

Paper 194-31

The Enigma of Survey Data Analysis: Comparison of SAS® Survey Procedures and SUDAAN Procedures

Katherine L. Baisden, SRI International, Menlo Park, California
Paul Hu, SRI International, Menlo Park, California

ABSTRACT

Stratified, clustered, or complex sample designs involved in most survey research impact the accuracy of variance estimates and test statistics. Until recently, SAS programmers had to rely on other statistical software packages, such as SUDAAN, WesVar, and STATA, to produce accurate estimates of such statistics. With the release of SAS/STAT® Version 9.1, there is now a survey module (e.g. PROC SURVEYMEANS, PROC SURVEYFREQ, PROC SURVEYREG and PROC SURVEYLOGISTIC) to address this issue. This paper examines these four basic procedures widely used in survey research, focusing on the differences between the SAS® procedures and the corresponding SAS-callable SUDAAN® V9.0 procedures. It will illustrate differences in syntax and output using data examples and it will also discuss the available options, limitations and recent updates of each package.

INTRODUCTION

To optimize the process of survey data collection, researchers continue to develop increasingly complex sample designs. These designs include stratification, clustering, unequal probabilities of selection, and a multitude of the combinations of all these techniques. Such complex designs impact the accuracy of variance estimates and test statistics. Until recently, SAS programmers had to use additional software packages, such as SUDAAN, to produce correct variance estimates. Now, with SAS survey modules, some of the tools needed to deal with this type of survey data are available in SAS software.

This paper compares and contrasts SAS/STAT Version 9.1 and SAS-callable SUDAAN software Version 9.0 in regards to syntax and output of four of the most common procedures: the cross-tabulation/frequency, means, regression, and logistic regression procedure. This will be demonstrated using data from a study of teachers in the state of California Schools. The study used a stratified sample with replacement design. Weights were developed for the data based on these stratification variables and teachers were then selected from each of the strata. The examples highlight the syntax features (not all options are included) for each of the procedures in SAS-callable SUDAAN and SAS software.

GENERAL COMMENTS ABOUT SAS AND SAS-CALLABLE SUDAAN

Both SAS and SUDAAN procedures are based on the Taylor linear approximation method to calculate the variance estimates. However, even though SUDAAN software does not have the capability to calculate balanced repeated replicates (BRR) or jackknife weights, it does offer the option of using these weights if they are provided in the data. SAS software, on the other hand, does not have this option. The SAS-callable version of SUDAAN is designed to use within the framework of SAS and it uses the SAS dataset format. Thus, much of the syntax of a procedure is similar. However, there are some important differences.

DETERMINING THE IMPACT OF SAMPLE DESIGNS ON VARIANCE ESTIMATES

An often used measure to gauge the impact of complex sample designs (CSD) on variance estimates is called the Design Effect (DE). DE is a ratio. It takes the variance from the CSD and compares it to the variance that would have occurred under the assumption of simple random sampling (SRS). If the DE is close to 1.0, then you can assume that the variances would remain the same whether it was a CSD or a SRS design. The DE for a CSD is usually greater than one. The larger the DE, the more correlated are your respondents within clusters, leading to underestimated variances if the database is analyzed with packages without the capabilities to go beyond the assumption of SRS.

Exhibit 1 shows a breakdown of the impact of point and variance estimates when you use weighted data in a complex sample design with various types of SAS procedures and SAS options (unweighted or weighted) and SUDAAN procedures.

	Point Estimates (Percents, Means, Etc.)	Variances (Standard Errors, Variances)
Unweighted Regular SAS Procedures	Incorrect	Incorrect
Weighted Regular SAS Procedures	Correct	Incorrect
SAS Proc Survey Procedures (Weighted)	Correct	Correct
SUDAAN Procedures (Weighted)	Correct	Correct

Exhibit 1. Impact of Point and Variance Estimates

The point estimates will be the same for weighted regular SAS procedures, SAS Proc Survey procedures, and SUDAAN procedures within rounding. The variance estimates will be the same for SAS Proc Survey procedures and SUDAAN procedures. Using unweighted regular SAS procedures produces incorrect point estimates and variance estimates. Using weighted regular SAS procedures creates correct point estimates but incorrect variance estimates. There may be a slight difference between the two programs because there are slight differences in computation and the handling of missing data. For example, different estimates and standard errors may be due to the different tolerances for matrix inversion or the number of iterations in regression procedures.

