_h = the unknown error for model with KEYOUT$_h$.

The usual error degrees of freedom, estimates of parameter coefficients, and R^2 values are also calculated and printed for each model computed on the data before and after swapping.

3.6 Utility Measures

3.6.1 Global Data Utility Measures for Tables

DataSwap computes the Hellinger's Distance (HD) as a global measure of data utility for tabular results**.** Gomatam et al (2005) give the HD formula as:

$$HD(\hat{N}_{orig}, \hat{N}_{swapped}) = \frac{1}{\sqrt{2}} \sqrt{\sum_c (\sqrt{\hat{N}_{orig}(c)} - \sqrt{\hat{N}_{swapped}(c)})^2}$$

where,

$\hat{N}_{orig}(c)$ = sum of weights for cell *c* on the original data; and,

$\hat{N}_{swapped}(c)$ = sum of weights for cell *c* on the swapped data.

3.6.2 Global Data Utility Measures for Pairwise Associations

The measure based on the Pearson product correlation is calculated as follows:

$$R\_ASED(Y_{orig}, Y_{swapped}) = \frac{\sum_{i \neq j} | r_{orig}(Y_i, Y_j) - r_{swapped}(Y_i, Y_j) | / SE(r_{orig}(Y_i, Y_j))}{n'}$$

where,

r = Pearson product correlation as computed in *DataSwap*;

n' = the total number of pairwise comparisons with pairwise comparison differences before and after swapping greater than 0; and,

SE = standard error (as given in Section 3.4).

The Pearson contingency coefficient (*C*) between two variables is equal to:

$$C(Y_i, Y_j) = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

where

$$\chi^2 = \sum_c \frac{(n_c - e_c)^2}{e_c};$$

n_c = actual frequency; and,

e_c = expected frequency.

The average absolute relative deviation (ARD) of the *C* measure is computed as follows:

$$C\_ARD(Y_{orig}, Y_{swapped}) = \frac{\sum_{i \neq j} | C_{orig}(Y_i, Y_j) - C_{swapped}(Y_i, Y_j) | / C_{orig}(Y_i, Y_j)}{n'}$$

where,

n' = the total number of pairwise *C* computations with differences before and after swapping greater than 0.

The Cramer’s V statistic (*V*) between two variables is equal to:

$$V(Y_i, Y_j) = \sqrt{\frac{\chi^2 / n}{\min(k-1, l-1)}}$$

where

k = number of categories for variable Y_i; and,

l = number of categories for variable Y_j.

The range is 0 ≤*V*≤1. The Cramer’s V is defined slightly differently for 2x2 tables and ranges from -1 to 1. The computation is equal to:

$$V(Y_i, Y_j) = \frac{(n_{11}n_{22} - n_{12}n_{21})}{\sqrt{n_{1\cdot}n_{2\cdot}n_{\cdot 2}n_{\cdot 1}}}$$

The average absolute relative deviation (*ARD*) of the *V* measure is computed as follows:

$$V\_ARD(Y_{orig}, Y_{swapped}) = \frac{\sum_{i \neq j} | V_{orig}(Y_i, Y_j) - V_{swapped}(Y_i, Y_j) / V_{orig}(Y_i, Y_j) |}{n'}$$

where,

n' = the total number of pairwise C computations with differences before and after swapping greater than 0.

3.6.3 Global Data Utility Measures for Multivariate Associations

A model-level measure is computed as the average absolute deviation between the weighted before and after swapping regression coefficients of the default weighted regression models, relative to the standard errors of the coefficients prior to swapping.

$$ASED_i(\beta_{orig}, \beta_{swapped}) = \frac{\sum_j | \beta_{orig}(i,j) - \beta_{swapped}(i,j) | / SE(\beta_{orig}(i,j))}{k_i}$$

where,

$\beta(i,j)$ = beta coefficient for model *i* and term *j*;

k_i = the total number of beta coefficients for model i ; and,

SE = standard error.

A global utility measure is computed as the unweighted average of the model-level measures:

$$ASED\_REG = \frac{\sum_i ASED_i(\beta_{orig}, \beta_{swapped})}{m}$$

where, *m* = number of models.

3.7 Measuring Sources of Error Due to Data Swapping

There are three sources of error introduced in estimates when a data file has been swapped: (1) bias in weighted distributions, (2) bias in correlations, and (3) bias in variances estimates. The following is a discussion of each of these biases and a description of how each is estimated. The discussion closely follows the information in Seastrom, Kaufman, and Roey (2005).

3.7.1 Bias in Weighted Distributions

If a swapping partner does not have the same weight as the target case, biases may be introduced into the weighted distribution of an individual swapping variable. A simple way of measuring this bias for an individual swapping variable is to compute the difference between the after-swapping distribution and the before-swapping distribution for each category in the distribution.

Additionally, a distributional bias can occur in averages of nonswapped variables that are cross-tabulated by a swapped variable. For example, the average math assessment score is tabulated by region, where region is a swapping variable, but scores were unaltered. Thus, the expected cross-tabulation swapping bias can also be measured for a number of nonswapped variables by each of the swapping variables; again, comparing these cross-tabulations before and after swapping is done. Given that the swapping procedure tries to minimize weight differences between the target case and its respective partner, it is expected that these biases will be relatively small even with a large swapping rate.

3.7.2 Bias in Correlations

Another source of bias is with weighted correlations between swapped and nonswapped variables, as well as correlation between two swapped variables. Again, the bias will be measured by computing the correlation difference between the after-swapping correlation and before-swapping correlation. The swapping procedure does not try to directly reduce this bias, so it might be large depending on the size of the swapping rate.

3.7.3 Bias in Variance Estimates

Data swapping can be viewed as an imputation procedure. Since the imputation (swapping) cases are a random sample, and because the swapping values are used in the imputation procedure, one should be able to develop an imputation (swapping) procedure that is unbiased or close to unbiased, at least in terms of distribution, but not necessarily correlations. Therefore, distributional biases should be expected to be relatively small.

However, another aspect of an imputation (swapping) procedure is that it necessarily adds a component of variance to the total variance known as the imputation variance. The imputation variance cannot be measured with standard variance methodologies. It can be measured only with specialized procedures like multiple imputations.

The magnitude of the imputation variance is a function of at least two elements: (1) the accuracy of the imputation (swapping) procedure and (2) the amount of imputation (swapping). If the imputation (swapping) procedure is accurate or the amount of imputation (swapping) is small, then the imputation variance may be small. However, with less accurate imputation (swapping) procedures or large imputation (swapping) rates, the imputation variance can be expected to be large.

Since swapped cases are not made available to the public, it becomes impossible for data users to measure the imputation variance. So, the imputation variance can be considered an unknown bias. However, since the *DataSwap* user knows which cases are swapped, the variance can be measured using a bootstrap procedure proposed by Shao and Sitter (1996). The basic approach is to replicate the swapping procedure for each set of bootstrap replicate weights and compute the simple variance of the replicate estimates.

When analyzing the swapping (imputation) variance bias, it should be noted that underestimation and overestimation might both occur within each swapping variable. However, an underestimation of the standard error has the serious potential of contributing to an incorrect rejection of the null hypothesis (e.g., assuming two estimates are different, when in fact the correctly adjusted standard error would not support this conclusion). This bias is largest for the swapping variables that change most often, especially when the perturbation rate is high, but even a small perturbation rate can introduce a large bias in the standard errors. It may be possible to reduce this bias by a more even spread of the actual changes due to the swaps across the swapping variables (as in the Balanced swap approach in *DataSwap*). In this case, it may be possible to maintain relatively large swapping rates.

3.8 Interpretation of the Data Utility Measures

When using statistical disclosure control approaches, there is a dual objective of reducing disclosure risk while maintaining data quality. When using *DataSwap* and its data utility measures, it is assumed that the first objective -- disclosure risk reduction -- is satisfied through the swapping rate assigned by the DRB. Since the data utility measures in *DataSwap* provide the level of distortion between the original data and the swapped data, they can be used to address the second objective of maintaining data quality.

In this section we explain the characteristics and limitations of the measures so that users may appropriately use them to evaluate various swapping scenarios. *DataSwap* includes data utility measures to help the user evaluate the impact of swapping on the following:

1. The weight distribution over tabulations of the swapping variables;
2. Selected pairwise associations; and,
3. Coefficients in regression models of the key output variables on the swapping variables.

For all measures, large values imply a reduction in data quality.

The objective of having these measures built into the program is to help users select the best swapping result from several replicates generated by using different seeds for target selection (with all other parameters unchanged). However, some of the measures are comparable when other parameters are changed as well. Note, however, that the parameters have been approved by the DRB as they appear in the Disclosure Analysis Plan, and therefore, it is important to notify the DRB of any alternative scenarios being considered since parameters may not be changed without further approval.

Users are cautioned not to compare data utility measures from different datasets, as each dataset is unique in its sensitivity to data swapping. The following paragraphs describe the utility measures in more detail including their comparability across runs of the *DataSwap* program on a given dataset.

3.8.1 Utility Measures for Tabulations of Swapping Variables

Multiple calculations of Hellinger’s Distance (HD) are used to determine the data quality of a swapped dataset in terms of the change in weight distributions across swapping variables. *DataSwap* calculates this measure for the weight distribution over the cross-tabulation of all the swapping variables and also separately for each individual swapping variable.

The full application of the HD measure (i.e. including all swapping cells) emphasizes differences in small cells. If the cross-tabulation of the swapping variables results in only a few cases in some cells, the differences between the weighted data in those cells before and after swapping may be dominated by the swapping impact on one or only a few records. In that event, the HD measure may underestimate the data utility since the one cell with few changes, is dominating the calculation. However, all the data are considered in the full application.

The restricted HD measure (i.e., excluding cells below the user's tolerance level) allows the user to examine the data utility after swapping without the impact of small cells. The user should place consideration on the restricted HD application, over the full HD application, when comparing replicated runs.

However, if a swapping scenario has a large number of cells relative to the sample size, then it may be more likely that there are cells with sizes below the user's tolerance level. In that event, the restricted HD measure (excluding small cells), may be based on relatively few cells, and be therefore less stable.

Users should also consider the HD measures for the individual swapping variables to ensure that a particular run does not result in a large amount of utility loss for a specific variable.

The HD measure may vary across each data run if specifications of any of the following parameters are changed: SEED, RATE, MOS, STRATUM, SORTVARS, SWAPMETH, SWAPVARS, BOUNDARY, BIASVAR, WGT, IMPUTE, MISSINGDEF, and TOLFLAG.

The nature of the HD calculation permits direct comparison across data runs with only the SEED parameter changed. We recommend not comparing the HD measures across swapping scenarios (i.e., runs with more than the SEED parameter changed) since each scenario will have a different number of small swapping cells, which impacts the values of the HD measures. Comparison of runs with more parameters changed is possible with all other *DataSwap* utility measures.

3.8.2 Utility Measures for Pairwise Associations

While the HD measures focus on the change in the weight distributions between the original and swapped datasets among the swapping variables only, the measures based on the pairwise associations may be used to evaluate how swapping influences the relationship between these and other important key variables. The three data utility measures of pairwise association in *DataSwap* are based on the Pearson correlations displayed in the *DataSwap* output, the Pearson contingency coefficient and Cramer's V.

All three measures incorporate the pairwise associations between the BOUNDARY, SWAPVARS, KEYVARS, and KEYOUT variables. These allow the user to examine the degree to which the associations differ between the original and swapped datasets. The measures are based on only non-zero differences between the pairwise associations calculated for the original and swapped data. Since all the calculations are average deviations relative to the before swapping standard error (using the number of pairwise associations included), they may be compared across swapping scenarios with more than just the SEED parameter changed.

The utility measure based on the Pearson correlations is calculated as the average absolute deviation between the weighted before and after swapping correlations, relative to the standard errors of the correlations prior to swapping. Since the Pearson correlation is appropriate only when both variables are ordinal, any nominal variables specified in the printed pairwise correlations are converted to dummy variables. The impact on the resulting data utility measure is that more correlations may potentially be included in a calculation for a swapping scenario with more nominal variables. Further, differences in correlations for individual levels from the same nominal variable may dominate a calculation. For this reason the users are cautioned to be certain that the BOUNDARY_T, SWAPVARS_T, and KEYVARS_T are appropriately specified.

The data utility measures based on the Pearson contingency coefficient and Cramer’s V are both calculated as the average absolute deviation of the estimates between the original and swapped data, relative to the values of the correlations prior to swapping. Since these measures treat all variables as nominal, the variables are used in their original form (i.e., dummy variables are not needed for this computation of the Pearson contingency coefficient and Cramer’s V).

All data utility measures based on pairwise associations may vary for a given dataset if specifications of any of the following parameters are changed: SEED, RATE, MOS, STRATUM, SORTVARS, SWAPMETH, SWAPVARS, BOUNDARY, BIASVAR, WGT, IMPUTE, MISSINGDEF, KEYVARS, and KEYOUT. As stated above, the measure based on Pearson correlation will further be sensitive to the specification of BOUNDARY_T, SWAPVARS_T, and KEYVARS_T.

The values of data utility measures based on pairwise associations depend on the values of the auxiliary key variables – variables not directly swapped, but whose relationships with the swapped variables are affected by the procedure. For this reason, it is important that all key variables are specified with care as their association with the swapping and boundary variables will be used to measure the data utility of the swapped dataset.

3.8.3 Utility Measures for Regression Coefficients

The regression coefficient utility measure makes use of the dummy variables created for any nominal swapping variables since linear regression is appropriate only when all variables lie on an ordinal scale. As a result, if the default and user specified regression models contain several nominal swapping variables, more coefficients will be in the models and incorporated into the utility measures than if they contained same number of ordinal variables. Further, differences in coefficients for individual levels associated with nominal variables may dominate a calculation. For this reason the users are cautioned to be certain that the BOUNDARY_T, SWAPVARS_T, and KEYVARS_T are appropriately specified. Since the calculation is standardized, they may be compared across swapping scenarios with more than just the SEED parameter changed.

All data utility measures based on the regression coefficients may vary for a given dataset if specifications of any of the following parameters are changed: SEED, RATE, MOS, STRATUM, SORTVARS, SWAPMETH, SWAPVARS, BOUNDARY, BIASVAR, WGT, IMPUTE, MISSINGDEF, KEYVARS, KEYOUT, BOUNDARY_T, SWAPVARS_T, and KEYVARS_T.

4. EXAMPLES

The examples in this section are presented to illustrate one without hard boundaries and one with a hard boundary restriction. The input dataset (EXAMPLEDATA) is a subset of the 1992 National Adult Literacy Study (NALS) prison study public-use microdata file. The values of the variance unit and variance stratum for this dataset are not those created for that release. The SAS dataset is called EXAMPLEDATA, and it has 20 variables and 182 records. The name, description, values, and value labels for each of the 20 variables are given in Table 4-1.

Table 4-1. Variable Information for example dataset

Variables	Description	Possible values
BIB1201	Ever been in program to improve basic skills?	1 (Yes), 2 (No)
BIC0501	Ever been placed on probation?	1 (Yes), 2 (No)
BID0101	Do you have work assignments inside or outside?	1 (Yes), 2 (No)
BIE0601	How often write letters/memos in English?	1 (Yes), 2 (No)
BORNUSA	Born in USA?	1 (Yes), 2 (No)
CASEID	Identification No.	ID from “90110104” to “93210309”
CENREG	Census region	1 (Northeast), 2 (Midwest), 3 (South), 4 (West)
DAGE	Age derived from date of birth	values ranging from 17 to 63
DAGE3	Derived age with three categories	1 (DAGE<30), 2 (30<=DAGE<50), 3 (DAGE>=50)
DIC0401	Derived years since admission	values ranging from 0.08 to 17.6
DRACE3	Derived Race/ethnicity with three categories	1 (Hispanic); 2 (NH Black); 3 (Other)
EDUC3	Recoded highest education level with three categories	1:less than high school, 2: high school , 3: >high school
EDUC_DET	Detailed highest level of education received	values ranging from 2 to 11
GENDER	Gender	1 (male), 2 (female)
RATE	Swapping rate	User loads into the file
RiskStratum	Risk stratum from risk assessment	0-4
SCORE	Average literacy score	values ranging from 13 to 400
VARSTRAT	Variance stratum	values ranging from 1 to 91
VARUNIT	Variance unit	1 or 2
WEIGHT	Full sample weight	continuous with values ranging from 136 to 1788

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Adult Literacy Study, 1992 Public-use File.

The first *DataSwap* example (described in Section 4.1) illustrates several features including a stratified random selection of target records, swapping among linked variables conditional on primary variables being swapped, and displaying the standard output. The second example (described in Section 4.2) illustrates an additional feature of *DataSwap* by swapping data within hard boundaries. In both examples, we illustrate the ability to create multiple runs with one specification of parameters. The results from the first run of the two examples (SEED = 22601) are discussed and summarized in Section 4.3. The output from these runs from the two examples is provided in appendices A and B.

4.1 *DataSwap* Example 1 – No Hard Boundary

DataSwap Example 1 has no restrictions on searching for swapping partners. The example is set up to select a certain percentage of the records as target records (i.e., the perturbation rate is considered two times the swapping rate) in the stratum found to have the highest risk in the previous example (RiskStratum=4) and to target a lower percentage of the records among the other strata (RiskStratum = 0, 1, 2, or 3). To do this, a RATE variable was created within the data file EXAMPLEDATA.

The swapping variables (SWAPVARS) are assigned as DRACE3, EDUC3 and DAGE3. The variable used to reduce the bias from the swapping procedure is DAGE3. The software finds a swapping partner for each target record from nearby cells created by the concatenation of DRACE3, EDUC3 and DAGE3.

The example also illustrates the conditional swapping that occurs using the LINKSWAP parameter. Whenever values of EDUC3 or DAGE3 are changed through swapping, the detailed education (EDUC_DET) and age (DAGE) variables are also swapped.

Finally, the filled-in parameter sheet and entire standard output for the first run of Example 1 are provided in appendix A, showing the variety of relevant output that can be reviewed as a user and the output that can be shared with the DRB chair and members as they review the swapping results.

In the output for *DataSwap* “Results by Pairs” beginning on page A-5, the change flags are shown. The change flags correspond to each swapping variable (SWAPVARS and LINKSWAP variables) and are set equal to one if the corresponding swapping variable has values that changed, otherwise they are set equal to zero.

4.2 *DataSwap* Example 2 – Hard Boundary

The set-up described in *DataSwap* Example 1 also exists in Example 2, except that the BOUNDARY parameter is used to limit the search for swapping partners. To do this, the DRACE3 and EDUC3 are used as the BOUNDARY variables and DAGE3 is the lone SWAPVARS variable. DAGE is assigned as the lone LINKSWAP variable. Therefore, only values of DAGE3 will change and, if DAGE3 changes value, then DAGE will also be swapped. The parameter sheet and standard output for the first run of Example 2 are provided in appendix B.

4.3 Summary of Results

Suppose a risk assessment created the variable RiskStratum. Records with RiskStratum = 4 were identified to have the highest disclosure risk in the dataset. As a result, these were selected at a higher rate for swapping than the other records on the file.

As a review, once the target records have been selected, the algorithm first selects two potential swapping partners with the closest weight from the two adjacent cells next to the cell that contains the target record. When two records are initially selected as potential swapping partners, they form two potential swapping pairs with the target record. The pair with the smallest swapping bias (discussed in Section 3.2.2) is selected as the swapping partner. If the cell is first or last, or if the cell is the first or last within a hard boundary, then only one potential swapping partner is found and it is designated as the swapping partner.

Five pairs of records are selected in both *DataSwap* examples. Both examples contain the same target selected records since the same random seed was specified and the STRATUM and RATE were identical. The search for the swapping partners for both examples is illustrated in Exhibit 4-1. The hard boundaries from Example 2 are identified by the heavier line, while the swapping cells are identified by the set of faint and heavy lines. Between the two examples, only one of the five pairs contains a different swapping partner (PAIR 2).

Focusing on PAIR 2 shown in Exhibit 4-1 for the example with the hard boundaries, the search for the swapping partner for target record (CASEID = 92210603) can go only to the cell below, since the cell above is cut off by the hard boundary. The search finds the partner with the closest weight in the cell below and identifies it as its swapping partner (CASEID = 90520102).

For this example, since the parameter MAXCAT is less than the number of levels of DAGE, the table for DAGE is suppressed from the output tables designated for the DRB. The parameter is present to reduce the amount of output and to include only variables most likely to be considered for reporting tables.

