**Estimation Macros**

Elisabeth Neusy, June 22 2022

Updated February 16 2024 (macros were developed and tested in R and Python)

**Contents**

[1. Introduction 1](#_Toc106812840)

[2. Description of the Estimation Macros 2](#_Toc106812841)

[2.1 Compute\_FREQ: Proportions and Estimated Counts of Binary / Categorical Variables 2](#_Toc106812842)

[2.2 Compute\_MEAN: Means and Totals of Quantitative Variables 3](#_Toc106812843)

[2.3 Compute\_PERCENTILE: Percentiles of Quantitative Variables 4](#_Toc106812844)

[2.4 Compute\_DIFFPCT: Compares Proportions or Percents 5](#_Toc106812845)

[2.5 Compute\_DIFFMEAN: Compares means 6](#_Toc106812846)

[3. Main Program 7](#_Toc106812847)

# **Introduction**

General macros were developed for producing estimates using SAS or SUDAAN. Two versions of the macros were developed: one version uses SAS procedures, and the other version uses SUDAAN procedures. The macro calls are the same for both versions. We have observed that the execution time of the SUDAAN procedures is substantially shorter than the SAS procedures.

Five macros were developed:

1. Compute\_FREQ: produces proportions and estimated counts of binary/categorical variables
2. Compute\_MEAN: produces means and totals of quantitative variables
3. Compute\_PERCENTILE: produces percentiles of quantitative variables (e.g., medians)
4. Compute\_DIFFPCT: compares proportions or percents
5. Compute\_DIFFMEAN: compares means

The macros are saved in the following program files:

* SAS\_estimation\_macros.sas: contains the SAS version of the macros
* SUDAAN\_estimation\_macros.sas: contains the SUDAAN version of the macros

|  |
| --- |
| **DISCLAIMER:**  There are no guarantees; use these macros at your own risk. The macros were developed and tested using SAS 9.4 and SUDAAN 11. There are no plans for further development, such as adding features. As well, there are no plans for developing similar macros for other statistical packages, nor for other versions of SAS or SUDAAN. |

**IMPORTANT NOTE:** The content of SAS\_estimation\_macros.sas and SUDAAN\_estimation\_macros.sas should NOT be updated by the user.

The user should create a separate main program with the survey specific information and the macro calls. These general macros can be used for surveys using CV-based release rules, and for surveys using the release rules recommended by SSMD in 2021. Section 3 provides more information on the main program and on the release rules.

# **Description of the Estimation Macros**

This section provides information on each of the estimation macros in SAS\_estimation\_macros.sas and SUDAAN\_estimation\_macros.sas.

## **2.1 Compute\_FREQ: Proportions and Estimated Counts of Binary / Categorical Variables**

Compute\_FREQ produces proportions and estimated counts of binary/categorical variables. The SAS version of the macro uses PROC SURVEYFREQ to produce the estimates, and the SUDAAN version of the macro uses PROC CROSSTAB.

Macro call: %Compute\_FREQ (infile,outfile,var,group,where,keepall)

Description of the macro parameters:

* infile: input SAS dataset
* outfile: optional output SAS dataset for the results. The results are automatically appended to a SAS dataset called \_results\_.
* var: the categorical (or binary) variable of interest.
* group: optional list of categorical variables within which results are to be produced. The parameter should be left blank to produce estimates with no break down.
* where: optional where statement to define the subset of records of interest. The parameter should be left blank when all records are to be included.
* keepall: optional yes/no parameter to indicate whether to keep extra records in the output. Default is no.

The macro output includes the following:

* Percent: estimated percent for the category of the variable of the interest
* LOWPCT, UPPCT: 95% confidence interval of the estimated percent (using the Wilson method)
* ReleaseCategory\_pct: release category of the estimated percent (A,E or F)
* WSUM: estimated count for the category of the variable of interest
* LOWWSUM, UPWSUM: 95% confidence interval of the estimated count
* ReleaseCategory\_wsum: release category of the estimated count (A,E or F)
* n\_numerator: unweighted count of the number of respondents in the numerator of the percent
* n\_denominator: unweighted count of the number of respondents in the denominator of the percent

Note regarding the SUDAAN version of Compute\_FREQ: Ideally the CROSSTAB procedure should be called with the SMCONF option in order to construct confidence intervals for proportions with good properties. However, CROSSTAB has a bug which produces an error message in certain situations with the SMCONF option. We therefore dropped the SMCONF option, and instead added program code to construct confidence intervals outside of the CROSSTAB procedure.