SAS assumes that first-stage sampling is with replacement although in reality that is often not the case. Sampling with replacement can result in a slight overestimate of the variance. SUDAAN, on the other hand, offers you a choice in determining what kind of sampling design the survey is based on:

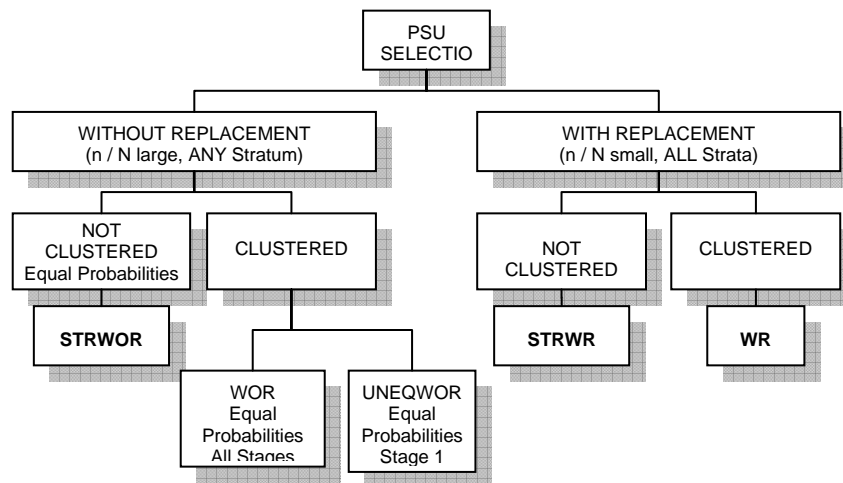


Exhibit 2. SUDAAN Flowchart

SAS and SUDAAN offer the following procedures:

SAS	SUDAAN	Purpose
	RECORDS	Print records from ASCII, SAS®, SPSS and SUDAAN
SURVEYFREQ	CROSSTAB	Produces weighted one-way and multi-way frequencies
	RATIO	Produces ratio estimates and their standard errors for correlated data
SURVEYMEANS	DESCRIPT	Produces means, medians and quantiles and their standard errors
SURVEYREG	REGRESS	Fits linear models
SURVEYLOGISTIC	RLOGIST	Fits logistic regression models
	MULTILOG	Fits logistic model with categorical dependent variables
	SURVIVAL	Fits the discrete proportional hazards model
SURVEYSELECT		Helps you select a sample

Exhibit 3. SAS and SUDAAN Procedures

A CLASS variable in SUDAAN can take on any values but must be numeric. A SAS CLASS variable can be either numeric or character.

Unlike SAS, SUDAAN has no default printing of output. You must specify exactly what statistic you want printed and in what format. SUDAAN also does not provide the variable names in the output unless they are specified in the label of the variable.

PROC SURVEYMEANS IN SAS

```
PROC SURVEYMEANS; VAR T4B; STRATA EMERG DISTSIZE SCHL_LVL; WEIGHT WGTD;
DOMAIN T40; RUN;
```

This analysis is requesting the overall mean of T4B (number of classes taught) and the mean for number of classes taught for each gender (T40). The stratification variables are EMERG, DISTSIZE, and SCHL_LVL as indicated on the STRATA statement. The DOMAIN statement indicates a breakdown of T4B by gender. SAS provides the NOBS, MEAN, STDERR and CLM statistics by default. A LIST option will provide basic information about the respondents in each stratum (N, number of missing, strata variable levels). Calculated design effects are not available in this procedure. If you want to calculate design effects, you need to run the same analysis as a normal weighted SAS MEANS procedure and then run it again as a PROC SURVEYMEANS. You then take the results from the two procedures and apply the DEFF formula ($DEFF = CDS \text{ Variance} / SRS \text{ Variance}$).