For the example without hard boundaries, the search for the swapping partner for the target case (CASEID=92210603) goes to the cell above and below and identifies one potential swapping partner with the closest weight from each adjacent cell (CASEID = 90311012 from the cell above, CASEID=90520102 from the cell below). This results because there is no hard boundary. The swapping bias is computed for each potential pair and compared; the potential swapping partner with the smallest swapping bias is chosen as the swapping partner (CASEID = 90311012).

The output is shorter for Example 2 since the BOUNDARY variables are not included in output tables; however, they are included in the output showing measures of association. More measures of association are different for Example 1 since more variables were swapped. Reviewing the plots of the measures of association for these runs (see appendix B-3) confirms this. There are more measures in the chart for Example 1, and more of these are further from the 45 degree reference line. However, since those furthest from the reference line have the smallest correlations before and after swapping, the impact on the resulting swapped data is likely minimal.

The charts summarizing the utility measures across all runs for Example 1 and 2 are also provided in appendices A and B, respectively. Reviewing all measures, across all runs, it appears that Run 5 is optimal for both examples. For Example 1, Run 5 appears optimal for the HD measures, the measures for Pearson's Contingency Coefficient and Cramer's V, and the regression utility measure. For Example 2, Run 5 was optimal for the Pearson's Contingency Coefficient and Cramer's V utility measures and one of the most optimal for the regression measure.

CASEID	RiskStratum	DRACE3	EDUC3	DAGE3	WEIGHT	
Removed 6 cases						
90320206	2	1	1		2 546.51	
90110307	2	1	1		2 595.61	
92820304	1	1	1		2 617.20	Pair 4 partner
92510312	3	1	1		2 647.65	
92210709	4	1	1		2 660.34	
Removed 5 cases						
92820224	2	1	1		3 629.26	Pair 4 target
Removed 11 cases						
92810503	2	1	3		2 143.55	
90520116	1	1	3		2 444.50	
90320112	1	1	3		2 540.92	
90311012	3	1	3		2 625.51	Pair 2 partner
Removed 15 cases						
92610107	2	2	1		1 661.28	
91510109	4	2	1		1 663.16	
92210603	3	2	1		1 675.40	Pair 2 target
91210102	1	2	1		1 697.05	
92210712	4	2	1		1 704.54	
92210502	4	2	1		1 717.71	Pair 1 target
92810103	2	2	1		1 764.06	
90710208	1	2	1		1 776.24	
Removed 5 cases						
90520102	1	2	1		2 613.14	Pair 2 partner with HB
91710301	2	2	1		2 769.97	Pair 1 partner
92310515	1	2	1		2 771.93	
90710102	2	2	1		2 795.08	
Removed 50 cases						
92510201	0	3	1		1 609.63	
90310411	0	3	1		1 614.99	
91210220	0	3	1		1 626.20	
92210206	0	3	1		1 648.22	Pair 5 partner
93010121	0	3	1		1 670.21	
Removed 5 cases						
91510309	0	3	1		2 528.74	
91110213	0	3	1		2 578.11	
92010114	0	3	1		2 582.61	
90410304	0	3	1		2 623.96	
93010101	0	3	1		2 653.47	Pair 5 target
91010108	0	3	1		2 668.24	
90810106	0	3	1		2 706.41	
Removed 30 cases						
91010318	0	3	3		1 452.40	
93210315	0	3	3		1 467.80	Pair 3 partner
92310201	0	3	3		1 484.59	
91010101	0	3	3		1 496.21	
90910108	0	3	3		1 520.37	
93110107	0	3	3		1 575.92	
90410114	0	3	3		2 467.78	
92720106	0	3	3		2 470.48	Pair 3 target
92720116	0	3	3		2 470.48	
91710206	0	3	3		2 472.45	
Removed 10 cases						
92410205	0	3	3		3 525.74	

Exhibit 4-1 - Illustration of Swapping Partner Search in Examples 1 and 2

REFERENCES

Dohrmann, S., Krenzke, T., Roey, S., and Russell, J. N. (2009). Evaluating the Impact of Data Swapping Using Global Utility Measures. *Proceedings of the Federal Committee on Statistical Methodology* (www.fcsm.gov/events/index.html).

Gomatam, S., Karr, A., and Sanil, P. (2005). Data swapping as a decision problem. *Journal of Official Statistics* 21:(4) 635-656.

Li, J., Krenzke, T., Brick, M., Judkins, D., and Larsen, M. (2011). Variance estimation for Census Transportation Planning Products with perturbed American Community Survey data. *Proceedings of the Section on Survey Research Methods of the American Statistical Association*.

Lohr, S. (1999). *Sampling:Ddesign and Analysis*, Duxbury Press (North Scituate, MA).

Karr, A., Kohnen, A., Oganian, J., Reiter, J., Sanil, A. (2006). A framework for evaluating the utility of data altered to protect confidentiality. *The American Statistician* 60:(3) 224-232.

Krenzke, T., Roey, S, Dohrmann, S., Mohadjer, L., Haung, W., Seastrom, M., and Kaufman, S. (2006). Tactics for Reducing the Risk of Disclosure Using the NCES *DataSwap* Software. *American Statistical Association, Proceedings of the Survey Research Methods Section*.

Särndal C. and Swensson B. and Wretman J. (1992), *Model Assisted Survey Sampling*. Springer-Verlag, New York.

Seastrom, M., Kaufman, S., and Roey, S. (2005). *Do Disclosure Controls to Protect Confidentiality Degrade the Quality of the Data? American Statistical Association, Proceedings of the Government Statistics Section*.

Shao J. and Sitter, R.R. (1996). Bootstrap for imputed survey data, *Journal of the American Statistical Association,* 91: 1278-1288.

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

APPENDIX A

Example 1 Parameter Sheet and Standard Output – No Hard Boundary

Appendix A-1. *DataSwap* Example 1 parameter sheet – no hard boundary

Data Swapping

3.2

Parameter specification form

Parameter		*	Entry	Default	Description
Input	**DATA=**	**R**	INFLE.EXAMPLEDATA		input file
Controllers	**SEED=**	**O**	22601 345 76 98 239	**0**	random seed
	ID=	**R**	CASEID		case identification - a single variable
	VARSTRAT=	**O**	VARSTRAT		Variance Stratum
	VARUNIT=	**O**	VARUNIT		Variance Unit
Sampling	**RATE=**	**R**	RATE		swapping rate - a single variable or a number
Controllers	**MOS=**	**O**		**1**	measure of size - a single variable
	STRATUM=	**O**	RiskStratum	**1**	stratum - a single variable
	SORTVARS=	**O**		**BOUNDARY|| SWAPVARS**[1]	sort order for non-certainty selection - a list of variables
	SWAPMETH=	**O**	1	**2**	swapping method – 1:original 2:balanced
	SWAPVARS=	**R**	DRACE3 EDUC3 DAGE3		primary swap variables – a list of variables
	SWAPVARS_T=	**O**	N O O	**O for all**	SWAPVARS variable type: Ordinal (O) Nominal (N)
	BOUNDARY=	**O**			hard boundary - a list of variables
Swap	**BOUNDARY_T=**	**O**		**O for all**	BOUNDARY variable type: Ordinal (O) Nominal (N)
Controllers	**BIASVAR=**	**O**	DAGE3	**the right- most var in SWAPVARS**	variable for calculating bias - a single variable (must be numeric)
	WGT =	**R**	WEIGHT		case survey weight - a single variable
	LINKSWAP=	**O**	NULL#EDUC_DET#DAGE		linked swap variables - a list of variables separated by #
	IMPUTE=	**O**	Y	**y**	impute if swapvars have missing values.
	MISSINGDEF=	**O**	NULL#NULL#NULL	**' ' or .**	values to be defined as missing – a list of values separated by #
	OUT=	**R**	C:\DataSwap\Data\example1		dataset name for swapped result
	KEYOUT=	**O**	SCORE		key outcome continuous variable for weighted means and correlations (must be numeric)
Output	**KEYVARS=**	**O**	BIB1201 BIC0501 BID0101 BIE0601 BORNUSA CENREG DAGE DIC0401 EDUC_DET GENDER		key variables for output tables and correlations
Controllers	**KEYVARS_T=**	**O**	N N N N N N O O O N	**O for all**	KEYVARS variable type: Ordinal (O) Nominal (N)
	MODELS=	**O**	SCORE : CENREG EDUC_DET DAGE GENDER		a list of models separated by # exp. X:Y Z#W : V1 V2
	MODELCLASS=	**O**	CENREG GENDER		a list of class variables separated by # exp. Y Z#V1
	TOLFLAG=	**O**	0.1#45#1.96#1.1	**0.10#45#1.96#1.1**	tolerance measure for flagging outliers
	MAXCAT=	**O**	20	**20**	Maximum # of categories for DRB table variables
	LISTPAIR=	**O**	S#1	**S#1.0**	Subset listing for swapping pairs

* O: Optional R: Required

[1] The double vertical bar denotes the concatenation of the list of BOUNDARY and SWAPVARS variables.

Version 3.3, December 2015

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary

```
DATASWAP   EXAMPLE1.SAS                                        09:22 Wednesday, December 23, 2015   1
DATASWAP REPORT
THE INFORMATION PAGE

DATASWAP MACRO VERSION #:                          Version 3.3

INPUT DATA SET:                                    INFLE.EXAMPLEDATA

OUTPUT DATA SET:                                   OUT.EXAMPLE1_Run1

OBSERVATIONS IN INPUT DATA SET:                    182

RANDOM SEED:                                       22061

CASE IDENTIFICATION:                               CASEID

MEASURE OF SIZE:                                   1 (Default)

STRATUM:                                           RISKSTRATUM

DESIRED SWAPPING RATE:                             RATE

NON-CERTAINTY SELECTION ORDER:                     NOT SPECIFIED

SWAP CELL DEFINITION:                              DRACE3 EDUC3 DAGE3

TOTAL NUMBER OF CELLS:                             23

HARD BOUNDARY:                                     NOT SPECIFIED

BOUNDARY VARIABLE TYPE:                            NOT APPLICABLE

PRIMARY SWAP VARIABLE(S):                          DRACE3 EDUC3 DAGE3

SWAPVARS VARIABLE TYPE:                            N O O

LINKED SWAP VARIABLE(S):                           NULL#EDUC_DET#DAGE

BIAS VARIABLE:                                     DAGE3

CASE WEIGHT:                                       WEIGHT

VARIANCE STRATUM:                                  VARSTRAT

VARIANCE UNIT:                                     VARUNIT

IMPUTE OPTION:                                     Yes (Default)

MISSINGDEF:                                        NULL#NULL#NULL

SWAPPING METHOD:                                   1 (Standard)
```

```
KEY OUTPUT VARIABLE:                               SCORE

OTHER KEY VARIABLES:                               BIB1201 BIC0501 BID0101 BIE0601 BORNUSA CENREG
                                                   DAGE DIC0401 EDUC_DET GENDER

KEYVARS VARIABLE TYPE:                             N N N N N N O O O N

USER-SPECIFIED MODELS:                             SCORE:CENREG EDUC_DET DAGE GENDER

CLASS VARIABLES IN MODELS:                         CENREG GENDER

TOLERANCE FLAG:                                    0.1#45#1.96#1.1 (Default)

SWAPPING PAIR OUTPUT CONTROL:                      S#1

MAXIMUM NUM. OF CATEGORIES FOR DRB TABLE VARIABLES:20 (Default)

TOTAL NUMBER OF ITERATIONS:                        1
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                        09:22 Wednesday, December 23, 2015   5
User Only Output
DataSwap Results by Pairs – [1st Pair Displayed Only]

Pair=1
                                      Recoded
                                      highest
                      Derived        education    Derived age
                  Race/ethnicity      level -      category -                 Age derived
                   - 1:Hispanic,    1:less HS,       1:<30,       Highest      fr date of
Identification      2:NH Black,       2:=HS,         2:<50,       level of      birth or     Literature
     no.              3:Other         3: >HS        3: >=50      education      Screener        score

   91510317              1               3             2             7             38          246.217
   90320112              2               1             1             3             29          295.805



  Ever been                     Nay work       How often
  in pgm to      Ever been    assignments        write        if born in                Derived
   improve       placed on     inside or     letters/memos    USA 1=Yes,    Census    years since
basic skills    probation?      outside?        in Engl?         2=No       region     admission

      1              1             1               1               1           3          1.33
      1              1             1               3               2           4          4.17



                       Initial     Change      Change      Change      Change      Change
Gender     Final     Risk: Risk    flag for    flag for    flag for    flag for    flag for
 (sex)     weight      Stratum     DRACE3      EDUC3       DAGE3      EDUC_DET      DAGE

   2      572.090         4           1           1           1           1           1
   1      540.924         4           1           1           1           1           1
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                                                  09:22 Wednesday, December 23, 2015   8
User Only Output
Aggregated Level -- Changes to Swapped Variable SWAPVARS=DRACE3

    Derived                                                               Estimated   Estimated
Race/ethnicity            Unweighted   Unweighted   Weighted   Weighted   Standard     Standard      Ratio of
 - 1:Hispanic,             Percents     Percents    Percents   Percents    Errors       Errors       Estimated
  2:NH Black,    Sample     Before       After       Before     After      Before       After     Standard Errors
    3:Other       Size     Swapping     Swapping    Swapping   Swapping   Swapping     Swapping   (After/Before)

       1           32      17.5824      17.5824     16.8029    16.8573      2.4978     2.49841       1.000237
       2           79      43.4066      43.4066     47.4255    47.3711      4.0623     4.06268       1.000090
       3           71      39.0110      39.0110     35.7716    35.7716      3.6551     3.65508       1.000000







* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
@ denotes SE ratios exceeds 1.1 and sample size exceeds 45
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                    09:22 Wednesday, December 23, 2015   9
User Only Output
Aggregated Level -- Changes to Swapped Variable SWAPVARS=DRACE3

                                                                         Ratio of Estimated
                          Before/After    Weighted       Estimated        Standard Errors
DRACE3    N     KEYOUT      Swapping        Mean      Standard Errors      (After/Before)

  1       32    SCORE        Before       200.6773        13.1005             1.021108
                             After        197.9714        13.3770
  2       79    SCORE        Before       231.1548         6.3947             1.012102
                             After        232.1527         6.4721
  3       71    SCORE        Before       261.7241         6.8832             1.000000
                             After        261.7241         6.8832

* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
@ denotes SE ratios exceeds 1.1 and sample size exceeds 45
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                                                   09:22 Wednesday, December 23, 2015  10
User Only Output
Aggregated Level -- Changes to Swapped Variable SWAPVARS=EDUC3

  Recoded
  highest
 education                                                                  Estimated    Estimated
  level -               Unweighted    Unweighted    Weighted    Weighted    Standard      Standard       Ratio of
1:less HS,               Percents      Percents     Percents    Percents     Errors        Errors        Estimated
  2:=HS,      Sample      Before        After        Before      After       Before        After      Standard Errors
  3: >HS       Size      Swapping      Swapping     Swapping    Swapping    Swapping      Swapping    (After/Before)

     1          87       47.8022       47.8022      49.6277     49.5733       4.0674      4.06777        1.000089
     2          53       29.1209       29.1209      31.6961     31.6961       3.7058      3.70579        1.000000
     3          42       23.0769       23.0769      18.6762     18.7306       2.7485      2.74905        1.000196

* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
@ denotes SE ratios exceeds 1.1 and sample size exceeds 45
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                 09:22 Wednesday, December 23, 2015  11
User Only Output
Aggregated Level -- Changes to Swapped Variable SWAPVARS=EDUC3

                                                                        Ratio of Estimated
                         Before/After    Weighted       Estimated        Standard Errors
EDUC3    N      KEYOUT     Swapping         Mean      Standard Errors      (After/Before)

  1      87     SCORE        Before      210.8865        7.55845              1.007563
                             After       211.8178        7.61561
  2      53     SCORE        Before      253.2086        7.51418              1.000000
                             After       253.2086        7.51418
  3      42     SCORE        Before      278.7154        5.69626              1.103795
                             After       276.0536        6.28750
```

```
* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
@ denotes SE ratios exceeds 1.1 and sample size exceeds 45
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                                               09:22 Wednesday, December 23, 2015  12
User Only Output
Aggregated Level -- Changes to Swapped Variable LINKSWAP=EDUC_DET

                                                                           Estimated    Estimated
                       Unweighted    Unweighted    Weighted    Weighted    Standard      Standard       Ratio of
 Highest                Percents      Percents     Percents    Percents     Errors        Errors        Estimated
 level of    Sample      Before        After        Before      After       Before        After      Standard Errors
education     Size      Swapping      Swapping     Swapping    Swapping    Swapping      Swapping    (After/Before)

     2         30       16.4835       16.4835      16.8840     16.8840       2.6985       2.69853        1.000000
     3         57       31.3187       31.3187      32.7437     32.6893       3.8948       3.89521        1.000097
     4         32       17.5824       17.5824      19.5312     19.5312       3.1976       3.19757        1.000000
     5         21       11.5385       11.5385      12.1649     12.1649       2.6270       2.62702        1.000000
     6          6        3.2967        3.2967       2.9218      2.9218       1.2773       1.27730        1.000000
     7         19       10.4396       10.4396       8.0742      8.0995       1.8137       1.81384        1.000097
     8          7        3.8462        3.8462       3.1669      3.1669       1.2179       1.21788        1.000000
     9          8        4.3956        4.3956       3.2857      3.3148       1.1875       1.18789        1.000299
    10          2        1.0989        1.0989       1.2275      1.2275       0.8711       0.87107        1.000000







* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
@ denotes SE ratios exceeds 1.1 and sample size exceeds 45
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                                                   09:22 Wednesday, December 23, 2015  13
User Only Output
Aggregated Level -- Changes to Swapped Variable LINKSWAP=EDUC_DET

                                                                           Ratio of Estimated
                            Before/After    Weighted       Estimated        Standard Errors
EDUC_DET    N     KEYOUT      Swapping        Mean      Standard Errors      (After/Before)

    2       30    SCORE        Before       177.4834        13.4480             1.000000
                               After        177.4834        13.4480
    3       57    SCORE        Before       228.1105         8.3165             1.014900
                               After        229.5515         8.4404
    4       32    SCORE        Before       266.0061         6.9660             1.000000
                               After        266.0061         6.9660
    5       21    SCORE        Before       232.6617        12.8376             1.000000
                               After        232.6617        12.8376
    6        6    SCORE        Before       273.4216        16.3528             1.000000
                               After        273.4216        16.3528
    7       19    SCORE        Before       272.7863         8.6800             1.049798
                               After        270.0131         9.1123
    8        7    SCORE        Before       293.3091        12.6352             1.000000
                               After        293.3091        12.6352
    9        8    SCORE        Before       296.5311         8.7201             1.386565
                               After        288.1554        12.0910
   10        2    SCORE        Before       244.9769         4.1836             1.000000
                               After        244.9769         4.1836






* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
@ denotes SE ratios exceeds 1.1 and sample size exceeds 45
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                                           09:22 Wednesday, December 23, 2015  14
User Only Output
Aggregated Level -- Changes to Swapped Variable SWAPVARS=DAGE3

Derived age                                                                  Estimated    Estimated
 category -             Unweighted    Unweighted    Weighted    Weighted    Standard      Standard       Ratio of
   1:<30,                Percents      Percents     Percents    Percents     Errors        Errors        Estimated
   2:<50,      Sample     Before        After        Before      After       Before        After      Standard Errors
  3: >=50       Size     Swapping      Swapping     Swapping    Swapping    Swapping      Swapping    (After/Before)

     1           87      47.8022       47.8022      44.6327     44.5885       3.9701      3.97038        1.000062
     2           88      48.3516       48.3516      51.2125     51.2566       4.1759      4.17615        1.000056
     3            7       3.8462        3.8462       4.1548      4.1548       1.6172      1.61717        1.000000
```

```
* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
@ denotes SE ratios exceeds 1.1 and sample size exceeds 45
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                  09:22 Wednesday, December 23, 2015  15
User Only Output
Aggregated Level -- Changes to Swapped Variable SWAPVARS=DAGE3

                                                                       Ratio of Estimated
                         Before/After    Weighted       Estimated        Standard Errors
DAGE3    N     KEYOUT      Swapping        Mean      Standard Errors      (After/Before)

  1      87    SCORE         Before       243.6602         5.2952             1.001603
                             After        243.9601         5.3037
  2      88    SCORE         Before       230.8760         7.8594             1.000505
                             After        230.6261         7.8634
  3       7    SCORE         Before       240.1879        18.6124             1.000000
                             After        240.1879        18.6124
```

```
* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
@ denotes SE ratios exceeds 1.1 and sample size exceeds 45
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                                      09:22 Wednesday, December 23, 2015  16
User Only Output
Aggregated Level -- Changes to Swapped Variable LINKSWAP=DAGE
                                                                             Estimated    Estimated
Age derived              Unweighted    Unweighted    Weighted    Weighted    Standard      Standard       Ratio of
 fr date of               Percents      Percents     Percents    Percents     Errors        Errors        Estimated
  birth or     Sample      Before        After        Before      After       Before        After      Standard Errors
  Screener      Size      Swapping      Swapping     Swapping    Swapping    Swapping      Swapping    (After/Before)

     18           2        1.0989        1.0989       0.8551      0.8551       0.6065      0.60647        1.000000
     19           7        3.8462        3.8462       3.6755      3.6755       1.4071      1.40711        1.000000
     20           9        4.9451        4.9451       4.7582      4.7582       1.6123      1.61229        1.000000
     21           6        3.2967        3.2967       3.1320      3.1320       1.2906      1.29060        1.000000
     22           6        3.2967        3.2967       2.7017      2.6727       1.1228      1.12314        1.000335
     23           8        4.3956        4.3956       4.3434      4.3261       1.5572      1.55727        1.000061
     24           5        2.7473        2.7473       2.2124      2.2124       0.7149      0.71494        1.000000
     25           7        3.8462        3.8462       3.5267      3.5267       1.3610      1.36102        1.000000
     26          12        6.5934        6.5934       6.6062      6.6111       1.6719      1.67192        1.000004
     27           7        3.8462        3.8462       3.6243      3.6470       1.3978      1.39798        1.000132
     28          12        6.5934        6.5934       6.3024      6.3024       1.9103      1.91028        1.000000
     29           6        3.2967        3.2967       2.8946      2.8693       1.2185      1.21879        1.000216
     30           6        3.2967        3.2967       3.5710      3.5710       1.4753      1.47526        1.000000
     31           7        3.8462        3.8462       4.6080      4.6080       1.7765      1.77646        1.000000
     32           6        3.2967        3.2967       3.1655      3.1655       1.3121      1.31206        1.000000
     33           7        3.8462        3.8462       4.4769      4.4769       1.7080      1.70800        1.000000
     34           6        3.2967        3.2967       3.7178      3.7178       1.5423      1.54231        1.000000
     35           8        4.3956        4.3956       4.5886      4.5886       1.7600      1.75999        1.000000
     36           2        1.0989        1.0989       1.4113      1.4113       0.9923      0.99233        1.000000
     37           7        3.8462        3.8462       4.4246      4.4537       1.6928      1.69307        1.000147
     38           5        2.7473        2.7473       3.1325      3.1578       1.4637      1.46394        1.000150
     39           4        2.1978        2.1978       2.0119      2.0291       1.0192      1.01938        1.000143
     40           5        2.7473        2.7473       2.5779      2.5779       1.1672      1.16723        1.000000
     41           9        4.9451        4.9451       4.8760      4.8760       1.7321      1.73212        1.000000
     42           4        2.1978        2.1978       2.2923      2.2923       1.1740      1.17401        1.000000
     43           3        1.6484        1.6484       1.6132      1.6132       0.9424      0.94240        1.000000
     44           1        0.5495        0.5495       0.6255      0.6207       0.6246      0.62463        1.000030
     45           3        1.6484        1.6484       1.4325      1.4098       0.8389      0.83924        1.000366
     47           1        0.5495        0.5495       0.6541      0.6541       0.6527      0.65270        1.000000
     48           1        0.5495        0.5495       0.5428      0.5428       0.5424      0.54236        1.000000
     49           3        1.6484        1.6484       1.4902      1.4902       0.8745      0.87450        1.000000
     50           1        0.5495        0.5495       0.4981      0.4981       0.4981      0.49813        1.000000
     53          2       1.0989        1.0989       1.3754      1.3754       0.9746      0.97460        1.000000
     54          1       0.5495        0.5495       0.8560      0.8560       0.8553      0.85531        1.000000
     56          1       0.5495        0.5495       0.5112      0.5112       0.5120      0.51203        1.000000
     58          1       0.5495        0.5495       0.4271      0.4271       0.4281      0.42809        1.000000
     68          1       0.5495        0.5495       0.4872      0.4872       0.4871      0.48710        1.000000
```

```
* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
@ denotes SE ratios exceeds 1.1 and sample size exceeds 45
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                    09:22 Wednesday, December 23, 2015  17
User Only Output
Aggregated Level -- Changes to Swapped Variable LINKSWAP=DAGE

                                                                       Ratio of Estimated
                        Before/After    Weighted       Estimated        Standard Errors
DAGE    N    KEYOUT      Swapping         Mean      Standard Errors      (After/Before)

 18      2    SCORE        Before       270.6371        13.7863             1.000000
                           After        270.6371        13.7863
 19      7    SCORE        Before       238.6627        18.1733             1.000000
                           After        238.6627        18.1733
 20      9    SCORE        Before       244.4130        19.5409             1.000000
                           After        244.4130        19.5409
 21      6    SCORE        Before       251.8129        25.1460             1.000000
                           After        251.8129        25.1460
 22      6    SCORE        Before       250.3557        18.1556             1.138635
                           After        260.2416        20.6726
 23      8    SCORE        Before       248.9454        13.7337             1.188903
                           After        240.1141        16.3281
 24      5    SCORE        Before       252.8639        27.9424             1.000000
                           After        252.8639        27.9424
 25      7    SCORE        Before       233.2826        15.1436             1.000000
                           After        233.2826        15.1436
 26     12    SCORE        Before       232.9787        12.1147             1.061938
                           After        228.6493        12.8650
 27      7    SCORE        Before       217.0387        27.5774             1.052442
                           After        226.0862        29.0236
 28     12    SCORE        Before       252.0289        12.3718             1.000000
                           After        252.0289        12.3718
 29      6    SCORE        Before       262.8960        16.1375             1.109099
                           After        270.6370        17.8981
 30      6    SCORE        Before       207.7303        36.5747             1.000000
                           After        207.7303        36.5747
 31      7    SCORE        Before       242.1075        19.9642             1.000000
                           After        242.1075        19.9642
 32      6    SCORE        Before       281.7871        20.2525             1.000000
                           After        281.7871        20.2525
 33      7    SCORE        Before       257.2944        14.0683             1.000000
                           After        257.2944        14.0683
 34      6    SCORE        Before       204.8023        26.3751             1.000000
                           After        204.8023        26.3751
 35      8    SCORE        Before       197.9061        27.5678             1.000000
                           After        197.9061        27.5678
 36      2    SCORE        Before       223.3839        19.5497             1.000000
                           After        223.3839        19.5497
 37      7    SCORE        Before       277.1285         9.9739             1.172579
                           After        271.0212        11.6952
 38      5    SCORE        Before       253.0645        19.3166             1.062843
                           After        246.1096        20.5305
 39      4    SCORE        Before       258.7778       21.3196              1.331556
                           After        277.5226       28.3882
 40      5    SCORE        Before       222.2087       19.6368              1.000000
                           After        222.2087       19.6368
 41      9    SCORE        Before       151.2064       29.4209              1.000000
                           After        151.2064       29.4209
 42      4    SCORE        Before       276.4834        7.8304              1.000000
                           After        276.4834        7.8304
 43      3    SCORE        Before       278.6246       13.7719              1.000000
                           After        278.6246       13.7719
 44      1    SCORE        Before       209.1415        0.0000
                           After        255.0729       45.9314
 45      3    SCORE        Before       264.1727        9.8261              2.512344
                           After        241.5261       24.6864
 47      1    SCORE        Before       220.8250        0.0000
                           After        220.8250        0.0000
 48      1    SCORE        Before        61.2798        0.0000
                           After         61.2798        0.0000
 49      3    SCORE        Before       213.2606       39.3590              1.000000
                           After        213.2606       39.3590
 50      1    SCORE        Before       266.6619        0.0000
                           After        266.6619        0.0000
 53      2    SCORE        Before       264.7618       44.2554              1.000000
                           After        264.7618       44.2554
 54      1    SCORE        Before       261.1270        0.0000
                           After        261.1270           0
 56      1    SCORE        Before       195.0395           0
                           After        195.0395           0
 58      1    SCORE        Before       203.0768           0
                           After        203.0768           0
 68      1    SCORE        Before       186.8552           0
                           After        186.8552           0
```

* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45

~ denotes value was zero before swapping and non-zero after swapping

@ denotes SE ratios exceeds 1.1 and sample size exceeds 45

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                    09:22 Wednesday, December 23, 2015  19
User Only Output
Weighted Percents Sorted by Relative Difference in SWAPVARS and LINKSWAP

                               Weighted    Weighted
                               Percents    Percents     Absolute
                     Sample     Before      After       Relative
Variable    Value     Size     Swapping    Swapping    Difference


DAGE         45         3        1.4325      1.4098     0.015842
DAGE         22         6        2.7017      2.6727     0.010757
EDUC_DET     9          8        3.2857      3.3148     0.008845
DAGE         29         6        2.8946      2.8693     0.008746
DAGE         39         4        2.0119      2.0291     0.008563
DAGE         38         5        3.1325      3.1578     0.008082
DAGE         44         1        0.6255      0.6207     0.007678
DAGE         37         7        4.4246      4.4537     0.006569
DAGE         27         7        3.6243      3.6470     0.006261
DAGE         23         8        4.3434      4.3261     0.003967
DRACE3       1         32       16.8029     16.8573     0.003236
EDUC_DET     7         19        8.0742      8.0995     0.003135
EDUC3        3         42       18.6762     18.7306     0.002912
EDUC_DET     3         57       32.7437     32.6893     0.001661
DRACE3       2         79       47.4255     47.3711     0.001147
EDUC3        1         87       49.6277     49.5733     0.001096
DAGE3        1         87       44.6327     44.5885     0.000988
DAGE3        2         88       51.2125     51.2566     0.000861
DAGE         26        12        6.6062      6.6111     0.000727
EDUC_DET     2         30       16.8840     16.8840     0.000000
EDUC_DET     4         32       19.5312     19.5312     0.000000
EDUC_DET     10         2        1.2275      1.2275     0.000000
EDUC_DET     6          6        2.9218      2.9218     0.000000
EDUC_DET     5         21       12.1649     12.1649     0.000000
EDUC_DET     8          7        3.1669      3.1669     0.000000















* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
@ denotes SE ratios exceeds 1.1 and sample size exceeds 45
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                    09:22 Wednesday, December 23, 2015  20
User Only Output
Weighted Means Sorted by Relative Difference in SWAPVARS and LINKSWAP

                                  Weighted         Weighted
                               Mean of SCORE    Mean of SCORE    Absolute
                     Sample        Before           After        Relative
Variable    Value     Size        Swapping         Swapping     Difference


DAGE         44         1         209.142          255.073        0.219619
DAGE         45         3         264.173          241.526        0.085727
DAGE         39         4         258.778          277.523        0.072436
DAGE         27         7         217.039          226.086        0.041686
DAGE         22         6         250.356          260.242        0.039487
DAGE         23         8         248.945          240.114        0.035475
DAGE         29         6         262.896          270.637        0.029445
EDUC_DET     9          8         296.531          288.155        0.028246
DAGE         38         5         253.065          246.110        0.027483
DAGE         37         7         277.129          271.021        0.022038
DAGE         26        12         232.979          228.649        0.018583
DRACE3       1         32         200.677          197.971        0.013484
EDUC_DET     7         19         272.786          270.013        0.010166
EDUC3        3         42         278.715          276.054        0.009550
EDUC_DET     3         57         228.110          229.552        0.006317
EDUC3        1         87         210.886          211.818        0.004416
DRACE3       2         79         231.155          232.153        0.004317
DAGE3        1         87         243.660          243.960        0.001231
DAGE3        2         88         230.876          230.626        0.001082




Only Absolute Relative Differences > 0 are shown
* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                                           09:22 Wednesday, December 23, 2015  23
Supplemental Tables for DRB Members
Aggregated Level -- Changes to Swapped Variable SWAPVARS=DRACE3

    Derived                                                               Estimated   Estimated
Race/ethnicity            Unweighted   Unweighted   Weighted   Weighted   Standard     Standard      Ratio of
 - 1:Hispanic,             Percents     Percents    Percents   Percents    Errors       Errors       Estimated
  2:NH Black,    Sample     Before       After       Before     After      Before       After     Standard Errors
    3:Other       Size     Swapping     Swapping    Swapping   Swapping   Swapping     Swapping   (After/Before)

       1           32      17.5824      17.5824     16.8029    16.8573      2.4978     2.49841       1.000237
       2           79      43.4066      43.4066     47.4255    47.3711      4.0623     4.06268       1.000090
       3           71      39.0110      39.0110     35.7716    35.7716      3.6551     3.65508       1.000000

* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
@ denotes SE ratios exceeds 1.1 and sample size exceeds 45
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                 09:22 Wednesday, December 23, 2015  24
Supplemental Tables for DRB Members
Aggregated Level -- Changes to Swapped Variable SWAPVARS=DRACE3

                                                                     Ratio of Estimated
                          Before/After    Weighted       Estimated        Standard Errors
DRACE3    N     KEYOUT      Swapping        Mean      Standard Errors      (After/Before)

  1       32    SCORE         Before       200.6773        13.1005             1.021108
                              After        197.9714        13.3770
  2       79    SCORE         Before       231.1548         6.3947             1.012102
                              After        232.1527         6.4721
  3       71    SCORE         Before       261.7241         6.8832             1.000000
                              After        261.7241         6.8832
```

```
* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
@ denotes SE ratios exceeds 1.1 and sample size exceeds 45
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                                   09:22 Wednesday, December 23, 2015  25
Supplemental Tables for DRB Members
Aggregated Level -- Changes to Swapped Variable SWAPVARS=EDUC3

  Recoded
  highest
 education                                                                          Estimated    Estimated
  level -             Unweighted    Unweighted    Weighted    Weighted    Standard     Standard       Ratio of
1:less HS,             Percents      Percents     Percents    Percents     Errors       Errors        Estimated
  2:=HS,      Sample    Before        After        Before      After       Before       After      Standard Errors
  3: >HS       Size     Swapping      Swapping     Swapping    Swapping    Swapping     Swapping    (After/Before)

     1          87      47.8022       47.8022      49.6277     49.5733      4.0674      4.06777        1.000089
     2          53      29.1209       29.1209      31.6961     31.6961      3.7058      3.70579        1.000000
     3          42      23.0769       23.0769      18.6762     18.7306      2.7485      2.74905        1.000196













* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
@ denotes SE ratios exceeds 1.1 and sample size exceeds 45
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                    09:22 Wednesday, December 23, 2015  26
Supplemental Tables for DRB Members
Aggregated Level -- Changes to Swapped Variable SWAPVARS=EDUC3

                                                                        Ratio of Estimated
                         Before/After    Weighted       Estimated        Standard Errors
EDUC3    N     KEYOUT      Swapping        Mean      Standard Errors      (After/Before)

  1      87    SCORE        Before       210.8865        7.55845             1.007563
                            After        211.8178        7.61561
  2      53    SCORE        Before       253.2086        7.51418             1.000000
                            After        253.2086        7.51418
  3      42    SCORE        Before       278.7154        5.69626             1.103795
                            After        276.0536        6.28750
```

```
* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
@ denotes SE ratios exceeds 1.1 and sample size exceeds 45
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                                                   09:22 Wednesday, December 23, 2015  27
Supplemental Tables for DRB Members
Aggregated Level -- Changes to Swapped Variables in LINKSWAP=EDUC_DET

                                                                           Estimated    Estimated
                       Unweighted    Unweighted    Weighted    Weighted    Standard      Standard       Ratio of
 Highest                Percents      Percents     Percents    Percents     Errors        Errors        Estimated
 level of    Sample      Before        After        Before      After       Before        After      Standard Errors
education     Size      Swapping      Swapping     Swapping    Swapping    Swapping      Swapping    (After/Before)

     2         30       16.4835       16.4835      16.8840     16.8840       2.6985      2.69853        1.000000
     3         57       31.3187       31.3187      32.7437     32.6893       3.8948      3.89521        1.000097
     4         32       17.5824       17.5824      19.5312     19.5312       3.1976      3.19757        1.000000
     5         21       11.5385       11.5385      12.1649     12.1649       2.6270      2.62702        1.000000
     6          6        3.2967        3.2967       2.9218      2.9218       1.2773      1.27730        1.000000
     7         19       10.4396       10.4396       8.0742      8.0995       1.8137      1.81384        1.000097
     8          7        3.8462        3.8462       3.1669      3.1669       1.2179      1.21788        1.000000
     9          8        4.3956        4.3956       3.2857      3.3148       1.1875      1.18789        1.000299
    10          2        1.0989        1.0989       1.2275      1.2275       0.8711      0.87107        1.000000
```