## **2.2 Compute\_MEAN: Means and Totals of Quantitative Variables**

Compute\_MEAN produces means and totals of quantitative variables. The SAS version of the macro uses PROC SURVEYMEANS to produce the estimates, and the SUDAAN version of the macro uses PROC DESCRIPT.

Macro call: %Compute\_MEAN (infile,outfile,var,group,where,keepall)

Description of the macro parameters:

* infile: input SAS dataset
* outfile: optional output SAS dataset for the results. The results are automatically appended to a SAS dataset called \_results\_.
* var: the quantitative variable of interest.
* group: optional list of categorical variables within which results are to be produced. The parameter should be left blank to produce estimates with no break down.
* where: optional where statement to define the subset of records of interest. The parameter should be left blank when all records are to be included.
* keepall: optional yes/no parameter to indicate whether to keep extra records in the output. Default is no. (Note that the keepall parameter does not do anything for the SAS version of the macro because PROC SURVEYMEANS does not produce any extra records.)

The macro output includes the following:

* varname: variable of interest
* MEAN: mean of the variable of interest. The variable should be numeric.
* LOWMEAN, UPMEAN, CVMEAN: 95% confidence interval and CV of the mean
* ReleaseCategory\_mean: release category of the mean (A,E or F)
* TOTAL: total of the variable of interest
* LOWTOTAL, UPTOTAL, CVTOTAL: 95% confidence interval and CV of the total
* ReleaseCategory\_total: release category of the total (A,E or F)
* n: unweighted count of the number of respondents that contributed to the estimate
* n\_nonzero: unweighted count of the number of respondents with a nonzero value that contributed to the estimate

## **2.3 Compute\_PERCENTILE: Percentiles of Quantitative Variables**

Compute\_PERCENTILE produces percentiles, such as medians, of quantitative variables. We observed that PROC SURVEYMEANS produces results that are potentially slightly different from the usual method. We therefore developed our own SAS code for computing percentiles; it produces estimates that are more in-line with how a median is usually defined for estimates produced by a statistical agency: i.e., the middle value, taking the weights into account. The macro produces estimates of the median that are consistent with bootvar and G-TAB.

Macro call: %Compute\_PERCENTILE (infile,outfile,var,group,where,perc)

Description of the macro parameters:

* infile: input SAS dataset
* outfile: optional output SAS dataset for the results. The results are automatically appended to a SAS dataset called \_results\_.
* var: the quantitative variable of interest.
* group: optional list of categorical variables within which results are to be produced. The parameter should be left blank to produce estimates with no break down.
* where: optional where statement to define the subset of records of interest. The parameter should be left blank when all records are to be included.
* perc: percentile of interest. Specify a value between 1 and 99; e.g., specify 50 for a median.

The macro output includes the following:

* VarName: variable of interest
* PXX: Estimate of the percentile; e.g., P50 for median
* LOW\_PERC, UP\_PERC, CV\_PERC: 95% confidence interval and CV of the percentile
* ReleaseCategory\_PERC: release category of the percentile (A, E or F)
* n: unweighted count of the number of respondents that contributed to the estimate
* nrep: number of non-empty bootstrap replicates used to compute the variance

Note regarding zeros:

In situations where the variable of interest contains many zeros, it is preferable to produce two estimates as follows: an estimate of the proportion of non-zero values, along with the estimated percentile for the non-zero values.

## **2.4 Compute\_DIFFPCT: Compares Proportions or Percents**

Compute\_DIFFPCT tests for significant differences between proportions or percents, and constructs confidence intervals for differences.