As in the PROC MEANS, when computing statistics for an analysis variable, SAS omits observations with missing values for that variable. In addition, it is important to note that in PROC SURVEYMEANS, if an observation has a missing value or nonpositive value for the weight, it will be excluded from the analysis. Observations are also excluded if there are missing values on the STRATA or CLUSTER statement, unless the MISSING option is used. When the MISSING option is used, the missing values are treated as a valid category. Output is shown in Exhibit A-1 in the appendix.

PROC DESCRIPT IN SUDAAN

```
PROC DESCRIPT DATA=ONE FILETYPE=SAS DESIGN=STRWR;
NEST EMERG DISTSIZE SCHL_LVL; WEIGHT WGTD; VAR T4B; CLASS T40;
SETENV LABWIDTH=28 COLSPCE=1 COLWIDTH=10 DECWIDTH=4;
PRINT NSUM="SAMPLE SIZE" WSUM="POPULATION SIZE" MEAN SEMEAN="S.E."
DEFFMEAN="DESIGN EFFECT" / STYLE=NCHS NSUMFMT=F6.0 WSUMFMT=F10.0
DEFFMEANFMT=F6.2 SEMEANFMT=F7.4;
RTITLE "MEAN OF T4B IN SUDAAN"; RUN;
```

This analysis is the same as presented earlier in the PROC SURVEYMEANS. The STWR design was used to correspond with the SAS assumptions. You must specify the file-type and the design. The NEST statement is similar to the STRATA statement in SAS. The SUBGROUP statement corresponds to the DOMAIN statement in SAS. In contrast to SAS, design effects can be requested. You have to execute the PROC DESCRIPT twice in order to get an overall mean of T4B (without the class variable) and the separate means for T4B by gender (with the class variable). Output shown in Exhibit A-2 in the appendix.

The SETENV statement sets the output environment parameters, similar to the options statement in SAS. The PRINT statement allows you to specify each statistic and a label for those statistics as appeared in the output. The STYLE option is a particular way SUDAAN prints the output. Output is printed in NCHS style, according to the standards of the National Center for Health Statistics. Format for each statistic must also be specified. The RTITLE statement is equivalent to the TITLE statement in SAS. You need to customize your SETENV and PRINT statement to set the statistics for each procedure individually. If you get **** in your output, that means that the format for a particular statistic or other information is not large enough.

SUDAAN handles missing values in a similar manner to SAS. Observations that have missing values for weights and required sample design variables will be excluded from the analysis. With the new CLASS statement you have the option of including missing values in your analysis.

In both programs the point estimates and the standard errors are the same (within reasonable rounding error). Unlike SAS, SUDAAN does not provide an option to obtain standard deviations.

PROC SURVEYFREQ IN SAS

```
PROC SURVEYFREQ; STRATA EMERG DISTSIZE SCHL_LVL;
TABLES T40*T6 / CHISQ WCHISQ ROW COL CHISQ1; WEIGHT WGTD; RUN;
```

This is a simple cross tabulation. When requesting a chi-square analysis with this procedure, you obtain a Rao-Scott chi-square test (CHISQ option), which applies a design effect correction to the Pearson chi-square that computes the design effect correction from proportion estimates instead of null proportions. The CHISQ1 option will give you a modified Rao-Scott chi-square test. The modified Rao-Scott chi-square bases the design effect correction on null hypothesis proportions. However, there is a known defect in the procedure which produces an incorrect Rao-Scott chisq if the WEIGHT statement is used. To correct this problem, you must normalize the weights (weights scaled so their sum equals total sample size). The WCHISQ is an option in the Tables statement that gives you a Wald chi-square. The default options are frequencies, weighted frequencies, standard error of the weighted frequencies, percentages, and standard error of the percentages. You must indicate specifically that you want row and column percentages and their standard errors; they are not given by default. Theoretically, the point estimates will not significantly differ from the SUDAAN output. Any observed differences can usually be accounted for by rounding errors. Output shown in Exhibit A-3 in the appendix.