```
* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
@ denotes SE ratios exceeds 1.1 and sample size exceeds 45
```

Appendix A-2.   *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                 09:22 Wednesday, December 23, 2015  28
Supplemental Tables for DRB Members
Aggregated Level -- Changes to Swapped Variables in LINKSWAP=EDUC_DET

                                                                            Ratio of Estimated
                             Before/After    Weighted       Estimated        Standard Errors
EDUC_DET    N     KEYOUT       Swapping        Mean      Standard Errors      (After/Before)

    2       30    SCORE         Before       177.4834        13.4480              1.000000
                                After        177.4834        13.4480
    3       57    SCORE         Before       228.1105         8.3165              1.014900
                                After        229.5515         8.4404
    4       32    SCORE         Before       266.0061         6.9660              1.000000
                                After        266.0061         6.9660
    5       21    SCORE         Before       232.6617        12.8376              1.000000
                                After        232.6617        12.8376
    6        6    SCORE         Before       273.4216        16.3528              1.000000
                                After        273.4216        16.3528
    7       19    SCORE         Before       272.7863         8.6800              1.049798
                                After        270.0131         9.1123
    8        7    SCORE         Before       293.3091        12.6352              1.000000
                                After        293.3091        12.6352
    9        8    SCORE         Before       296.5311         8.7201              1.386565
                                After        288.1554        12.0910
   10        2    SCORE         Before       244.9769         4.1836              1.000000
                                After        244.9769         4.1836
```

```
* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
@ denotes SE ratios exceeds 1.1 and sample size exceeds 45
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                                                   09:22 Wednesday, December 23, 2015  29
Supplemental Tables for DRB Members
Aggregated Level -- Changes to Swapped Variable SWAPVARS=DAGE3

Derived age                                                                  Estimated    Estimated
 category -              Unweighted    Unweighted    Weighted    Weighted    Standard      Standard       Ratio of
   1:<30,                 Percents      Percents     Percents    Percents     Errors        Errors        Estimated
   2:<50,      Sample      Before        After        Before      After       Before        After      Standard Errors
  3: >=50       Size      Swapping      Swapping     Swapping    Swapping    Swapping      Swapping    (After/Before)

     1           87       47.8022       47.8022      44.6327     44.5885       3.9701      3.97038        1.000062
     2           88       48.3516       48.3516      51.2125     51.2566       4.1759      4.17615        1.000056
     3            7        3.8462        3.8462       4.1548      4.1548       1.6172      1.61717        1.000000
```

```
* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
@ denotes SE ratios exceeds 1.1 and sample size exceeds 45
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                    09:22 Wednesday, December 23, 2015  30
Supplemental Tables for DRB Members
Aggregated Level -- Changes to Swapped Variable SWAPVARS=DAGE3

                                                                        Ratio of Estimated
                         Before/After    Weighted       Estimated        Standard Errors
DAGE3    N     KEYOUT      Swapping        Mean      Standard Errors      (After/Before)

  1      87    SCORE        Before       243.6602         5.2952             1.001603
                            After        243.9601         5.3037
  2      88    SCORE        Before       230.8760         7.8594             1.000505
                            After        230.6261         7.8634
  3       7    SCORE        Before       240.1879        18.6124             1.000000
                            After        240.1879        18.6124
```

```
* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
@ denotes SE ratios exceeds 1.1 and sample size exceeds 45
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                    09:22 Wednesday, December 23, 2015  31
Supplemental Tables for DRB Members
Aggregated Level -- Changes to Swapped Variables in LINKSWAP=DAGE


          Report for DAGE was not generated because it has 37 levels which exceeds MAXCAT=20
                    Increase the value of MAXCAT to allow the report to be printed
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                     09:22 Wednesday, December 23, 2015  32
Supplemental Tables for DRB Members
Weighted Percents Sorted by Relative Difference in SWAPVARS and LINKSWAP

                               Weighted    Weighted
                               Percents    Percents     Absolute
                     Sample     Before      After       Relative
Variable    Value     Size     Swapping    Swapping    Difference

EDUC_DET     9          8        3.2857      3.3148     0.008845
DRACE3       1         32       16.8029     16.8573     0.003236
EDUC_DET     7         19        8.0742      8.0995     0.003135
EDUC3        3         42       18.6762     18.7306     0.002912
EDUC_DET     3         57       32.7437     32.6893     0.001661
DRACE3       2         79       47.4255     47.3711     0.001147
EDUC3        1         87       49.6277     49.5733     0.001096
DAGE3        1         87       44.6327     44.5885     0.000988
DAGE3        2         88       51.2125     51.2566     0.000861
EDUC_DET     2         30       16.8840     16.8840     0.000000
EDUC_DET     4         32       19.5312     19.5312     0.000000
EDUC_DET     10         2        1.2275      1.2275     0.000000
EDUC_DET     6          6        2.9218      2.9218     0.000000
EDUC_DET     5         21       12.1649     12.1649     0.000000
EDUC_DET     8          7        3.1669      3.1669     0.000000
```

```
* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
@ denotes SE ratios exceeds 1.1 and sample size exceeds 45
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                    09:22 Wednesday, December 23, 2015  33
Supplemental Tables for DRB Members
Weighted Means Sorted by Relative Difference in SWAPVARS and LINKSWAP

                                  Weighted         Weighted
                               Mean of SCORE    Mean of SCORE     Absolute
                     Sample       Before            After         Relative
Variable    Value     Size       Swapping         Swapping      Difference


EDUC_DET      9         8         296.531          288.155        0.028246
DRACE3        1        32         200.677          197.971        0.013484
EDUC_DET      7        19         272.786          270.013        0.010166
EDUC3         3        42         278.715          276.054        0.009550
EDUC_DET      3        57         228.110          229.552        0.006317
EDUC3         1        87         210.886          211.818        0.004416
DRACE3        2        79         231.155          232.153        0.004317
DAGE3         1        87         243.660          243.960        0.001231
DAGE3         2        88         230.876          230.626        0.001082
```

```
* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
@ denotes SE ratios exceeds 1.1 and sample size exceeds 45
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                    09:22 Wednesday, December 23, 2015  34
Supplemental Tables for DRB Members
After swapping
Check creation of indicator variables for DRACE3

The FREQ Procedure

                                                                       Cumulative    Cumulative
DRACE3    DRACE3_1    DRACE3_2    DRACE3_3    Frequency     Percent     Frequency      Percent
-----------------------------------------------------------------------------------------------
     1           1           0           0          32       17.58            32        17.58
     2           0           1           0          79       43.41           111        60.99
     3           0           0           1          71       39.01           182       100.00
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                     09:22 Wednesday, December 23, 2015  35
Supplemental Tables for DRB Members
After swapping
Check creation of indicator variables for BIB1201

The FREQ Procedure

                                                               Cumulative    Cumulative
BIB1201    BIB1201_1    BIB1201_2    Frequency     Percent     Frequency      Percent
-------------------------------------------------------------------------------------
      1            1            0           29       15.93            29        15.93
      2            0            1          153       84.07           182       100.00
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                    09:22 Wednesday, December 23, 2015  36
Supplemental Tables for DRB Members
After swapping
Check creation of indicator variables for BIC0501

The FREQ Procedure

                                                              Cumulative    Cumulative
BIC0501    BIC0501_1    BIC0501_2    Frequency     Percent     Frequency      Percent
--------------------------------------------------------------------------------------
      1            1            0         108       59.34           108        59.34
      2            0            1          74       40.66           182       100.00
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                    09:22 Wednesday, December 23, 2015  37
Supplemental Tables for DRB Members
After swapping
Check creation of indicator variables for BID0101

The FREQ Procedure

                                                              Cumulative    Cumulative
BID0101    BID0101_1    BID0101_2    Frequency     Percent     Frequency      Percent
--------------------------------------------------------------------------------------
      1            1            0         128       70.33           128        70.33
      2            0            1          54       29.67           182       100.00
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                                                09:22 Wednesday, December 23, 2015  38
Supplemental Tables for DRB Members
After swapping
Check creation of indicator variables for BIE0601

The FREQ Procedure

                                                                                              Cumulative   Cumulative
BIE0601   BIE0601_1   BIE0601_2   BIE0601_3   BIE0601_4   BIE0601_5   Frequency    Percent    Frequency      Percent
----------------------------------------------------------------------------------------------------------------------
      1           1           0           0           0           0          53      29.12           53        29.12
      2           0           1           0           0           0          66      36.26          119        65.38
      3           0           0           1           0           0          28      15.38          147        80.77
      4           0           0           0           1           0          20      10.99          167        91.76
      5           0           0           0           0           1          15       8.24          182       100.00
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                    09:22 Wednesday, December 23, 2015  39
Supplemental Tables for DRB Members
After swapping
Check creation of indicator variables for BORNUSA

The FREQ Procedure

                                                              Cumulative    Cumulative
BORNUSA    BORNUSA_1    BORNUSA_2    Frequency     Percent     Frequency      Percent
--------------------------------------------------------------------------------------
      1            1            0         162       89.01           162        89.01
      2            0            1          20       10.99           182       100.00
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                                                   09:22 Wednesday, December 23, 2015  40
Supplemental Tables for DRB Members
After swapping
Check creation of indicator variables for CENREG

The FREQ Procedure

                                                                                   Cumulative    Cumulative
CENREG    CENREG_1    CENREG_2    CENREG_3    CENREG_4    Frequency     Percent     Frequency      Percent
-----------------------------------------------------------------------------------------------------------
     1           1           0           0           0          29       15.93            29        15.93
     2           0           1           0           0          44       24.18            73        40.11
     3           0           0           1           0          68       37.36           141        77.47
     4           0           0           0           1          41       22.53           182       100.00
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                               09:22 Wednesday, December 23, 2015  41
Supplemental Tables for DRB Members
After swapping
Check creation of indicator variables for GENDER

The FREQ Procedure

                                                           Cumulative    Cumulative
GENDER    GENDER_1    GENDER_2    Frequency     Percent     Frequency      Percent
-----------------------------------------------------------------------------------
     1           1           0         168       92.31           168        92.31
     2           0           1          14        7.69           182       100.00
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                    09:22 Wednesday, December 23, 2015  42
Supplemental Tables for DRB Members
Unweighted Measures of Association in Swapping Variables and Key Output Variables
Where r(before)-r(after) is not 0

Variable     Variable       Unweighted         Unweighted
1            2            Before Swapping    After Swapping


BIB1201_1    DAGE            -0.00218           -0.03185
BIC0501_1    DAGE            -0.13292           -0.07275
BIC0501_1    EDUC_DET        -0.12277           -0.08996
BID0101_1    DAGE             0.07806            0.09919
BIE0601_1    DAGE            -0.00729           -0.01525
BIE0601_1    EDUC_DET         0.15706            0.14524
BIE0601_2    DAGE            -0.21793           -0.26059
BIE0601_3    DAGE             0.16769            0.15264
BIE0601_3    EDUC_DET         0.00630           -0.02348
BIE0601_4    DAGE             0.05084            0.11449
BIE0601_4    EDUC_DET        -0.10155           -0.05002
BIE0601_5    DAGE             0.11520            0.15029
BORNUSA_1    DAGE             0.03209            0.07837
BORNUSA_1    EDUC_DET         0.01567            0.10155
CENREG_1     DAGE            -0.10601           -0.10931
CENREG_3     DAGE             0.05490            0.10727
CENREG_3     EDUC_DET        -0.09329           -0.03777
CENREG_4     DAGE             0.07954            0.02179
CENREG_4     EDUC_DET         0.00682           -0.05748
DAGE         DIC0401          0.37930            0.40219
DAGE         EDUC_DET         0.08507            0.08979
DAGE         GENDER_1         0.02455            0.00418
DAGE3        BIB1201_1       -0.03308           -0.05948
DAGE3        BIC0501_1       -0.14809           -0.08907
DAGE3        BID0101_1        0.11135            0.13251
DAGE3        BIE0601_2       -0.18069           -0.22089
DAGE3        BIE0601_3        0.11538            0.08859
DAGE3        BIE0601_4        0.05535            0.11714
DAGE3        BIE0601_5        0.09113            0.12628
DAGE3        BORNUSA_1       -0.02445            0.03735
DAGE3        CENREG_3         0.01778            0.07771
DAGE3        CENREG_4         0.04677           -0.02262
DAGE3        DIC0401          0.32468            0.34215
DAGE3        EDUC_DET         0.06754            0.07226
DAGE3        GENDER_1         0.03069           -0.00558
DAGE3        SCORE           -0.08039           -0.08839
DIC0401      EDUC_DET         0.12472            0.13173
DRACE3_1     BIC0501_1       -0.08783           -0.05845
DRACE3_1     BIE0601_3        0.04308            0.00308
DRACE3_1     BIE0601_4        0.02231            0.06847
DRACE3_1     BORNUSA_1       -0.43767           -0.34537
DRACE3_1     CENREG_3        -0.05836            0.00131
DRACE3_1     CENREG_4         0.16554            0.09644
```

```
DRACE3_1     DIC0401         -0.03579           -0.03363
DRACE3_1     GENDER_1         0.02500           -0.02917
DRACE3_1     SCORE           -0.26833           -0.29130
DRACE3_2     BIC0501_1        0.00273           -0.01984
DRACE3_2     BIE0601_3        0.02600            0.05672
DRACE3_2     BIE0601_4       -0.02415           -0.05960
DRACE3_2     BORNUSA_1        0.23683            0.16593
DRACE3_2     CENREG_3         0.14857            0.10274
DRACE3_2     CENREG_4        -0.20689           -0.15382
DRACE3_2     DIC0401         -0.00143           -0.00307
DRACE3_2     GENDER_1        -0.03840            0.00320
DRACE3_2     SCORE           -0.08417           -0.06652
EDUC3        BIC0501_1       -0.11532           -0.08752
EDUC3        BIE0601_3       -0.00146           -0.03930
EDUC3        BIE0601_4       -0.11035           -0.06669
EDUC3        BORNUSA_1        0.02303            0.11035
EDUC3        CENREG_3        -0.08731           -0.03086
EDUC3        CENREG_4         0.01859           -0.04678
EDUC3        DIC0401          0.11225            0.11427
EDUC3        GENDER_1         0.09066            0.03942
EDUC3        SCORE            0.44680            0.42507
EDUC_DET     GENDER_1         0.10158            0.06126
SCORE        DAGE            -0.10646           -0.11102
SCORE        EDUC_DET         0.44552            0.42413
```

```
Number of Pairwise Associations computed : 253
* denotes the after swapping correlation changed by more than 1.96 standard errors
Variable(s) 'DRACE3_* BIB1201_* BIC0501_* BID0101_* BIE0601_* BORNUSA_* CENREG_* GENDER_*' are
            indicator variables
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                    09:22 Wednesday, December 23, 2015  43
Supplemental Tables for DRB Members
Weighted Measures of Association in Swapping Variables and Key Output Variables
Where r(before)-r(after) is not equal to 0

Variable     Variable        Weighted           Weighted
1            2            Before Swapping    After Swapping


BIB1201_1    DAGE             0.00397           -0.03045
BIB1201_1    EDUC_DET         0.07744            0.07821
BIC0501_1    DAGE            -0.16214           -0.10447
BIC0501_1    EDUC_DET        -0.11769           -0.08414
BID0101_1    DAGE             0.06821            0.09598
BID0101_1    EDUC_DET         0.05982            0.06070
BIE0601_1    DAGE             0.00649           -0.00226
BIE0601_1    EDUC_DET         0.16738            0.15269
BIE0601_2    DAGE            -0.21811           -0.26466
BIE0601_2    EDUC_DET         0.03155            0.03044
BIE0601_3    DAGE             0.15329            0.13946
BIE0601_3    EDUC_DET         0.00142           -0.02484
BIE0601_4    DAGE             0.03506            0.10158
BIE0601_4    EDUC_DET        -0.11108           -0.05942
BIE0601_5    DAGE             0.12730            0.16210
BIE0601_5    EDUC_DET        -0.19664           -0.19687
BORNUSA_1    DAGE             0.05151            0.09468
BORNUSA_1    EDUC_DET        -0.00575            0.07784
CENREG_1     DAGE            -0.11462           -0.11502
CENREG_1     EDUC_DET        -0.05624           -0.05681
CENREG_2     DAGE            -0.04712           -0.05464
CENREG_2     EDUC_DET         0.16198            0.16104
CENREG_3     DAGE             0.04973            0.10298
CENREG_3     EDUC_DET        -0.10610           -0.05289
CENREG_4     DAGE             0.08860            0.03553
CENREG_4     EDUC_DET         0.01211           -0.04718
DAGE         DIC0401          0.36650            0.39023
DAGE         EDUC_DET         0.05384            0.05966
DAGE         GENDER_1         0.04994            0.03195
DAGE3        BIB1201_1       -0.03038           -0.06013
DAGE3        BIC0501_1       -0.16598           -0.10984
DAGE3        BID0101_1        0.10118            0.12806
DAGE3        BIE0601_1        0.02936            0.02719
DAGE3        BIE0601_2       -0.17566           -0.21894
DAGE3        BIE0601_3        0.09530            0.07209
DAGE3        BIE0601_4        0.03364            0.09683
DAGE3        BIE0601_5        0.09309            0.12735
DAGE3        BORNUSA_1        0.00653            0.06240
DAGE3        CENREG_1        -0.06654           -0.06406
DAGE3        CENREG_2        -0.02326           -0.03004
DAGE3        CENREG_3         0.01915            0.07729
DAGE3        CENREG_4         0.05853           -0.00332
DAGE3        DAGE             0.87329            0.87333
DAGE3        DIC0401          0.31296            0.33156
```

```
DAGE3        EDUC_DET         0.03914            0.04489
DAGE3        GENDER_1         0.06116            0.02952
DAGE3        SCORE           -0.07921           -0.08286
DIC0401      EDUC_DET         0.09163            0.09930
DRACE3_1     BIB1201_1        0.16207            0.16314
DRACE3_1     BIC0501_1       -0.08689           -0.05786
DRACE3_1     BID0101_1        0.05651            0.05740
DRACE3_1     BIE0601_1       -0.06368           -0.06708
DRACE3_1     BIE0601_2       -0.06247           -0.06347
DRACE3_1     BIE0601_3        0.03063           -0.00312
DRACE3_1     BIE0601_4        0.04425            0.08843
DRACE3_1     BIE0601_5        0.11894            0.11833
DRACE3_1     BORNUSA_1       -0.43037           -0.34479
DRACE3_1     CENREG_1         0.10542            0.10464
DRACE3_1     CENREG_2        -0.19892           -0.19944
DRACE3_1     CENREG_3        -0.07125           -0.01715
DRACE3_1     CENREG_4         0.18349            0.12296
DRACE3_1     DAGE             0.03252            0.03315
DRACE3_1     DAGE3            0.08730            0.08788
DRACE3_1     DIC0401         -0.04992           -0.04636
DRACE3_1     DRACE3_2        -0.42683           -0.42720
DRACE3_1     DRACE3_3        -0.33539           -0.33604
DRACE3_1     EDUC3           -0.05473           -0.05279
```

```
Number of Pairwise Associations computed : 253
* denotes the after swapping correlation changed by more than 1.96 standard errors
Variable(s) 'DRACE3_* BIB1201_* BIC0501_* BID0101_* BIE0601_* BORNUSA_* CENREG_* GENDER_*' are
            indicator variables
```

Appendix A-2.   *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                    09:22 Wednesday, December 23, 2015  44
Supplemental Tables for DRB Members
Weighted Measures of Association in Swapping Variables and Key Output Variables
Where r(before)-r(after) is not equal to 0

Variable    Variable        Weighted           Weighted
   1        2            Before Swapping    After Swapping

DRACE3_1    EDUC_DET        -0.08098           -0.07846
DRACE3_1    GENDER_1         0.03273           -0.01536
DRACE3_1    SCORE           -0.25779           -0.27755
DRACE3_2    BIB1201_1       -0.08123           -0.08219
DRACE3_2    BIC0501_1        0.00114           -0.02055
DRACE3_2    BID0101_1       -0.10379           -0.10452
DRACE3_2    BIE0601_1       -0.02001           -0.01740
DRACE3_2    BIE0601_2        0.03450            0.03532
DRACE3_2    BIE0601_3        0.05105            0.07632
DRACE3_2    BIE0601_4       -0.03733           -0.07050
DRACE3_2    BIE0601_5       -0.04799           -0.04766
DRACE3_2    BORNUSA_1        0.25477            0.19103
DRACE3_2    CENREG_1         0.01102            0.01150
DRACE3_2    CENREG_2         0.01076            0.01134
DRACE3_2    CENREG_3         0.16679            0.12631
DRACE3_2    CENREG_4        -0.21029           -0.16509
DRACE3_2    DAGE            -0.01220           -0.01296
DRACE3_2    DAGE3           -0.04099           -0.04156
DRACE3_2    DIC0401          0.04239            0.03980
DRACE3_2    DRACE3_3        -0.70880           -0.70803
DRACE3_2    EDUC3           -0.16466           -0.16494
DRACE3_2    EDUC_DET        -0.12629           -0.12687
DRACE3_2    GENDER_1        -0.03462            0.00140
DRACE3_2    SCORE           -0.08728           -0.07222
DRACE3_3    DAGE            -0.01265           -0.01239
DRACE3_3    DAGE3           -0.02539           -0.02534
DRACE3_3    EDUC3            0.21423            0.21305
DRACE3_3    EDUC_DET         0.19472            0.19344
EDUC3       BIB1201_1        0.12957            0.13074
EDUC3       BIC0501_1       -0.12104           -0.09273
EDUC3       BID0101_1        0.01640            0.01733
EDUC3       BIE0601_1        0.17895            0.17545
EDUC3       BIE0601_2        0.03337            0.03229
EDUC3       BIE0601_3       -0.01584           -0.04875
EDUC3       BIE0601_4       -0.11818           -0.07491
EDUC3       BIE0601_5       -0.18862           -0.18895
EDUC3       BORNUSA_1        0.00692            0.08994
EDUC3       CENREG_1        -0.11109           -0.11166
EDUC3       CENREG_2         0.19092            0.19006
EDUC3       CENREG_3        -0.09826           -0.04547
EDUC3       CENREG_4         0.02258           -0.03631
EDUC3       DAGE             0.10681            0.10761
EDUC3       DAGE3            0.08359            0.08426
```

```
EDUC3       DIC0401          0.08029           0.08362
EDUC3       EDUC_DET         0.92523           0.92528
EDUC3       GENDER_1         0.07166           0.02471
EDUC3       SCORE            0.42780           0.40794
EDUC_DET    GENDER_1         0.07601           0.03877
SCORE       DAGE            -0.10997          -0.11072
SCORE       EDUC_DET         0.41599           0.39569
```

```
Number of Pairwise Associations computed : 253
* denotes the after swapping correlation changed by more than 1.96 standard errors
Variable(s) 'DRACE3_* BIB1201_* BIC0501_* BID0101_* BIE0601_* BORNUSA_* CENREG_* GENDER_*' are
            indicator variables
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                    09:22 Wednesday, December 23, 2015  45
Supplemental Tables for DRB Members
Unweighted Multiple Regression Coefficients for model KEYOUT = INTERCEPT SWAPVARS

                                         Estimate    Estimate
Dependent                                 Before      After
Variable     Parameter                   Swapping    Swapping

  SCORE      Error degrees of freedom    177.0000    177.0000
             Intercept                   217.7043    221.6826
             DRACE3 1                    -47.6002    -51.2375
             DRACE3 2                    -16.5107    -15.9490
             EDUC3                        32.6248     30.9327
             DAGE3                       -12.4106    -12.8059
             R-squared                     0.2849      0.2782




* denotes the after swapping beta coefficient changed by more than 1.96 standard errors
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                    09:22 Wednesday, December 23, 2015  46
Supplemental Tables for DRB Members
Weighted Multiple Regression Coefficients for model KEYOUT = INTERCEPT SWAPVARS

                                         Estimate    Estimate
Dependent                                 Before      After
Variable     Parameter                   Swapping    Swapping

  SCORE      Error degrees of freedom    177.0000    177.0000
             Intercept                   215.7333    218.9424
             DRACE3 1                    -49.4962    -52.8515
             DRACE3 2                    -19.1253    -18.7350
             EDUC3                        32.5861     30.9325
             DAGE3                       -10.3206    -10.3493
             R-squared                     0.2648      0.2581





* denotes the after swapping beta coefficient changed by more than 1.96 standard errors
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                    09:22 Wednesday, December 23, 2015  47
Supplemental Tables for DRB Members
Unweighted Multiple Regression Coefficients for User-specified Models