PROC SURVEYFREQ excludes an observation from a cross-tabulation table if that observation has a missing value for any of the table, weight, or required sample design variables unless you specify the MISSING option. In contrast to PROC FREQ, you cannot specify a MISSPRINT option that gives the number of missing observations in each cell and still not include the missing values in the calculations of the percentages and other statistics.

PROC CROSSTAB IN SUDAAN

```
PROC CROSSTAB DATA=ONE FILETYPE=SAS DESIGN=STRWR;
NEST EMERG DISTSIZE SCHL_LVL; WEIGHT WGTD; CLASS T40 T6;
TABLES T40*T6; RUN;
```

The PROC CROSSTAB in SUDAAN follows the logic of the syntax presented in the PROC DESCRIPT. SUDAAN will produce several types of chi-square tests including the Cochran-Mantel-Haenszel and the Pearson chi-square. Output shown in Exhibit A-4 in the appendix.

The cross-tabulation output prints out the totals on the left, reversed from the traditional SAS output. One of the disadvantages of SUDAAN output is that it produces a single page for every table and every test statistic you request. In addition, SUDAAN in the default mode does not provide a count for the missing values for each table cell.

PROC SURVEYREG IN SAS

```
PROC SURVEYREG; STRATA EMERG DISTSIZE SCHL_LVL / LIST;
CLASS T40 ; MODEL T36=T40 T41 / ANOVA DEFF ADJRSQ SOLUTION ;
WEIGHT WGTD; RUN;
```

This procedure performs linear regression taking into account the survey design variables. The dependent variable must be continuous (or assume so) and the independent variables can be either continuous or categorical. Any categorical variable on the model statement must appear on the CLASS statement. In addition, the CLASS statement must precede the model statement. PROC SURVEYREG forms dummy indicator variables (coded 1 or 0) for categorical independent variables with the highest coded value of variable defined as reference group. By specifying ANOVA, you will get a traditional ANOVA table. You have the ability to specify DEFF to get the design effects, which is important for understanding how the stratification or clustering sampling frame affected your data. SAS produces an estimated regression coefficient table by default if there is no CLASS statement. If you have a CLASS statement in your code, you must provide a SOLUTION option on the MODEL statement to produce the estimated regression coefficient table. To match the parameters of the SUDAAN procedure, we ran the analysis to include the intercept. Output in Exhibit A-5 in the appendix.

If an observation has a missing value or a nonpositive value for the WEIGHT variable, then PROC SURVEYREG excludes that observation from the analysis. An observation is also excluded if it has a missing value for any STRATA

variable, CLUSTER variable, dependent variable, or any variable used in the independent effects. The analysis includes all observations in the data set that have non-missing values for all these design and analysis variables.

PROC REGRESS IN SUDAAN

```
PROC REGRESS DATA=ONEA FILETYPE=SAS DESIGN=STRWR;
  NEST EMERG DISTSIZE SCHL_LVL; WEIGHT WGTD; CLASS T40;
  TEST SATADJCHI ADJWALDF; MODEL T36=T40 T41; RUN;
```

PROC REGRESS in SUDAAN is similar to SAS SURVEYREG except for the fact that there is no option for obtaining a traditional ANOVA table directly. Comparable pieces of information have been highlighted in Exhibit A-5. The R-square in SAS PROC SURVEYREG output is an adjusted multiple R-square and is found in the SUDAAN output as well. The information in the “Tests of Model Effects” table found in the SAS PROC SURVEYREG output can be found in the “Contrast” table in the SUDAAN output. The beta coefficients and their standard errors are found in the “Estimated Regression Coefficient” table in the SAS PROC SURVEYREG output and the “Independent Variables and Effects” table in SUDAAN. We have chosen to run the PROC SURVEYREG and PROC SURVEYLOGISTIC with an intercept. The NOINT (no intercept) option in both of these procedures uses the uncorrected sum of squares as opposed to an intercept option which uses a corrected sum of squares.