                                         Estimate    Estimate
Dependent                                 Before      After
Variable     Parameter                   Swapping    Swapping

  SCORE      Error degrees of freedom    175.0000    175.0000
             Intercept                   231.9791    233.1957
             CENREG 1                    -27.2338    -30.1073
             CENREG 2                     -5.9335     -7.8830
             CENREG 3                     -9.8158    -13.6471
             EDUC_DET                     13.7469     13.1653
             DAGE                         -1.0904     -1.1013
             GENDER 1                     -7.6383     -3.3220
             R-squared                     0.2399      0.2256





* denotes the after swapping beta coefficient changed by more than 1.96 standard errors
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                   09:22 Wednesday, December 23, 2015  48
Supplemental Tables for DRB Members
Weighted Multiple Regression Coefficients for User-specified Models

                                         Estimate    Estimate
Dependent                                 Before      After
Variable     Parameter                   Swapping    Swapping

  SCORE      Error degrees of freedom    175.0000    175.0000
             Intercept                   229.4148    229.5338
             CENREG 1                    -32.0225    -34.0545
             CENREG 2                     -7.3854     -8.9345
             CENREG 3                    -12.6968    -15.8622
             EDUC_DET                     13.5072     12.9724
             DAGE                         -1.0575     -1.0406
             GENDER 1                     -2.8653      0.7807
             R-squared                     0.2166      0.2040
```

```
* denotes the after swapping beta coefficient changed by more than 1.96 standard errors
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                    09:22 Wednesday, December 23, 2015  49
Supplemental Tables for DRB Members
Data Utility Measure for Tables
SWAPVARS=DRACE3 EDUC3 DAGE3

                                                   For         Utility     Cell Count/
                Application                    Variable(s)      value      Small Cell

Hellinger's Distance, all cells                Across All     0.590329!       23/23
Hellinger's Distance, excluding small cells    Across All     0.000000        0/0
Hellinger's Distance, all cells                DRACE3         0.191411!       3/1
Hellinger's Distance, all cells                EDUC3          0.183041!       3/1
Hellinger's Distance, all cells                DAGE3          0.112057!       3/1
Hellinger's Distance, excluding small cells    DRACE3         0.097982        2/0
Hellinger's Distance, excluding small cells    EDUC3          0.095782        2/0
Hellinger's Distance, excluding small cells    DAGE3          0.112057        2/0
```

```
! denotes small cell(s)
```

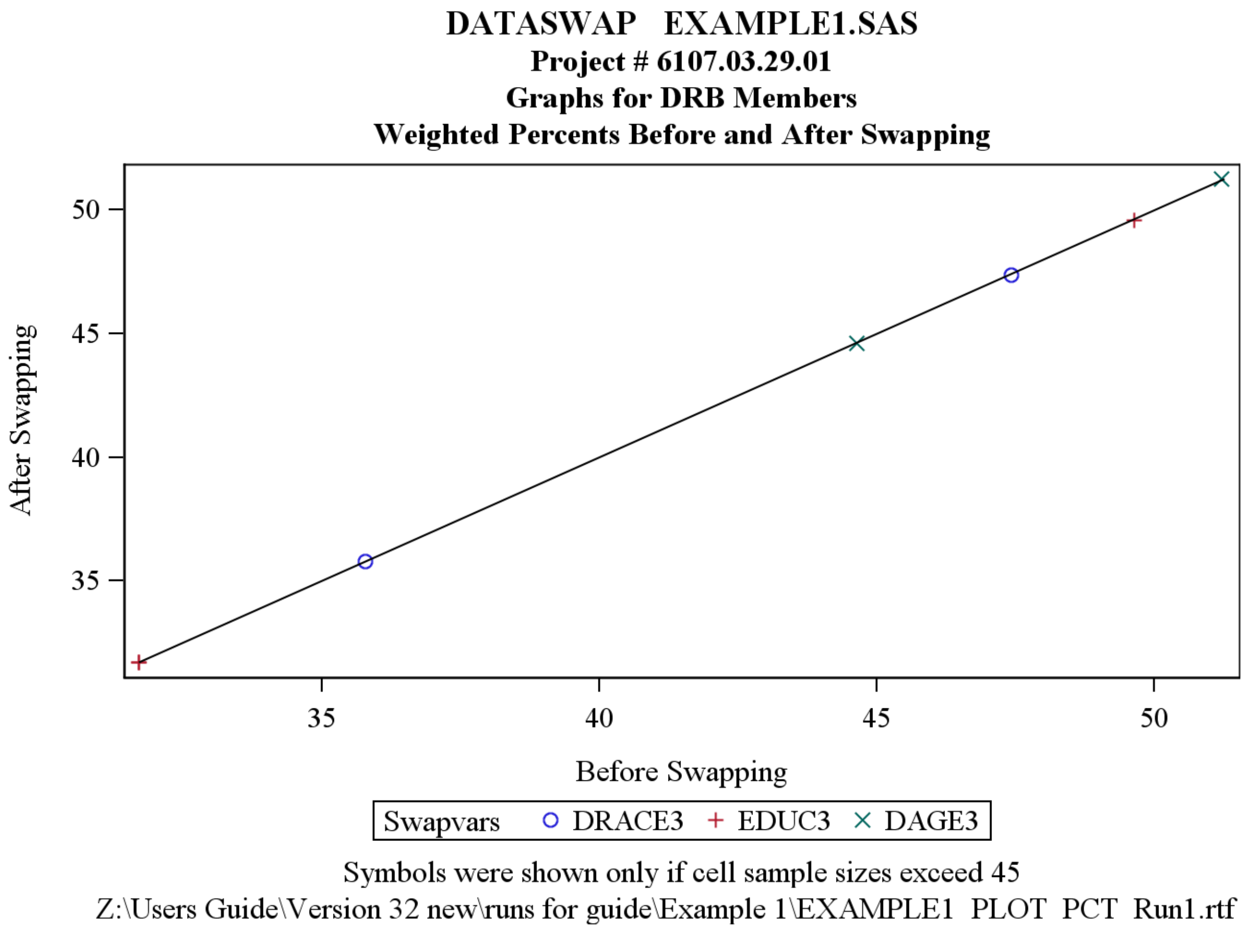
Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

```
DATASWAP   EXAMPLE1.SAS                                     09:22 Wednesday, December 23, 2015  50
Supplemental Tables for DRB Members
Data Utility Measure for Pairwise Associations
Including BOUNDARY, SWAPVARS and KEYVARS

                                     Utility
Application                           value

Pearson's Product Correlation        0.29307
Pearson's Contingency Coefficient    0.10402
Cramer's V                           0.10598
```

Appendix A-2. *DataSwap* Example 1, run 1 output – no hard boundary (continued)

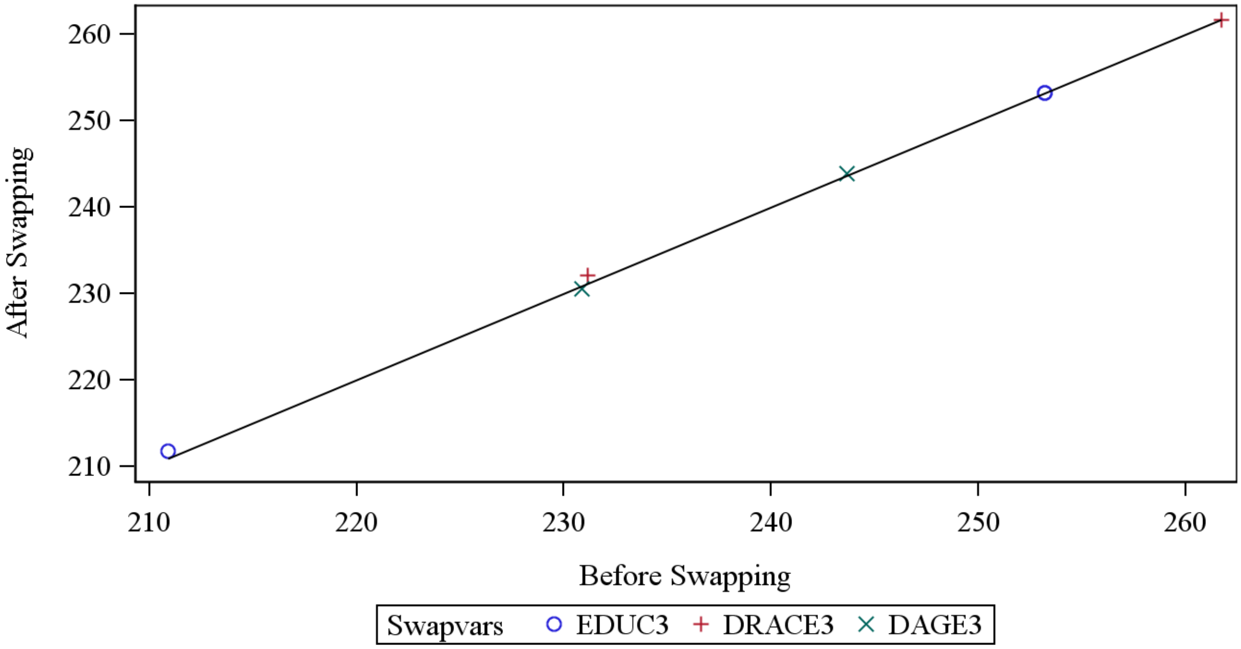
```
DATASWAP   EXAMPLE1.SAS                                09:22 Wednesday, December 23, 2015  51
Supplemental Tables for DRB Members
Data Utility Measure for Regression Coefficients - Default and User-defined models
KEYOUT=SCORE
SWAPVARS=DRACE3 EDUC3 DAGE3

                                     Utility
              Model                   value

SCORE=DRACE3 EDUC3 DAGE3             0.11721
SCORE=CENREG EDUC_DET DAGE GENDER    0.11685
Across All Regression Models         0.11703
```

Appendix A-3. *DataSwap* Example 1, run 1 graphs – no hard boundary



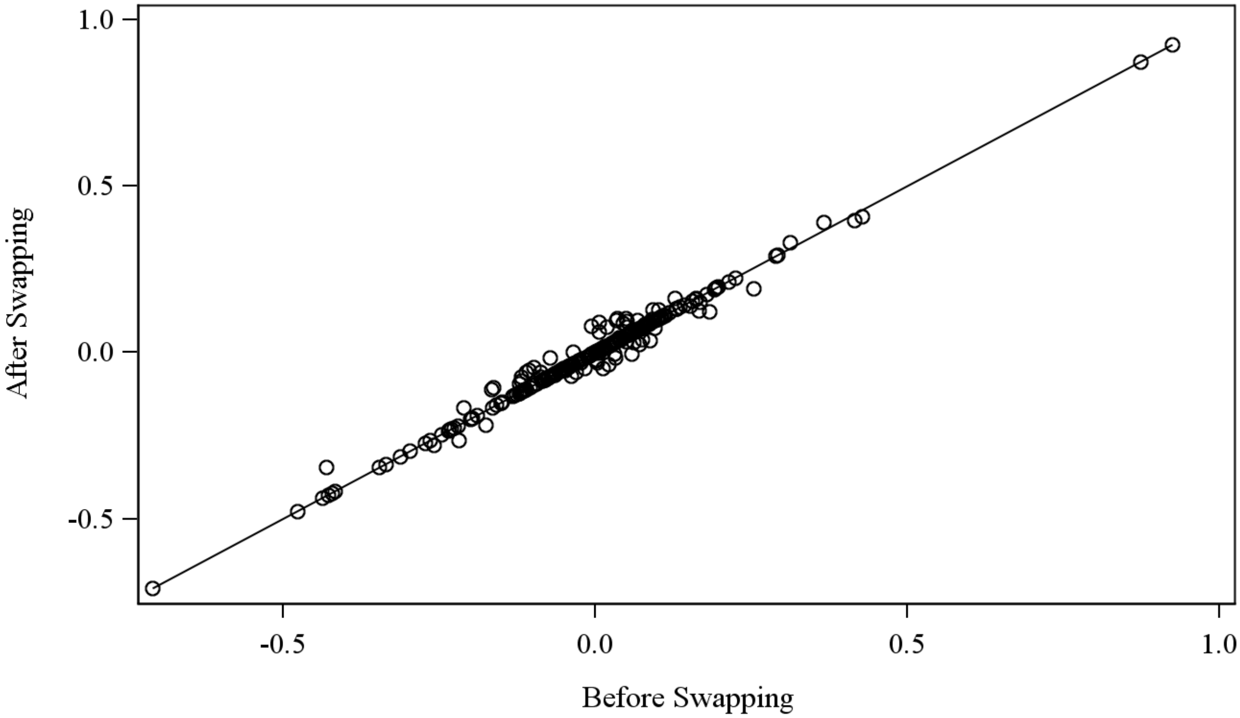
Appendix A-3. *DataSwap* Example 1, run 1 graphs – no hard boundary (continued)

DATASWAP EXAMPLE1.SAS
Project # 6107.03.29.01
Graphs for DRB Members
Weighted Means Before and After Swapping
KEYOUT=SCORE



Symbols were shown only if cell sample sizes exceed 45
Z:\Users Guide\Version 32 new\runs for guide\Example 1\EXAMPLE1 PLOT MEAN Run1.rtf

DATASWAP EXAMPLE1.SAS
Project # 6107.03.29.01
Graphs for DRB Members
Weighted Measures of Association Before and After Swapping

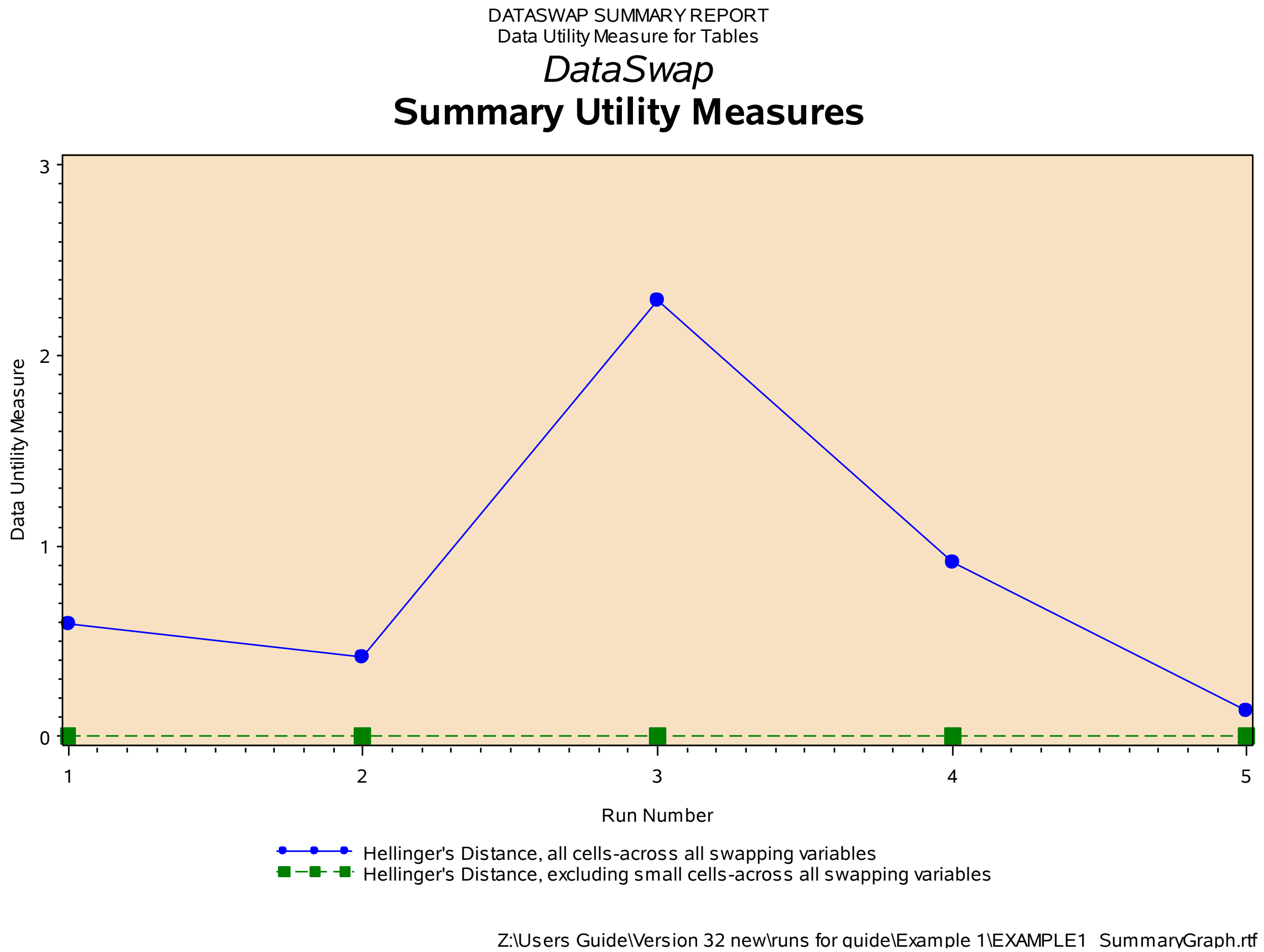


Z:\Users Guide\Version 32 new\runs for guide\Example 1\EXAMPLE1 PLOT CORR Run1.rtf

Appendix A-4. *DataSwap* Example 1, summary output – no hard boundary (continued)

```
DATASWAP SUMMARY REPORT                                                         09:22 Wednesday, December 23, 2015  52
Data Utility Measure for Tables

                                                For       Run 1      Run 2     Run 3     Run 4     Run 5   Cell Count/
Application                                 Variable(s) Seed=22061 Seed=345   Seed=76   Seed=98  Seed=239  Small Cell

Utility Measures for Tables

Hellinger's Distance, all cells             Across All  0.590329!  0.412697! 2.290697! 0.915342! 0.131213!    23/23
Hellinger's Distance, excluding small cells Across All  0.000000   0.000000  0.000000  0.000000  0.000000     0/0
Hellinger's Distance, all cells             DRACE3      0.191411!  0.018379! 0.200794! 0.022579! 0.004181!    3/1
Hellinger's Distance, all cells             EDUC3       0.183041!  0.022532! 0.210491! 0.021590! 0.003998!    3/1
Hellinger's Distance, all cells             DAGE3       0.112057!  0.198985! 0.519130! 0.027803! 0.064575!    3/1
Hellinger's Distance, excluding small cells DRACE3      0.097982   0.018379  0.097173  0.011550  0.002138     2/0
Hellinger's Distance, excluding small cells EDUC3       0.095782   0.011781  0.110158  0.011290  0.002090     2/0
Hellinger's Distance, excluding small cells DAGE3       0.112057   0.192075  0.519130  0.027803  0.048919     2/0


Utility Measures for Pairwise Associations

Pearson's Product Correlation                           0.293071   0.134749  0.199399  0.186394  0.171779
Pearson's Contingency Coefficient                       0.104016   0.053716  0.079003  0.064585  0.042294
Cramer's V                                              0.105980   0.055460  0.080827  0.065172  0.043452


Utility Measures for Multivariate Assoc.

SCORE=DRACE3 EDUC3 DAGE3                                0.117214   0.222812  0.336456  0.090312  0.028253
SCORE=CENREG EDUC_DET DAGE GENDER                       0.116848   0.145233  0.158416  0.108499  0.134646
Across All Regression Models                            0.117031   0.184022  0.247436  0.099406  0.081450
```

Appendix A-5. *DataSwap* Example 1, summary graphs – no hard boundary



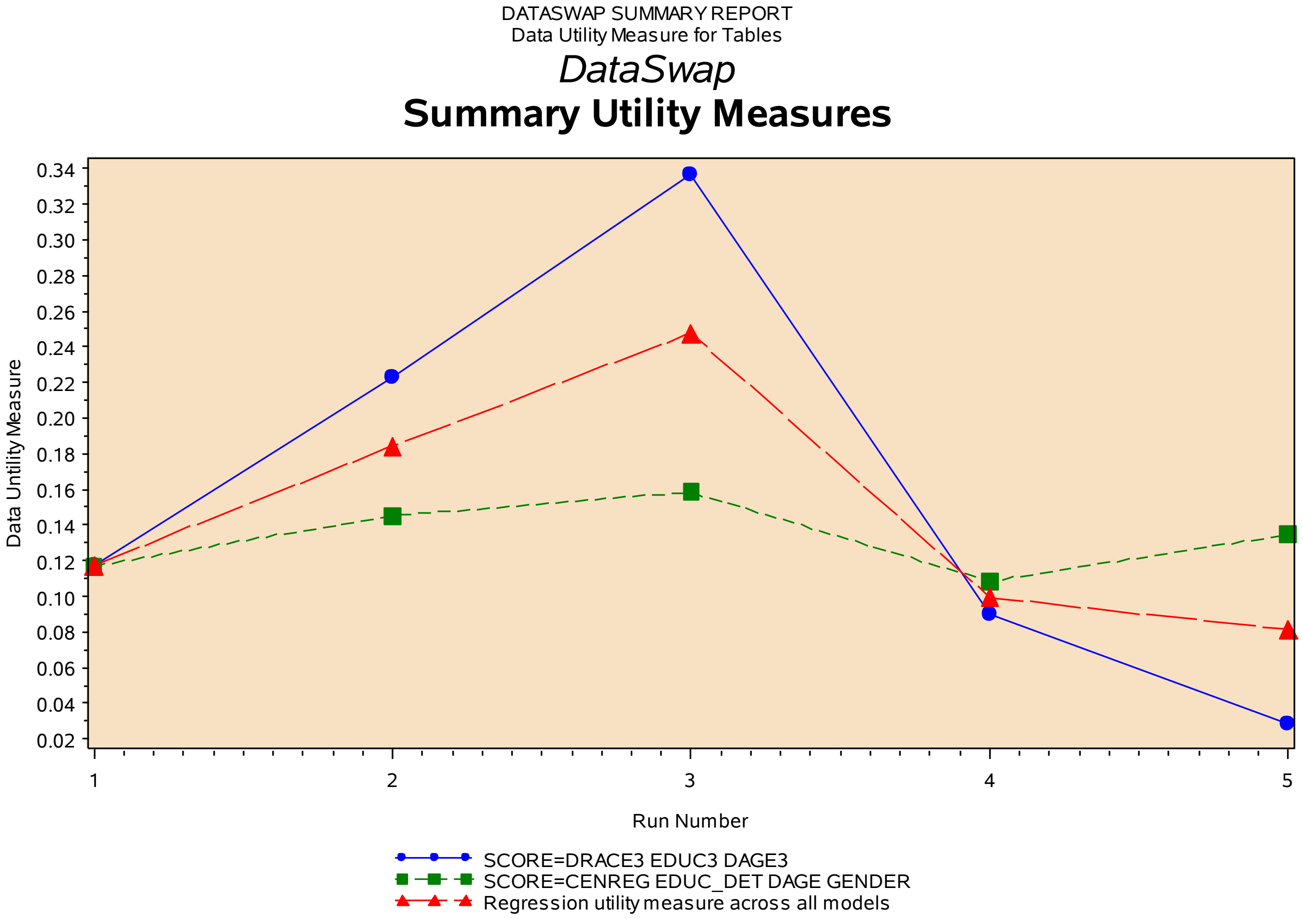
DATASWAP SUMMARY REPORT
Data Utility Measure for Tables

DataSwap

Summary Utility Measures

Z:\Users Guide\Version 32 new\runs for guide\Example 1\EXAMPLE1_SummaryGraph.rtf

Appendix A-5. *DataSwap* Example 1, summary graphs – no hard boundary (continued)



DATASWAP SUMMARY REPORT
Data Utility Measure for Tables

DataSwap

Summary Utility Measures

Z:\Users Guide\Version 32 new\runs for guide\Example 1\EXAMPLE1 SummaryGraph.rtf

APPENDIX B

Example 2   Parameter Sheet and Standard Output – Hard Boundary

Appendix B-1. *DataSwap* Example 2 parameter sheet –hard boundary

Data Swapping

3.2

Parameter specification form

Parameter		*	Entry	Default	Description
Input	**DATA=**	**R**	INFLE.EXAMPLEDATA		input file
Controllers	**SEED=**	**O**	22601 345 76 98 239	**0**	random seed
	ID=	**R**	CASEID		case identification - a single variable
	VARSTRAT=	**O**	VARSTRAT		Variance Stratum
	VARUNIT=	**O**	VARUNIT		Variance Unit
Sampling	**RATE=**	**R**	RATE		swapping rate - a single variable or a number
Controllers	**MOS=**	**O**		**1**	measure of size - a single variable
	STRATUM=	**O**	RiskStratum	**1**	stratum - a single variable
	SORTVARS=	**O**		**BOUNDARY|| SWAPVARS[1]**	sort order for non-certainty selection - a list of variables
	SWAPMETH=	**O**	1	**2**	swapping method – 1:original 2:balanced
	SWAPVARS=	**R**	DAGE3		primary swap variables – a list of variables
	SWAPVARS_T=	**O**	O	**O for all**	SWAPVARS variable type: Ordinal (O) Nominal (N)
	BOUNDARY=	**O**	DRACE3 EDUC3		hard boundary - a list of variables
Swap	**BOUNDARY_T=**	**O**	N O	**O for all**	BOUNDARY variable type: Ordinal (O) Nominal (N)
Controllers	**BIASVAR=**	**O**	DAGE3	**the right- most var in SWAPVARS**	variable for calculating bias - a single variable (must be numeric)
	WGT =	**R**	WEIGHT		case survey weight - a single variable
	LINKSWAP=	**O**	DAGE		linked swap variables - a list of variables separated by #
	IMPUTE=	**O**	Y	**y**	impute if swapvars have missing values.
	MISSINGDEF=	**O**	NULL	**‘ ‘ or .**	values to be defined as missing – a list of values separated by #
	OUT=	**R**	C:\DataSwap\Data\example2		dataset name for swapped result
	KEYOUT=	**O**	SCORE		key outcome continuous variable for weighted means and correlations (must be numeric)
Output	**KEYVARS=**	**O**	BIB1201 BIC0501 BID0101 BIE0601 BORNUSA CENREG DAGE DIC0401 EDUC_DET GENDER		key variables for output tables and correlations
Controllers	**KEYVARS_T=**	**O**	N N N N N N O O O N	**O for all**	KEYVARS variable type: Ordinal (O) Nominal (N)
	MODELS=	**O**	SCORE : CENREG EDUC_DET DAGE GENDER		a list of models separated by # exp. X:Y Z#W : V1 V2
	MODELCLASS=	**O**	CENREG GENDER		a list of class variables separated by # exp. Y Z#V1
	TOLFLAG=	**O**	0.1#45#1.96#1.1	**0.10#45#1.96#1.1**	tolerance measure for flagging outliers
	MAXCAT=	**O**	20	**20**	Maximum # of categories for DRB table variables
	LISTPAIR=	**O**	S#1	**S#1.0**	Subset listing for swapping pairs

* O: Optional R: Required

[1] The double vertical bar denotes the concatenation of the list of BOUNDARY and SWAPVARS variables.

Version 3.3, December 2015

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary

```
DATASWAP   EXAMPLE2.SAS                                           09:23 Wednesday, December 23, 2015   1
THE INFORMATION PAGE

DATASWAP MACRO VERSION #:                          Version 3.3

INPUT DATA SET:                                    INFLE.EXAMPLEDATA

OUTPUT DATA SET:                                   OUT.EXAMPLE2_Run1

OBSERVATIONS IN INPUT DATA SET:                    182

RANDOM SEED:                                       22061

CASE IDENTIFICATION:                               CASEID

MEASURE OF SIZE:                                   1 (Default)

STRATUM:                                           RISKSTRATUM

DESIRED SWAPPING RATE:                             RATE

NON-CERTAINTY SELECTION ORDER:                     NOT SPECIFIED

SWAP CELL DEFINITION:                              DRACE3 EDUC3 DAGE3

TOTAL NUMBER OF CELLS:                             23

HARD BOUNDARY:                                     DRACE3 EDUC3

BOUNDARY VARIABLE TYPE:                            N O

PRIMARY SWAP VARIABLE(S):                          DAGE3

SWAPVARS VARIABLE TYPE:                            O

LINKED SWAP VARIABLE(S):                           DAGE

BIAS VARIABLE:                                     DAGE3

CASE WEIGHT:                                       WEIGHT

VARIANCE STRATUM:                                  VARSTRAT

VARIANCE UNIT:                                     VARUNIT

IMPUTE OPTION:                                     Yes (Default)

MISSINGDEF:                                        NULL

SWAPPING METHOD:                                   1 (Standard)

KEY OUTPUT VARIABLE:                               SCORE

OTHER KEY VARIABLES:                               BIB1201 BIC0501 BID0101 BIE0601 BORNUSA CENREG DAGE
DIC0401 EDUC_DET
                                                   GENDER

KEYVARS VARIABLE TYPE:                             N N N N N N O O O N

USER-SPECIFIED MODELS:                             SCORE:CENREG EDUC_DET DAGE GENDER

CLASS VARIABLES IN MODELS:                         CENREG GENDER

TOLERANCE FLAG:                                    0.1#45#1.96#1.1 (Default)

SWAPPING PAIR OUTPUT CONTROL:                      S#1

MAXIMUM NUM. OF CATEGORIES FOR DRB TABLE VARIABLES:20 (Default)
```

TOTAL NUMBER OF ITERATIONS: 3

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                                      09:23 Wednesday, December 23, 2015   5
User Only Output
DataSwap Results by Pairs [1st Pair Displayed Only]

Pair=1

                                  Recoded
                                 highest
                    Derived      education  Derived age
                Race/ethnicity    level -    category -  Age derived                Ever been                  Nay work
                 - 1:Hispanic,  1:less HS,     1:<30,     fr date of                in pgm to    Ever been   assignments
Identification    2:NH Black,      2:=HS,      2:<50,      birth or   Literature     improve     placed on    inside or
     no.            3:Other       3: >HS      3: >=50      Screener      score    basic skills  probation?     outside?

   91510317            2             1           2            31       246.217      1                1           1
   90710102            2             1           1            29       250.451      2                1           1



  How often
    write       if born in              Derived      Highest                        Initial     Change     Change
letters/memos   USA 1=Yes,   Census   years since    level of   Gender    Final    Risk: Risk   flag for   flag for
   in Engl?        2=No      region    admission    education    (sex)    weight     Stratum     DAGE3       DAGE

      1              1          3         1.33          3          2     572.090     4              1          1
      3              1          3         1.58          3          1     795.079     1              1          1
```

Appendix B-2.   *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                                           09:23 Wednesday, December 23, 2015   8
User Only Output
Aggregated Level -- Changes to Swapped Variable SWAPVARS=DAGE3

Derived age                                                                  Estimated    Estimated
 category -              Unweighted    Unweighted    Weighted    Weighted    Standard      Standard       Ratio of
   1:<30,                 Percents      Percents     Percents    Percents     Errors        Errors        Estimated
   2:<50,      Sample      Before        After        Before      After       Before        After      Standard Errors
  3: >=50       Size      Swapping      Swapping     Swapping    Swapping    Swapping      Swapping    (After/Before)

     1           87       47.8022       47.8022      44.6327     44.9139       3.9701      3.98009        1.002507
     2           88       48.3516       48.3516      51.2125     50.9312       4.1759      4.18538        1.002266
     3            7        3.8462        3.8462       4.1548      4.1548       1.6172      1.61717        1.000000
```

```
* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
@ denotes SE ratios exceeds 1.1 and sample size exceeds 45
```

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                    09:23 Wednesday, December 23, 2015   9
User Only Output
Aggregated Level -- Changes to Swapped Variable SWAPVARS=DAGE3

                                                                       Ratio of Estimated
                         Before/After    Weighted       Estimated        Standard Errors
DAGE3    N     KEYOUT      Swapping        Mean      Standard Errors      (After/Before)

  1      87    SCORE        Before       243.6602         5.2952              1.054501
                            After        241.8882         5.5838
  2      88    SCORE        Before       230.8760         7.8594              1.017859
                            After        232.3680         7.9998
  3       7    SCORE        Before       240.1879        18.6124              1.000000
                            After        240.1879        18.6124
```

```
* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
@ denotes SE ratios exceeds 1.1 and sample size exceeds 45
```

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                                                        09:23 Wednesday, December 23, 2015  10
User Only Output
Aggregated Level -- Changes to Swapped Variable LINKSWAP=DAGE
                                                                             Estimated    Estimated
Age derived              Unweighted    Unweighted    Weighted    Weighted    Standard      Standard       Ratio of
 fr date of               Percents      Percents     Percents    Percents     Errors        Errors        Estimated
  birth or     Sample     Before         After        Before      After       Before        After      Standard Errors
  Screener      Size      Swapping      Swapping     Swapping    Swapping    Swapping      Swapping    (After/Before)
     18           2        1.0989        1.0989       0.8551      0.8551       0.6065       0.60647        1.000000
     19           7        3.8462        3.8462       3.6755      3.6755       1.4071       1.40711        1.000000
     20           9        4.9451        4.9451       4.7582      4.7582       1.6123       1.61229        1.000000
     21           6        3.2967        3.2967       3.1320      3.1320       1.2906       1.29060        1.000000
     22           6        3.2967        3.2967       2.7017      2.7916       1.1228       1.12635        1.003199
     23           8        4.3956        4.3956       4.3434      4.3261       1.5572       1.55727        1.000061
     24           5        2.7473        2.7473       2.2124      2.2124       0.7149       0.71494        1.000000
     25           7        3.8462        3.8462       3.5267      3.5267       1.3610       1.36102        1.000000
     26          12        6.5934        6.5934       6.6062      6.6111       1.6719       1.67192        1.000004
     27           7        3.8462        3.8462       3.6243      3.6470       1.3978       1.39798        1.000132
     28          12        6.5934        6.5934       6.3024      6.3024       1.9103       1.91028        1.000000
     29           6        3.2967        3.2967       2.8946      3.0757       1.2185       1.23191        1.010988
     30           6        3.2967        3.2967       3.5710      3.5710       1.4753       1.47526        1.000000
     31           7        3.8462        3.8462       4.6080      4.4268       1.7765       1.78567        1.005185
     32           6        3.2967        3.2967       3.1655      3.1655       1.3121       1.31206        1.000000
     33           7        3.8462        3.8462       4.4769      4.4769       1.7080       1.70800        1.000000
     34           6        3.2967        3.2967       3.7178      3.7178       1.5423       1.54231        1.000000
     35           8        4.3956        4.3956       4.5886      4.5886       1.7600       1.75999        1.000000
     36           2        1.0989        1.0989       1.4113      1.4113       0.9923       0.99233        1.000000
     37           7        3.8462        3.8462       4.4246      4.4246       1.6928       1.69282        1.000000
     38           5        2.7473        2.7473       3.1325      3.1325       1.4637       1.46372        1.000000
     39           4        2.1978        2.1978       2.0119      2.0291       1.0192       1.01938        1.000143
     40           5        2.7473        2.7473       2.5779      2.5779       1.1672       1.16723        1.000000
     41           9        4.9451        4.9451       4.8760      4.8760       1.7321       1.73212        1.000000
     42           4        2.1978        2.1978       2.2923      2.2923       1.1740       1.17401        1.000000
     43           3        1.6484        1.6484       1.6132      1.6132       0.9424       0.94240        1.000000
     44           1        0.5495        0.5495       0.6255      0.6207       0.6246       0.62463        1.000030
     45           3        1.6484        1.6484       1.4325      1.4098       0.8389       0.83924        1.000366
     47           1        0.5495        0.5495       0.6541      0.6541       0.6527       0.65270        1.000000
     48           1        0.5495        0.5495       0.5428      0.5428       0.5424       0.54236        1.000000
     49           3        1.6484        1.6484       1.4902      1.4003       0.8745       0.87910        1.005268
     50           1        0.5495        0.5495       0.4981      0.4981       0.4981       0.49813        1.000000
     53           2        1.0989        1.0989       1.3754      1.3754       0.9746       0.97460        1.000000
     54           1        0.5495        0.5495       0.8560      0.8560       0.8553       0.85531        1.000000
     56           1        0.5495        0.5495       0.5112      0.5112       0.5120       0.51203        1.000000
     58           1        0.5495        0.5495       0.4271      0.4271       0.4281       0.42809        1.000000
     68           1        0.5495        0.5495       0.4872      0.4872       0.4871       0.48710        1.000000

* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
@ denotes SE ratios exceeds 1.1 and sample size exceeds 45
```

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                    09:23 Wednesday, December 23, 2015  11
User Only Output
Aggregated Level -- Changes to Swapped Variable LINKSWAP=DAGE

                                                                       Ratio of Estimated
                        Before/After    Weighted       Estimated        Standard Errors
DAGE    N     KEYOUT      Swapping        Mean      Standard Errors      (After/Before)

 18      2    SCORE        Before       270.6371        13.7863             1.000000
                           After        270.6371        13.7863
 19      7    SCORE        Before       238.6627        18.1733             1.000000
                           After        238.6627        18.1733
 20      9    SCORE        Before       244.4130        19.5409             1.000000
                           After        244.4130        19.5409
 21      6    SCORE        Before       251.8129        25.1460             1.000000
                           After        251.8129        25.1460
 22      6    SCORE        Before       250.3557        18.1556             1.388271
                           After        232.8726        25.2049
 23      8    SCORE        Before       248.9454        13.7337             1.188903
                           After        240.1141        16.3281
 24      5    SCORE        Before       252.8639        27.9424             1.000000
                           After        252.8639        27.9424
 25      7    SCORE        Before       233.2826        15.1436             1.000000
                           After        233.2826        15.1436
 26     12    SCORE        Before       232.9787        12.1147             1.061938
                           After        228.6493        12.8650
 27      7    SCORE        Before       217.0387        27.5774             1.052442
                           After        226.0862        29.0236
 28     12    SCORE        Before       252.0289        12.3718             1.000000
                           After        252.0289        12.3718
 29      6    SCORE        Before       262.8960        16.1375             1.000017
                           After        262.8028        16.1378
 30      6    SCORE        Before       207.7303        36.5747             1.000000
                           After        207.7303        36.5747
 31      7    SCORE        Before       242.1075        19.9642             1.000774
                           After        241.3216        19.9797
 32      6    SCORE        Before       281.7871        20.2525             1.000000
                           After        281.7871        20.2525
 33      7    SCORE        Before       257.2944        14.0683             1.000000
                           After        257.2944        14.0683
 34      6    SCORE        Before       204.8023        26.3751             1.000000
                           After        204.8023        26.3751
 35      8    SCORE        Before       197.9061        27.5678             1.000000
                           After        197.9061        27.5678
 36      2    SCORE        Before       223.3839        19.5497             1.000000
                           After        223.3839        19.5497
 37      7    SCORE        Before       277.1285         9.9739             1.000000
                           After        277.1285         9.9739
 38      5    SCORE        Before       253.0645        19.3166             1.000000
                           After        253.0645        19.3166
 39      4    SCORE        Before       258.7778        21.3196             1.331556
                           After        277.5226        28.3882
 40      5    SCORE        Before       222.2087        19.6368             1.000000
                           After        222.2087        19.6368
 41      9    SCORE        Before       151.2064        29.4209             1.000000
                           After        151.2064        29.4209
 42      4    SCORE        Before       276.4834         7.8304             1.000000
                           After        276.4834         7.8304
 43      3    SCORE        Before       278.6246        13.7719             1.000000
                           After        278.6246        13.7719
 44      1    SCORE        Before       209.1415         0.0000
                           After        255.0729        45.9314
 45      3    SCORE        Before       264.1727         9.8261             2.512344
                           After        241.5261        24.6864
 47      1    SCORE        Before       220.8250         0.0000
                           After        220.8250         0.0000
 48      1    SCORE        Before        61.2798         0.0000
                           After         61.2798         0.0000
 49      3    SCORE        Before       213.2606        39.3590             1.296403
```

```
                           After        245.7327        51.0252
 50      1    SCORE        Before       266.6619         0.0000
                           After        266.6619         0.0000
 53      2    SCORE        Before       264.7618        44.2554             1.000000
                           After        264.7618        44.2554
 54      1    SCORE        Before       261.1270         0.0000


* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
```

@ denotes SE ratios exceeds 1.1 and sample size exceeds 45

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                    09:23 Wednesday, December 23, 2015  12
User Only Output
Aggregated Level -- Changes to Swapped Variable LINKSWAP=DAGE

                                                                      Ratio of Estimated
                       Before/After    Weighted       Estimated        Standard Errors
DAGE    N    KEYOUT      Swapping        Mean      Standard Errors      (After/Before)

                          After        261.1270           0
 56     1    SCORE        Before       195.0395           0
                          After        195.0395           0
 58     1    SCORE        Before       203.0768           0
                          After        203.0768           0
 68     1    SCORE        Before       186.8552           0
                          After        186.8552           0
```