Currently in the SUDAAN modeling procedures, a record with missing values for any of the model variables is excluded from the analysis. With the new CLASS statement, records with missing values can now be included in the analysis provided the variable names are listed in the CLASS statement and INCLUDE=MISSING is used. The default is NOMISSING. SAS does not have this option.

PROC SURVEYLOGISTIC IN SAS

```
PROC SURVEYLOGISTIC; STRATUM EMERG DISTSIZE SCHL_LVL;
  MODEL T42AB(EVENT='1')=T41 T40 / STB RSQ ;
  WEIGHT WGTD; RUN;
```

This procedure performs a logistic regression taking into account the survey design variables. The dependent variable can be binary (0,1) or ordinal (small, medium, large) and the independent variables can be either continuous or categorical. When you have a binary dependent variable, you have the capability to determine which category you would like to be the “event” category in the model statement. The option RSQ on the MODEL statement will give you a generalized R square for the fitted model. Output in Exhibit A-6 in the appendix.

In PROC SURVEYLOGISTIC, any observation with missing values for the response, offset or explanatory variables or any required sample design variable is excluded from the analysis. The estimated linear predictor, its standard error estimate, the fitted probabilities, and their confidence limits are not computed for any observation with missing offset or explanatory variable values. The MISSING option can be used in the same manner as with PROC SURVEYFREQ and PROC SURVEYMEANS.

PROC RLOGIST IN SUDAAN

```
PROC RLOGIST DATA=ONEA FILETYPE=SAS DESIGN=STRWR;
  NEST EMERG DISTSIZE SCHL_LVL; WEIGHT WGTD; CLASS T40;
  MODEL T42AB=T41 T40; TEST SATADJCHI WALDCHI; RUN;
```

There are several procedure names in SAS-callable SUDAAN that are similar to SAS syntax. In order to not create confusion for SAS, SUDAAN has used a naming convention to start such procedures with the letter R. The syntax of this procedure is comparable to the PROC REGRESS.

Information about the response (dependent) variable is found in the “Response Profile” in the SAS PROC SURVEYLOGISTIC output (Exhibit A-6) and the same information is found in the “Sample and Population Counts Table” in the SUDAAN PROC RLOGIST output. Beta coefficients and standard errors are found in the “Analysis of

Maximum Likelihood Estimates (SAS Output)” table and the “Independent Variables and Effects (SUDAAN Output)” table in Exhibit A-6. Each package produces odds ratios.

LIMITATIONS OF EACH PACKAGE

One of the major limitations in SAS is the package does not offer the option of using balanced repeated replicates (BRR) or jackknife weights. Why is this so important? It is common as a programmer/analyst to inherit data sets or secondary dataset analysis. In many cases you do not have access to the actual information of the sampling design. It is essential, especially in SUDAAN, to be able to designate the sample design based on such information. With the use of balanced repeated replicates or jackknife weights, the syntax does not require any further information other than the supplied weights. Designation of the design makes it much more usable for the analyst.

Although SUDAAN offers more options in terms of survey sampling designs and procedures, it is a cumbersome program to code. SUDAAN documentation is not the easiest to comprehend, especially if you are a novice. From an economic viewpoint, using SUDAAN is an additional expense in terms of licensing and training. SAS gives you the ease of coding and more print control for output, especially when using the ODS system; however at this time SAS is very limited in what it offers an analyst in terms of design and procedures.