```
* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
@ denotes SE ratios exceeds 1.1 and sample size exceeds 45
```

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                   09:23 Wednesday, December 23, 2015  13
User Only Output
Weighted Percents Sorted by Relative Difference in SWAPVARS and LINKSWAP

                               Weighted    Weighted
                               Percents    Percents     Absolute
                     Sample     Before       After      Relative
Variable    Value     Size     Swapping    Swapping    Difference

 DAGE        29          6       2.8946      3.0757     0.062577
 DAGE        49          3       1.4902      1.4003     0.060312
 DAGE        31          7       4.6080      4.4268     0.039309
 DAGE        22          6       2.7017      2.7916     0.033267
 DAGE        45          3       1.4325      1.4098     0.015842
 DAGE        39          4       2.0119      2.0291     0.008563
 DAGE        44          1       0.6255      0.6207     0.007678
 DAGE3       1          87      44.6327     44.9139     0.006302
 DAGE        27          7       3.6243      3.6470     0.006261
 DAGE3       2          88      51.2125     50.9312     0.005492
 DAGE        23          8       4.3434      4.3261     0.003967
 DAGE        26         12       6.6062      6.6111     0.000727
```

```
Only Absolute Relative Differences > 0 are shown
* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
```

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                09:23 Wednesday, December 23, 2015  14
User Only Output
Weighted Means Sorted by Relative Difference in SWAPVARS and LINKSWAP

                                  Weighted         Weighted
                               Mean of SCORE    Mean of SCORE     Absolute
                     Sample        Before           After         Relative
Variable    Value     Size        Swapping         Swapping      Difference

 DAGE        44         1         209.142          255.073        0.219619
 DAGE        49         3         213.261          245.733        0.152265
 DAGE        45         3         264.173          241.526        0.085727
 DAGE        39         4         258.778          277.523        0.072436
 DAGE        22         6         250.356          232.873        0.069833
 DAGE        27         7         217.039          226.086        0.041686
 DAGE        23         8         248.945          240.114        0.035475
 DAGE        26        12         232.979          228.649        0.018583
 DAGE3       1         87         243.660          241.888        0.007272
 DAGE3       2         88         230.876          232.368        0.006462
 DAGE        31         7         242.107          241.322        0.003246
 DAGE        29         6         262.896          262.803        0.000355
```

```
Only Absolute Relative Differences > 0 are shown
* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
```

~ denotes value was zero before swapping and non-zero after swapping

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                                                   09:23 Wednesday, December 23, 2015  17
Supplemental Tables for DRB Members
Aggregated Level -- Changes to Swapped Variable SWAPVARS=DAGE3

Derived age                                                                  Estimated    Estimated
 category -              Unweighted    Unweighted    Weighted    Weighted    Standard      Standard       Ratio of
   1:<30,                 Percents      Percents     Percents    Percents     Errors        Errors        Estimated
   2:<50,      Sample      Before        After        Before      After       Before        After      Standard Errors
  3: >=50       Size      Swapping      Swapping     Swapping    Swapping    Swapping      Swapping    (After/Before)

     1           87       47.8022       47.8022      44.6327     44.9139       3.9701      3.98009        1.002507
     2           88       48.3516       48.3516      51.2125     50.9312       4.1759      4.18538        1.002266
     3            7        3.8462        3.8462       4.1548      4.1548       1.6172      1.61717        1.000000
```

```
* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
@ denotes SE ratios exceeds 1.1 and sample size exceeds 45
```

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                  09:23 Wednesday, December 23, 2015  18
Supplemental Tables for DRB Members
Aggregated Level -- Changes to Swapped Variable SWAPVARS=DAGE3

                                                                      Ratio of Estimated
                         Before/After    Weighted       Estimated       Standard Errors
DAGE3    N     KEYOUT      Swapping        Mean      Standard Errors     (After/Before)

  1      87    SCORE         Before       243.6602         5.2952            1.054501
                             After        241.8882         5.5838
  2      88    SCORE         Before       230.8760         7.8594            1.017859
                             After        232.3680         7.9998
  3       7    SCORE         Before       240.1879        18.6124            1.000000
                             After        240.1879        18.6124
```

```
* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
@ denotes SE ratios exceeds 1.1 and sample size exceeds 45
```

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                 09:23 Wednesday, December 23, 2015  19
Supplemental Tables for DRB Members
Aggregated Level -- Changes to Swapped Variables in LINKSWAP=DAGE


          Report for DAGE was not generated because it has 37 levels which exceeds MAXCAT=20
                    Increase the value of MAXCAT to allow the report to be printed
```

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                    09:23 Wednesday, December 23, 2015  20
Supplemental Tables for DRB Members
Weighted Percents Sorted by Relative Difference in SWAPVARS and LINKSWAP

                               Weighted    Weighted
                               Percents    Percents     Absolute
                     Sample     Before       After      Relative
Variable    Value     Size     Swapping    Swapping    Difference

 DAGE3        1        87       44.6327     44.9139     0.006302
 DAGE3        2        88       51.2125     50.9312     0.005492
```

```
Only Absolute Relative Differences > 0 are shown
* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
```

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                    09:23 Wednesday, December 23, 2015  21
Supplemental Tables for DRB Members
Weighted Means Sorted by Relative Difference in SWAPVARS and LINKSWAP

                                  Weighted         Weighted
                               Mean of SCORE    Mean of SCORE     Absolute
                     Sample        Before           After          Relative
Variable    Value     Size        Swapping         Swapping       Difference

 DAGE3        1        87          243.660          241.888         0.007272
 DAGE3        2        88          230.876          232.368         0.006462
```

```
Only Absolute Relative Differences > 0 are shown
* denotes absolute relative difference exceeds 0.1 and sample size exceeds 45
~ denotes value was zero before swapping and non-zero after swapping
```

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                    09:23 Wednesday, December 23, 2015  22
Supplemental Tables for DRB Members
After swapping
Check creation of indicator variables for DRACE3

The FREQ Procedure

                                                                       Cumulative    Cumulative
DRACE3    DRACE3_1    DRACE3_2    DRACE3_3    Frequency     Percent     Frequency      Percent
-----------------------------------------------------------------------------------------------
     1           1           0           0          32       17.58            32        17.58
     2           0           1           0          79       43.41           111        60.99
     3           0           0           1          71       39.01           182       100.00
```

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                    09:23 Wednesday, December 23, 2015  23
Supplemental Tables for DRB Members
After swapping
Check creation of indicator variables for BIB1201

The FREQ Procedure

                                                              Cumulative    Cumulative
BIB1201    BIB1201_1    BIB1201_2    Frequency     Percent     Frequency      Percent
--------------------------------------------------------------------------------------
      1            1            0          29       15.93            29        15.93
      2            0            1         153       84.07           182       100.00
```

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                    09:23 Wednesday, December 23, 2015  24
Supplemental Tables for DRB Members
After swapping
Check creation of indicator variables for BIC0501

The FREQ Procedure

                                                              Cumulative    Cumulative
BIC0501    BIC0501_1    BIC0501_2    Frequency     Percent     Frequency      Percent
--------------------------------------------------------------------------------------
      1            1            0         108       59.34           108        59.34
      2            0            1          74       40.66           182       100.00
```

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                    09:23 Wednesday, December 23, 2015  25
Supplemental Tables for DRB Members
After swapping
Check creation of indicator variables for BID0101

The FREQ Procedure

                                                              Cumulative    Cumulative
BID0101    BID0101_1    BID0101_2    Frequency     Percent     Frequency      Percent
--------------------------------------------------------------------------------------
      1            1            0         128       70.33           128        70.33
      2            0            1          54       29.67           182       100.00
```

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                                                   09:23 Wednesday, December 23, 2015  26
Supplemental Tables for DRB Members
After swapping
Check creation of indicator variables for BIE0601

The FREQ Procedure

                                                                                             Cumulative   Cumulative
BIE0601   BIE0601_1   BIE0601_2   BIE0601_3   BIE0601_4   BIE0601_5   Frequency    Percent    Frequency     Percent
--------------------------------------------------------------------------------------------------------------------
      1           1           0           0           0           0          53      29.12           53       29.12
      2           0           1           0           0           0          66      36.26          119       65.38
      3           0           0           1           0           0          28      15.38          147       80.77
      4           0           0           0           1           0          20      10.99          167       91.76
      5           0           0           0           0           1          15       8.24          182      100.00
```

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                    09:23 Wednesday, December 23, 2015  27
Supplemental Tables for DRB Members
After swapping
Check creation of indicator variables for BORNUSA

The FREQ Procedure

                                                              Cumulative    Cumulative
BORNUSA    BORNUSA_1    BORNUSA_2    Frequency     Percent     Frequency      Percent
--------------------------------------------------------------------------------------
      1            1            0         162       89.01           162        89.01
      2            0            1          20       10.99           182       100.00
```

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                                                   09:23 Wednesday, December 23, 2015  28
Supplemental Tables for DRB Members
After swapping
Check creation of indicator variables for CENREG

The FREQ Procedure

                                                                                   Cumulative    Cumulative
CENREG    CENREG_1    CENREG_2    CENREG_3    CENREG_4    Frequency     Percent     Frequency      Percent
-----------------------------------------------------------------------------------------------------------
     1           1           0           0           0          29       15.93            29        15.93
     2           0           1           0           0          44       24.18            73        40.11
     3           0           0           1           0          68       37.36           141        77.47
     4           0           0           0           1          41       22.53           182       100.00
```

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                 09:23 Wednesday, December 23, 2015  29
Supplemental Tables for DRB Members
After swapping
Check creation of indicator variables for GENDER

The FREQ Procedure

                                                            Cumulative    Cumulative
GENDER    GENDER_1    GENDER_2    Frequency      Percent     Frequency      Percent
-----------------------------------------------------------------------------------
     1           1           0          168        92.31           168        92.31
     2           0           1           14         7.69           182       100.00
```

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                09:23 Wednesday, December 23, 2015  30
Supplemental Tables for DRB Members
Unweighted Measures of Association in Swapping Variables and Key Output Variables
Where r(before)-r(after) is not 0

Variable     Variable       Unweighted         Unweighted
1            2            Before Swapping    After Swapping

BIB1201_1    DAGE            -0.00218           -0.02855
BIC0501_1    DAGE            -0.13292           -0.05801
BID0101_1    DAGE             0.07806            0.09919
BIE0601_1    DAGE            -0.00729           -0.00463
BIE0601_2    DAGE            -0.21793           -0.26059
BIE0601_3    DAGE             0.16769            0.11921
BIE0601_4    DAGE             0.05084            0.13763
BIE0601_5    DAGE             0.11520            0.15029
CENREG_1     DAGE            -0.10601           -0.10931
CENREG_3     DAGE             0.05490            0.07734
CENREG_4     DAGE             0.07954            0.05644
DAGE         DIC0401          0.37930            0.35435
DAGE         EDUC_DET         0.08507            0.09775
DAGE         GENDER_1         0.02455            0.02002
DAGE3        BIC0501_1       -0.14809           -0.08907
DAGE3        BID0101_1        0.11135            0.13251
DAGE3        BIE0601_1        0.00631            0.02758
DAGE3        BIE0601_2       -0.18069           -0.22089
DAGE3        BIE0601_3        0.11538            0.06181
DAGE3        BIE0601_4        0.05535            0.11714
DAGE3        BIE0601_5        0.09113            0.12628
DAGE3        CENREG_3         0.01778            0.03776
DAGE3        CENREG_4         0.04677            0.02364
DAGE3        DIC0401          0.32468            0.31442
DAGE3        EDUC_DET         0.06754            0.07699
DAGE3        GENDER_1         0.03069           -0.00558
DAGE3        SCORE           -0.08039           -0.06238
SCORE        DAGE            -0.10646           -0.08055
















Number of Pairwise Associations computed : 253
* denotes the after swapping correlation changed by more than 1.96 standard errors
Variable(s) 'DRACE3_* BIB1201_* BIC0501_* BID0101_* BIE0601_* BORNUSA_* CENREG_* GENDER_*' are
indicator variables
```

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                09:23 Wednesday, December 23, 2015  31
Supplemental Tables for DRB Members
Weighted Measures of Association in Swapping Variables and Key Output Variables
Where r(before)-r(after) is not equal to 0

Variable     Variable        Weighted           Weighted
1            2            Before Swapping    After Swapping

BIB1201_1    DAGE            0.00397           -0.02668
BIC0501_1    DAGE           -0.16214           -0.08651
BID0101_1    DAGE            0.06821            0.09352
BIE0601_1    DAGE            0.00649            0.01091
BIE0601_2    DAGE           -0.21811           -0.26202
BIE0601_3    DAGE            0.15329            0.09641
BIE0601_4    DAGE            0.03506            0.12544
BIE0601_5    DAGE            0.12730            0.16344
BORNUSA_1    DAGE            0.05151            0.05046
CENREG_1     DAGE           -0.11462           -0.11342
CENREG_2     DAGE           -0.04712           -0.05265
CENREG_3     DAGE            0.04973            0.07163
CENREG_4     DAGE            0.08860            0.06798
DAGE         DIC0401         0.36650            0.33809
DAGE         EDUC_DET        0.05384            0.07092
DAGE         GENDER_1        0.04994            0.04505
DAGE3        BIB1201_1      -0.03038           -0.03615
DAGE3        BIC0501_1      -0.16598           -0.11010
DAGE3        BID0101_1       0.10118            0.12417
DAGE3        BIE0601_1       0.02936            0.05049
DAGE3        BIE0601_2      -0.17566           -0.21445
DAGE3        BIE0601_3       0.09530            0.03313
DAGE3        BIE0601_4       0.03364            0.09881
DAGE3        BIE0601_5       0.09309            0.12901
DAGE3        BORNUSA_1       0.00653            0.00488
DAGE3        CENREG_1       -0.06654           -0.06149
DAGE3        CENREG_2       -0.02326           -0.02698
DAGE3        CENREG_3        0.01915            0.03565
DAGE3        CENREG_4        0.05853            0.03896
DAGE3        DAGE            0.87329            0.87405
DAGE3        DIC0401         0.31296            0.30175
DAGE3        EDUC_DET        0.03914            0.05291
DAGE3        GENDER_1        0.06116            0.02792
DAGE3        SCORE          -0.07921           -0.05759
DRACE3_1     DAGE            0.03252            0.03406
DRACE3_1     DAGE3           0.08730            0.08945
DRACE3_2     DAGE           -0.01220           -0.01635
DRACE3_2     DAGE3          -0.04099           -0.04676
DRACE3_3     DAGE           -0.01265           -0.00953
DRACE3_3     DAGE3          -0.02539           -0.02106
EDUC3        DAGE            0.10681            0.11036
EDUC3        DAGE3           0.08359            0.08837
SCORE        DAGE           -0.10997           -0.07809










Number of Pairwise Associations computed : 253
* denotes the after swapping correlation changed by more than 1.96 standard errors
Variable(s) 'DRACE3_* BIB1201_* BIC0501_* BID0101_* BIE0601_* BORNUSA_* CENREG_* GENDER_*' are
indicator variables
```

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                  09:23 Wednesday, December 23, 2015  32
Supplemental Tables for DRB Members
Unweighted Multiple Regression Coefficients for model KEYOUT = INTERCEPT SWAPVARS

                                         Estimate    Estimate
Dependent                                 Before      After
Variable     Parameter                   Swapping    Swapping

  SCORE      Error degrees of freedom    180.0000    180.0000
             Intercept                   253.8102    250.7130
             DAGE3                        -8.8596     -6.8748
             R-squared                      0.0065      0.0039
```

```
* denotes the after swapping beta coefficient changed by more than 1.96 standard errors
```

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                  09:23 Wednesday, December 23, 2015  33
Supplemental Tables for DRB Members
Weighted Multiple Regression Coefficients for model KEYOUT = INTERCEPT SWAPVARS

                                        Estimate    Estimate
Dependent                                Before      After
Variable     Parameter                  Swapping    Swapping

  SCORE      Error degrees of freedom    180.0000    180.0000
             Intercept                   251.0132    247.1533
             DAGE3                        -8.8040     -6.3957
             R-squared                     0.0063      0.0033
```

```
* denotes the after swapping beta coefficient changed by more than 1.96 standard errors
```

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                    09:23 Wednesday, December 23, 2015  34
Supplemental Tables for DRB Members
Unweighted Multiple Regression Coefficients for User-specified Models

                                        Estimate    Estimate
Dependent                                Before      After
Variable     Parameter                  Swapping    Swapping

  SCORE      Error degrees of freedom    175.0000    175.0000
             Intercept                   231.9791    227.0320
             CENREG 1                    -27.2338    -26.3645
             CENREG 2                     -5.9335     -5.2407
             CENREG 3                     -9.8158     -9.0686
             EDUC_DET                     13.7469     13.7539
             DAGE                         -1.0904     -0.9419
             GENDER 1                     -7.6383     -8.0351
             R-squared                     0.2399      0.2336
```

```
* denotes the after swapping beta coefficient changed by more than 1.96 standard errors
```

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                09:23 Wednesday, December 23, 2015  35
Supplemental Tables for DRB Members
Weighted Multiple Regression Coefficients for User-specified Models

                                        Estimate    Estimate
Dependent                                Before      After
Variable     Parameter                  Swapping    Swapping

  SCORE      Error degrees of freedom    175.0000    175.0000
             Intercept                   229.4148    223.0960
             CENREG 1                    -32.0225    -30.9431
             CENREG 2                     -7.3854     -6.7667
             CENREG 3                    -12.6968    -11.9845
             EDUC_DET                     13.5072     13.5429
             DAGE                         -1.0575     -0.8656
             GENDER 1                     -2.8653     -3.4780
             R-squared                     0.2166      0.2093
```

```
* denotes the after swapping beta coefficient changed by more than 1.96 standard errors
```

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                    09:23 Wednesday, December 23, 2015  36
Supplemental Tables for DRB Members
Data Utility Measure for Tables
SWAPVARS=DAGE3

                                                   For         Utility     Cell Count/
                Application                    Variable(s)      value      Small Cell

Hellinger's Distance, all cells                Across All     0.714367!        3/1
Hellinger's Distance, excluding small cells    Across All     0.714367         2/0
Hellinger's Distance, all cells                DAGE3          0.714367!        3/1
Hellinger's Distance, excluding small cells    DAGE3          0.714367         2/0
```

```
! denotes small cell(s)
```

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                 09:23 Wednesday, December 23, 2015  37
Supplemental Tables for DRB Members
Data Utility Measure for Pairwise Associations
Including BOUNDARY, SWAPVARS and KEYVARS

                                     Utility
Application                           value

Pearson's Product Correlation        0.30731
Pearson's Contingency Coefficient    0.06290
Cramer's V                           0.06810
```

Appendix B-2. *DataSwap* Example 2, run 1 output – hard boundary (continued)

```
DATASWAP   EXAMPLE2.SAS                                 09:23 Wednesday, December 23, 2015  38
Supplemental Tables for DRB Members
Data Utility Measure for Regression Coefficients - Default and User-defined models
KEYOUT=SCORE
SWAPVARS=DAGE3

                                     Utility
Model                                 value

SCORE=DAGE3                          0.14190
SCORE=CENREG EDUC_DET DAGE GENDER    0.10106
Across All Regression Models         0.12148
```

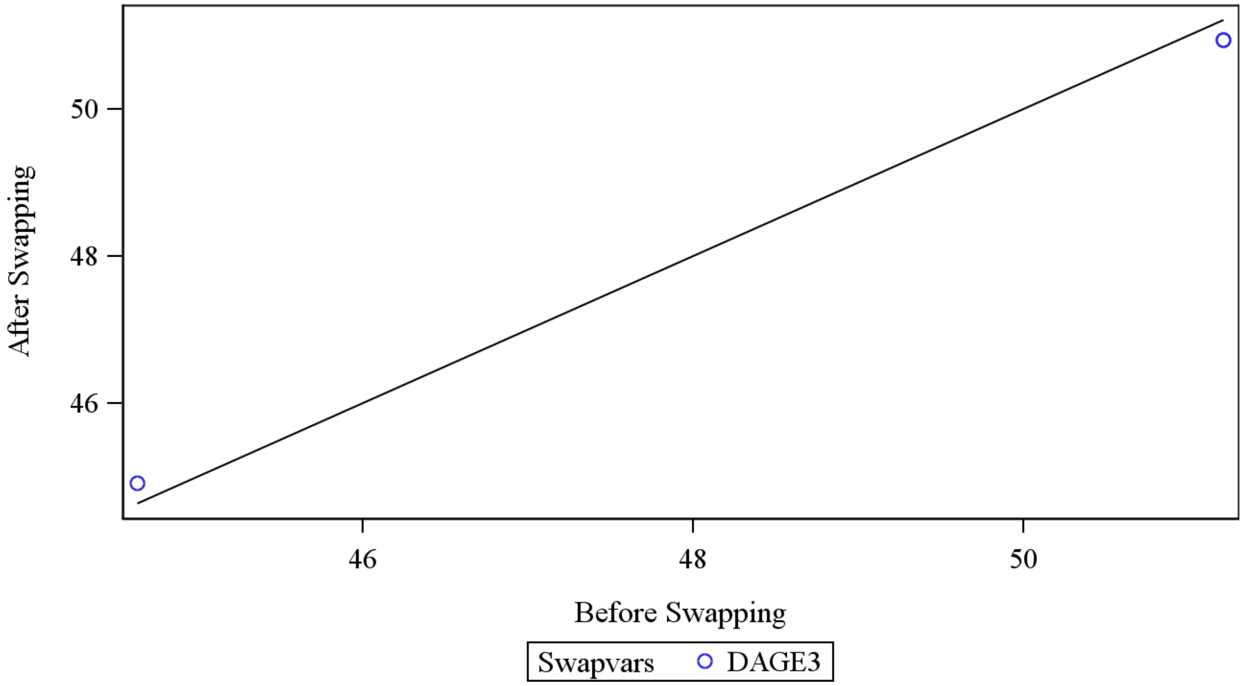
Appendix B-3. *DataSwap* Example 2, run 1 graphs – hard boundary

DATASWAP EXAMPLE2.SAS

Project # 6107.03.29.01

Graphs for DRB Members

Weighted Percents Before and After Swapping

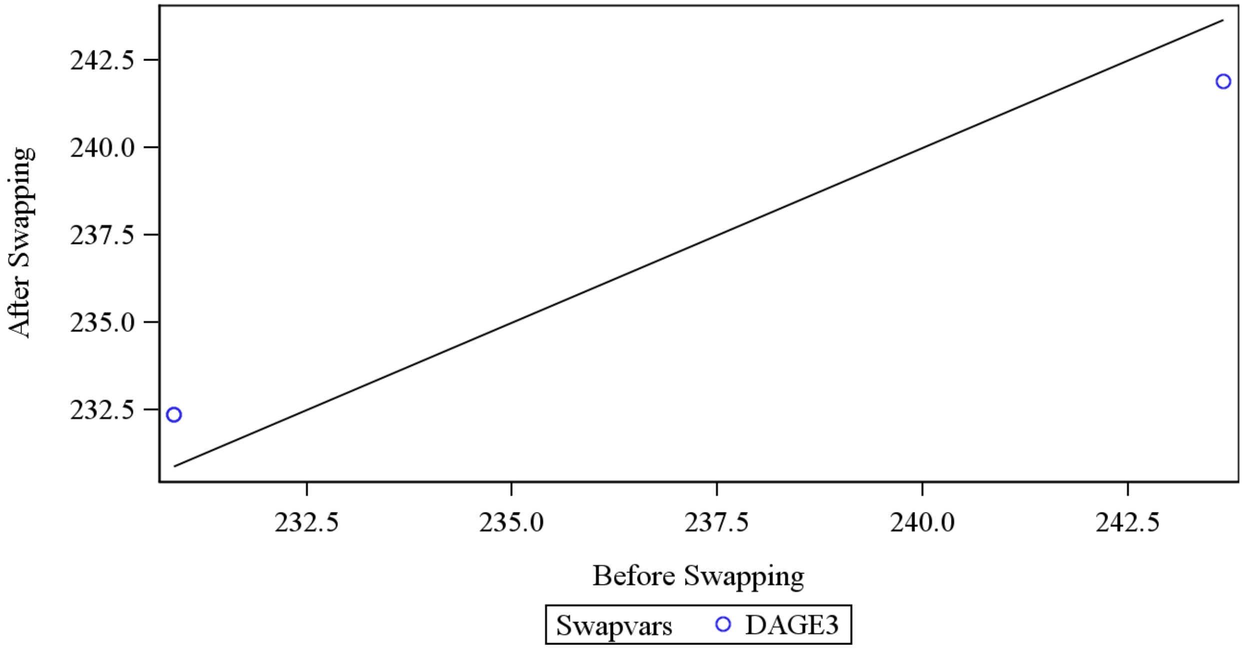


Symbols were shown only if cell sample sizes exceed 45

Z:\Users Guide\Version 32 new\runs for guide\Example 2\EXAMPLE2 PLOT PCT Run1.rtf

Appendix B-3. *DataSwap* Example 2, run 1 graphs – hard boundary (continued)

DATASWAP EXAMPLE2.SAS
Project # 6107.03.29.01
Graphs for DRB Members
Weighted Means Before and After Swapping
KEYOUT=SCORE



Symbols were shown only if cell sample sizes exceed 45
Z:\Users Guide\Version 32 new\runs for guide\Example 2\EXAMPLE2 PLOT MEAN Run1.rtf

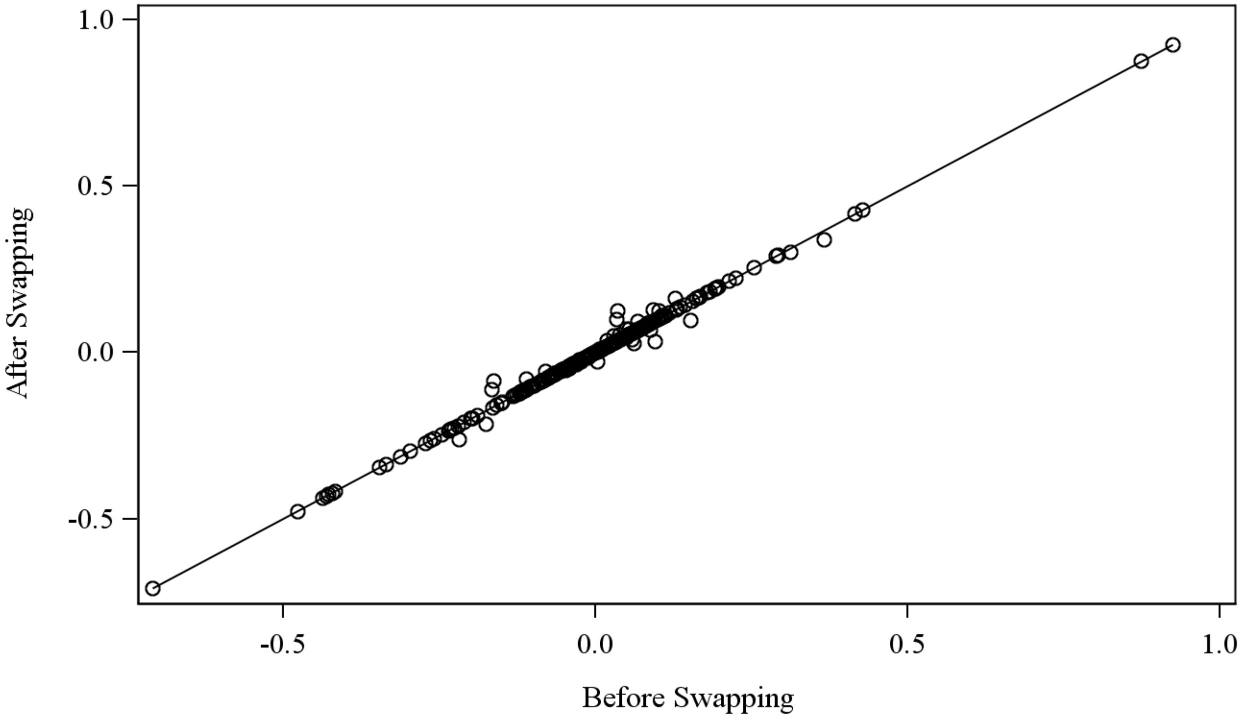
Appendix B-3. *DataSwap* Example 2, run 1 graphs – hard boundary (continued)

DATASWAP EXAMPLE2.SAS

Project # 6107.03.29.01

Graphs for DRB Members

Weighted Measures of Association Before and After Swapping

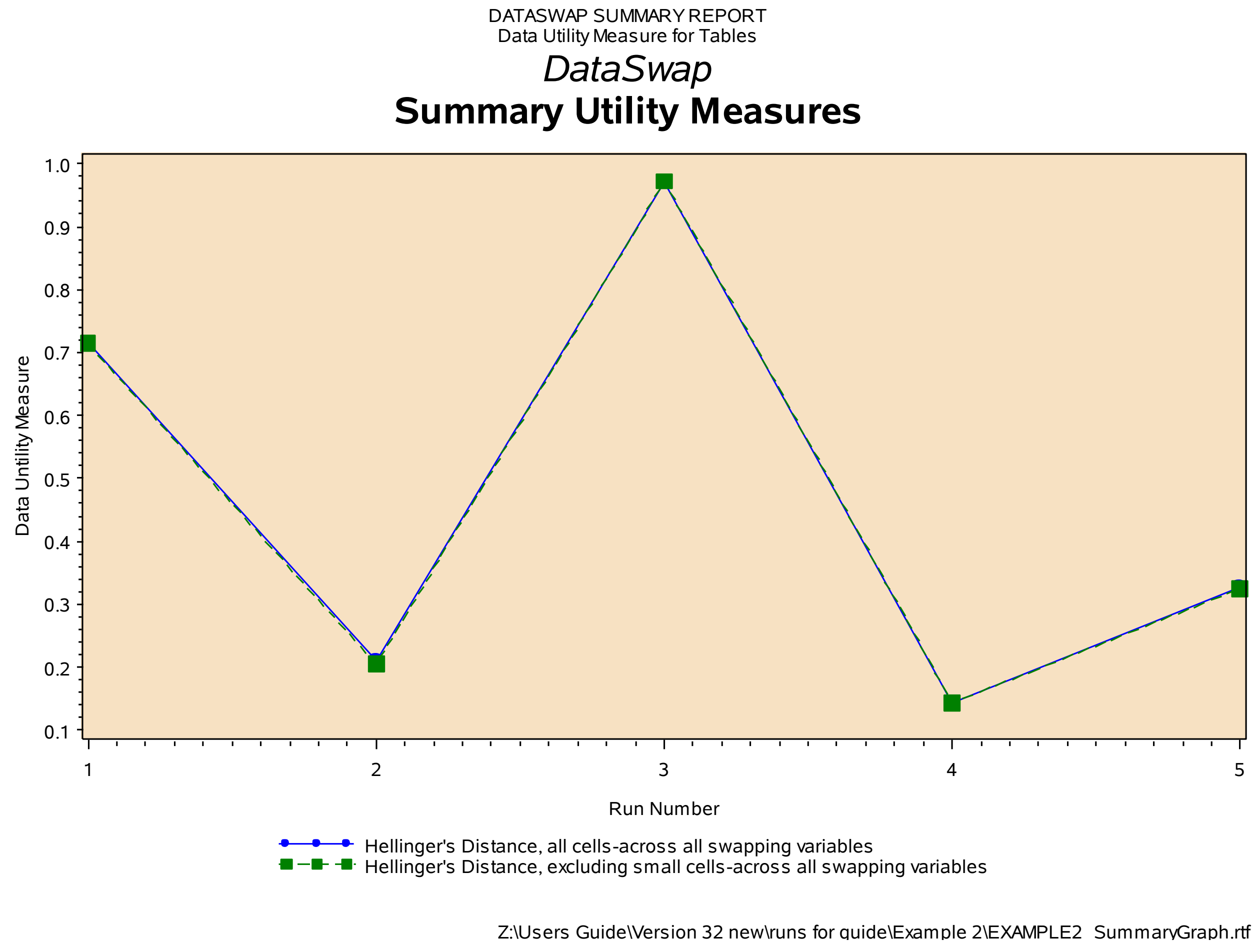


Z:\Users Guide\Version 32 new\runs for guide\Example 2\EXAMPLE2 PLOT CORR Run1.rtf

Appendix B-4. *DataSwap* Example 2 summary output – hard boundary

```
DATASWAP SUMMARY REPORT                                                         09:23 Wednesday, December 23, 2015  39
Data Utility Measure for Tables

                                                For       Run 1      Run 2     Run 3     Run 4     Run 5   Cell Count/
Application                                 Variable(s) Seed=22061 Seed=345   Seed=76   Seed=98  Seed=239  Small Cell

Utility Measures for Tables

Hellinger's Distance, all cells             Across All  0.714367!  0.211131! 0.972131! 0.142631! 0.327048!     3/1
Hellinger's Distance, excluding small cells Across All  0.714367   0.204631  0.972131  0.142631  0.324320      2/0
Hellinger's Distance, all cells             DAGE3       0.714367!  0.211131! 0.972131! 0.142631! 0.327048!     3/1
Hellinger's Distance, excluding small cells DAGE3       0.714367   0.204631  0.972131  0.142631  0.324320      2/0


Utility Measures for Pairwise Associations

Pearson's Product Correlation                           0.307309   0.215874  0.232436  0.262020  0.226633
Pearson's Contingency Coefficient                       0.062899   0.080739  0.088904  0.075099  0.047888
Cramer's V                                              0.068101   0.081967  0.088213  0.079012  0.049909


Utility Measures for Multivariate Assoc.

SCORE=DAGE3                                             0.141902   0.248226  0.069406  0.116341  0.061478
SCORE=CENREG EDUC_DET DAGE GENDER                       0.101058   0.158896  0.030017  0.067343  0.057482
Across All Regression Models                            0.121480   0.203561  0.049711  0.091842  0.059480
```

Appendix B-5. *DataSwap* Example 2 summary graphs – hard boundary

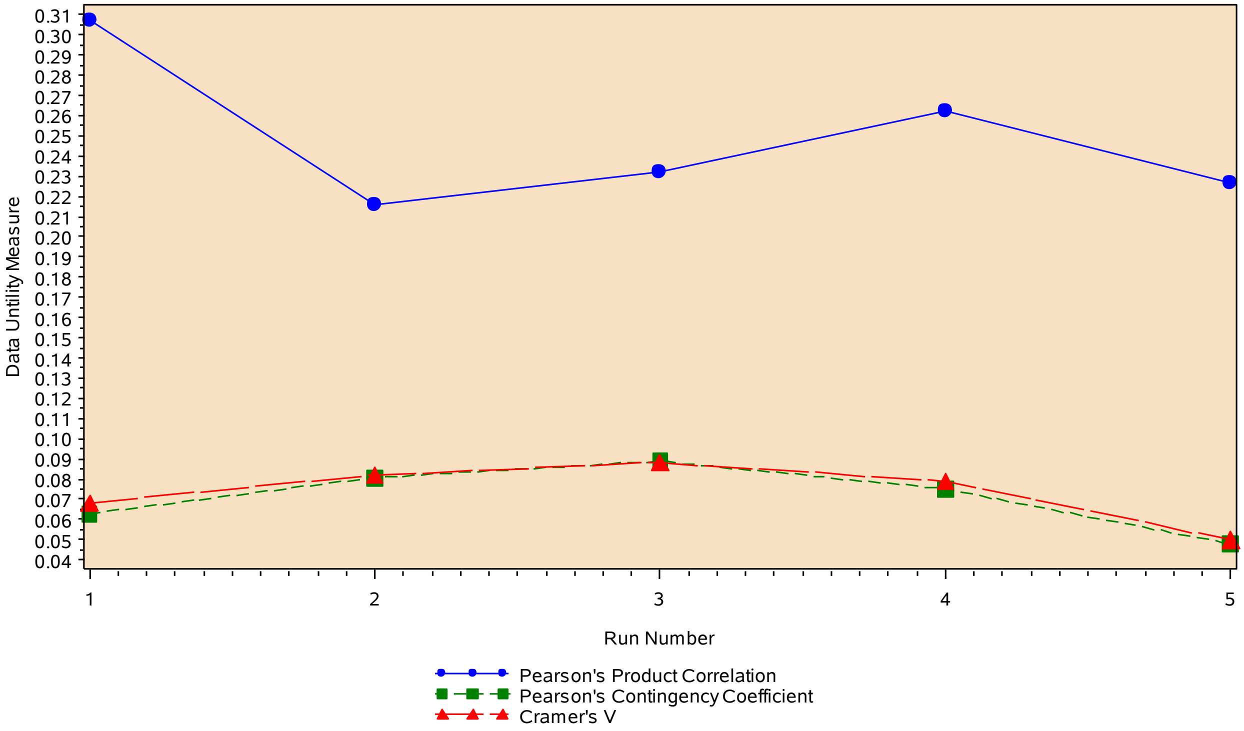


Appendix B-5. *DataSwap* Example 2 summary graphs – hard boundary (continued)

DATASWAP SUMMARY REPORT
Data Utility Measure for Tables

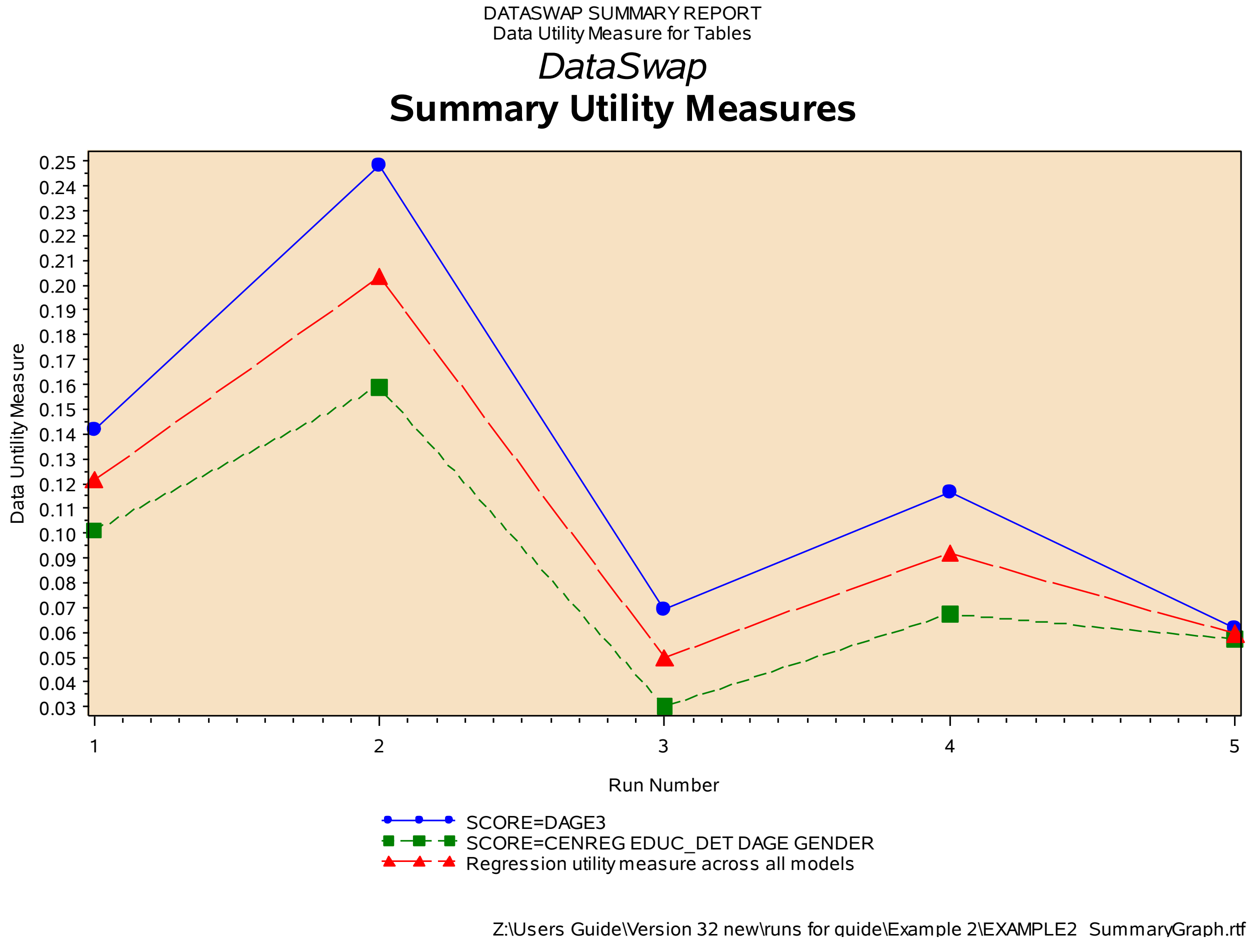
DataSwap

Summary Utility Measures



Z:\Users Guide\Version 32 new\runs for guide\Example 2\EXAMPLE2_SummaryGraph.rtf

Appendix B-5. *DataSwap* Example 2 summary graphs – hard boundary (continued)



DATASWAP SUMMARY REPORT
Data Utility Measure for Tables

DataSwap

Summary Utility Measures

Z:\Users Guide\Version 32 new\runs for guide\Example 2\EXAMPLE2_SummaryGraph.rtf