CONCLUSION

Simple random sampling is a rare gem in this day of social science research. Analysts are dealing with increasingly more complex sample designs. These designs require the sophistication of SAS survey procedures or SUDAAN procedures. You must balance variety of choice with ease of coding. At this time SUDAAN is the most desirable package to use because of the variety of choice it offers for sample designs and the number of procedures available to analyze the data. However, it is a package that is cumbersome to program, creating a more labor-intensive task than its counterpart SAS. The inclusion of the CLASS statement in SUDAAN version 9 resolves some of these issues. You will still have to deal with the specification of the print options in SUDAAN. SAS is moving in the right direction, and we hope to see it incorporate the power of SUDAAN in terms of choice and number of procedures in the future. The single most important SAS survey procedure to be included in SAS version 9.1 is PROC SURVEYFREQ. Contingency analysis is a mainstay of any data analysis. With the incorporation of these survey-based procedures in SAS, analysts look forward to greater ease in coding when dealing with complex sample designs.

In conclusion, because neither the SAS survey procedures nor SUDAAN provide all the tools necessary to analyze most survey research, thus analysts must use both products together.

ACKNOWLEDGEMENTS

We wish to thank the Center for Education Policy at SRI International. Special thanks go to Andrea Lash for her mentoring and support. We also want to thank Patrick Shields, director of the Center for Education Policy, for his continued support of programmers.

REFERENCES

- An, Anthony and Donna Watts. 1998 “New SAS Procedures for Analysis of Sample Survey Data.” Proceedings of the Twenty-Third Annual SAS® Users Group International (SUGI) Conference, Nashville, TN.
- Baisden, Katherine. 2003. “The Quandary of Survey Data: Comparison of SAS Procedures and SUDAAN Procedures.” Proceedings of the Eleventh Annual Western Users of SAS® Software Conference, San Francisco, CA.
- Baisden, Katherine. 2004. “The Continuing Quandary of Survey Data: Comparison of SAS Procedures and SUDAAN Procedures.”, Proceedings of the Twelfth Annual Western Users of SAS® Software Conference, Pasadena, CA.
- Cassell, David L. and AnnMaria Rousey. 2003. “Complex Sampling Designs Meet the Flaming Turkey of Glory.” Proceedings of the Twenty-Eighth Annual SAS® Users Group International (SUGI) Conference, Seattle, WA.
- Research Triangle Institute (2004). SUDDAN User’s Manual, Release 9.0, Research Triangle Park, NC: Research Triangle Institute.
- SAS® Institute, Inc (1999), SAS/STAT® User’s Guide, Version 8, Volumes 1,2,3, Cary, NC: SAS Institute Inc. 3884PP.

SAS® Institute, Inc (2002-2003), SAS9.1 Help and Documentation, "SURVEYFREQ and SURVEYLOGISTIC Procedures", Cary, NC: SAS Institute Inc.

CONTACT INFORMATION

Katherine Baisden
SRI International
333 Ravenswood Ave, BS381
Menlo Park, CA 94025
Phone: (650) 859-5944
Fax: (650) 859-3375
katherine.baisden@sri.com
www.sri.com/policy/cep

Paul Hu
SRI International
333 Ravenswood Ave, BS366
Menlo Park, CA 94025
Phone: (650) 859-4734
Fax: (650) 859-3375
Paul.hu@sri.com
www.sri.com/policy/cep

SAS and all other SAS Institute Inc. products or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and products names are trademarks of their respective companies.

APPENDIX

Statistics

Variable	N	Mean	Std Error of Mean	Lower 95% CL for Mean	Upper 95% CL for Mean
T4b	489	2.889167	0.058885	2.7345094	3.00488367

Domain Analysis: T40						
T40	Variable	N	Mean	Std Error of Mean	Lower 95% CL for Mean	Upper 95% CL for Mean
(1) FEMALE	T4b	361	2.402571	0.094611	2.21664496	2.58849614
(2) MALE	T4b	124	4.336014	0.216710	3.91014493	4.76188363

EXHIBIT A-1. SAS SURVEYMEANS Procedure

Variance Estimation Method: Taylor Series (STRWR) (OVERALL MEAN)
 Mean of T4b (# Classes Taught)
 by: Variable, One.

Variable	Sample Size	Population size	Mean	S.E.	Design effect
T4b	489	272615	2.8892	0.0636	0.46

Variance Estimation Method: Taylor Series (STRWR) (MEAN BY GENDER)
 #

Mean of T4b by T40 by: Variable, T40:GENDER.

Variable	T40:GENDER	Sample Size	Population size	Mean	S.E.	Design effect
T4b:TOTAL NUMBER OF CLASSES TAUGHT						
	Total	485	271054	2.8982	0.0641	0.47
	(1) FEMALE	361	201576	2.4026	0.0961	0.91
	(2) MALE	124	69478	4.3360	0.2167	1.78

EXHIBIT A-2. SUDAAN PROC DESCRIPT Procedure

T40	T6	Freq	Wted Freq	Std Wted	%	SE %	Row %	SE R
1 (Females)	1 (Yes)	16	10006	3742	4.03	1.51	5.4	1.99
	2 (No)	320	175435	7305	70.7	2.92	94.6	1.99
Total		336	185441	6922	74.76	2.76	100	
2 (Males)	1 (Yes)	9	2333	1053	.94	.423	3.73	1.69
	2 (No)	107	60248	6922	24.29	2.75	96.3	1.69
Total		116	62582	6943	25.33	2.76	100	
Total	1 (Yes)	25	12340	3885	4.97	1.56		
Total	2 (No)	427	235683	4580	95.02	1.56		
Total		452	248022	2576	100			

Frequency Missing = 39

Wald Chi-Square Test Chi-Square 0.4118 F Value 0.4118 Num DF 1
 Pr > F 0.5214 Sample Size = 471

EXHIBIT A-3. SAS SURVEYFREQ Procedure

Variance Estimation Method: Taylor Series (STRWR)
 Crosstab of T40 (GENDER) by T6 (PREP PGM)
 by: T40:GENDER, T6:LEAVE MA OR PREP PGM FOR FT PAID POSITION.

T40:GENDER		T6:LEAVE MA OR PREP PGM FOR FT PAID POSITION		
		Total	1 (YES)	2 (NO)
Total	Sample Size	452	25	427
	Weighted Size	248022	12340	235683
	Row Percent	100.00	4.98	95.02
	Tot Percent	100.00	4.98	95.02
	SE Row Percent	0.00	1.56	1.56
(1) FEMALE	Sample Size	336	16	320
	Weighted Size	185441	10006	175435
	Row Percent	100.00	5.40	94.60
	Tot Percent	74.77	4.03	70.73
	SE Row Percent	0.00	1.99	1.99
(2) MALE	Sample Size	116	9	107
	Weighted Size	62581	2333	60248
	Row Percent	100.00	3.73	96.27
	Tot Percent	25.23	0.94	24.29
	SE Row Percent	0.00	1.70	1.70

Chi-Square Test Chi-Square 0.4118 F Value 0.4118 Num DF 1
 Pr > F 0.5214 Sample Size = 471 (similar to Pearson's Chisq)

EXHIBIT A-4. SUDAAN PROC CROSSTAB

Due to space limitations, complete output for regression and logistic regression procedures are not be presented in their entirety. However, we will point out similarities of certain statistics. If you would like to see examples of complete output refer to the WUSS 2004 Conference Paper: Baisden, Katherine. 2004. "The Continuing Quandary of Survey Data: Comparison of SAS Procedures and SUDAAN Procedures.", Proceedings of the Twelfth Annual Western Users of SAS® Software Conference, Pasadena, CA. The "Estimated Regression Coefficients" Table in SAS contains the same information as the "Independent Variables and Effects" Table in SUDAAN.

Regression Model Statement T36 (Years Teaching)=T40(Gender M/F) + T41(Year of Birth)

Adjusted R-square 0.5181 (Same for both SAS and SUDAAN within rounding)

Estimated Regression Coefficients (Same for both SAS and SUDAAN within rounding)

Parameter	Estimate	Standard Error	t Value	Pr > t	Design Effect
Intercept	52.0321874	2.93655735	17.72	<.0001	2.99
T40 1 (fem)	-0.4235516	1.15500148	-0.37	0.7140	2.78
T40 2 (male)	0.0000000	0.00000000	.	.	.
T41(birthyr)	-0.6496751	0.04190682	-15.50	<.0001	2.20

ANOVA for Dependent Variable T36 (Only Available in SAS)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	13689069	6844535	269.19	<.0001
Error	497	12636761	25426		
Corrected Total	499	26325830			

Tests of Model Effects (Available in SAS output)

Effect	Num DF	F Value	Pr > F
Model	2	120.28	<.0001
Intercept	1	339.25	<.0001
T40	1	0.13	0.7140
T41	1	240.34	<.0001

The test of Model Effects table in SAS is similar to the Contrast Table in SUDAAN

Contrast (Available in SUDAAN output)

	DF	Adj DF	Chi-sq(SAT)	F-test(W-ALD)	P-Value(-SAT)	P-Value(-Wald-F)
OVERALL MODEL	3.00	2.220	812.26	352.23	0.0000	0.0000
MODEL MINUS INTERCEPT	2.00	1.953	210.99	121.24	0.0000	0.0000
INTERCEPT
T40	1.00	1.000	0.14	0.14	0.7125	0.7126
T41	1.00	1.000	242.76	242.76	0.0000	0.0000

Exhibit A-5. Highlights of SAS SURVEYREG and SUDAAN REGRESS Output

There is much information given in both the SAS and SUDAAN output. Again, if you would like to see complete output please refer to the WUSS 2004 paper: Baisden, Katherine. 2004. "The Continuing Quandary of Survey Data: Comparison of SAS Procedures and SUDAAN Procedures." Proceedings of the Twelfth Annual Western Users of SAS® Software Conference, Pasadena, CA. The same information is contained in the "Response Profile" (SAS) and the "Sample and Population Counts" (SUDAAN). To match information from the SAS "Analysis of Maximum Likelihood Estimate" table you must look at both the "Independent Variables and Effects" table and the "Contrast table" in SUDAAN.

Logistic Model Statement T42Ab(Master Degree Y/N)=T41(Gender M/F)+T40(Year of Birth)

Response Profile (SAS Output)

Ordered Value	T42Ab	Total Frequency	Total Weight
1	0	128	70499.60
2	1	161	100603.30

Sample and Population Counts Table (SUDAAN)

Sample and Population Counts for Response Variable T42AB

0: Sample Count	128	Population Count	70500
1: Sample Count	161	Population Count	100603

Analysis of Maximum Likelihood Estimates (SAS Output)

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate
Intercept	1	-0.3926	1.0329	0.1445	0.7039	
T41	1	-0.00049	0.0149	0.0011	0.9740	-0.0689
T40	1	0.6172	0.3843	2.5798	0.1082	3.6871

Independent Variables and Effects (SUDAAN Output)

	BETA	S.E.	DESIGN		P-VALUE
			EFFECT	T: BETA=0	
Intercept	0.842	0.924474	1.80	0.91	0.363
T41: YEAR OF BIRTH	-0.000	0.014905	1.69	-0.03	0.974
T40: GENDER					
(1) FEMALE	-0.617	0.392954	1.93	-1.57	0.117
(2) MALE	0.000	0.000000	.	.	.

Contrast Table (SUDAAN Output)

Contrast	DF	ADJ DF	CHI-SQ		P-VALUE (WALD)	P-VALUE (SAT.)
			(WALD)	(SAT.)		
OVERALL MODEL	3	2.97	6.48	6.75	0.090	0.079
MODEL MINUS INTERCEPT	2	1.99	2.48	2.64	0.289	0.266
INTERCEPT
T41	1	1.00	0.00	0.00	0.974	0.974
T40	1	1.00	2.47	2.47	0.116	0.116

Exhibit A-6. Highlights of SAS SURVEYLOGISTIC and SUDAAN RLOGIST Output