Macro call: %Compute\_DIFFPCT (infile,outfile,var,varcat,diffgroup,group,where,keepall)

Description of the macro parameters:

* infile: input SAS dataset
* outfile: optional output SAS dataset for the results. The results are automatically appended to a SAS dataset called \_results\_.
* var: the categorical variable of interest. For the SUDAAN version of the macro, the variable should be numeric.
* varcat: category of interest for the variable of interest
* diffgroup: categorical variable. The percent of the variable of interest is compared between the categories of diffgroup.
* group: optional list of categorical variables within which results are to be produced. The parameter should be left blank to produce estimates with no break down.
* where: optional where statement to define the subset of records of interest. The parameter should be left blank when all records are to be included.
* keepall: optional yes/no parameter to indicate whether to keep extra records in the output. Default is no. (Note that the keepall parameter does not do anything for the SAS version of the macro because the SAS procedure does not produce any extra records.)

The macro output includes the following:

* varname: variable of interest
* diffgroup\_1, diffgroup\_2: the two categories that are being compared
* DIFFPCT: difference of the two estimated percents for the variable of interest
* LOWDIFF, UPDIFF: 95% confidence interval of the difference
* pvalue: p-value for the null hypothesis that there is no difference
* significant: \* indicates that the percents are significantly different

## **2.5 Compute\_DIFFMEAN: Compares means**

Compute\_DIFFMEAN tests for significant differences between means, and constructs confidence intervals for differences.

Macro call: %Compute\_DIFFMEAN (infile,outfile,var,diffgroup,group,where,keepall)

Description of the macro parameters:

* infile: input SAS dataset
* outfile: optional output SAS dataset for the results. The results are automatically appended to a SAS dataset called \_results\_.
* var: the quantitative variable of interest. The variable should be numeric.
* diffgroup: categorical variable. The mean of the variable of interest is compared between the categories of diffgroup.
* group: optional list of categorical variables within which results are to be produced. The parameter should be left blank to produce estimates with no break down.
* where: optional where statement to define the subset of records of interest. The parameter should be left blank when all records are to be included.
* keepall: optional yes/no parameter to indicate whether to keep extra records in the output. Default is no. (Note that the keepall parameter does not do anything for the SAS version of the macro because the SAS procedure does not produce any extra records.)

The macro output includes the following:

* varname: variable of interest
* diffgroup\_1, diffgroup\_2: the two categories that are being compared
* DIFFMEAN: difference of the two estimated means for the variable of interest
* LOWDIFF, UPDIFF: 95% confidence interval of the difference
* pvalue: p-value for the null hypothesis that there is no difference
* significant: \* indicates that the means are significantly different

# **Main Program**

An example of a main program is provided in the file called ‘MainProgram\_Example.sas’. The main program should include the following:

* Input the survey data and the bootstrap weights; merge them into one dataset
* Specify the variable name of the survey weight through macro variable ‘wt’. Specify the variable name of the first bootstrap replicate weight and the last bootstrap replicate weight through macro variables ‘rep\_first’ and ‘rep\_last’ respectively. Specify the number of bootstrap weights through macro variable ‘b’.
* For the SAS version of the macros: specify the fay parameter for the bootstrap weights through macro variable ‘fay’; the fay parameter should be 0 for regular bootstrap weights. For the SUDAAN version of the macros: specify the adjfay parameter for the bootstrap weights through macro variable ‘adjfay’; the adjfay parameter should be 1 for regular bootstrap weights.
* Input the general macros, SAS\_estimation\_macros.sas or SUDAAN\_estimation\_macros.sas, through a %include statement.
* Produce estimates by calling the appropriate macros with the appropriate parameters, and save the estimation results.

Release Rules for Quality

Until recently, most surveys at Statistics Canada used CV-based rules to measure and report the quality of survey estimates in terms of their sampling error. Concern with these rules has led SSMD to recommend a new set of release rules in 2021; see the accompanying documents for information. The specific release rules that apply to the survey at hand are usually provided by Methodology and specified in the User Guide of the survey. The estimation macros apply the SSMD 2021 release rules by default.

For Users using the SSMD 2021 Release Rules for Quality

The release rules recommended by SSMD in 2021 are automatically applied by the macros *Compute\_FREQ, Compute\_MEAN* and *Compute\_PERCENTILE*. In order to apply the release rules, the user must specify the sample size thresholds applicable to the survey; these are usually provided by the User Guide of the survey. The sample size thresholds are specified in the main program through a macro called *THRESHOLD*. The macro *THRESHOLD* should be defined before calling *Compute\_FREQ, Compute\_MEAN* or *Compute\_PERCENTILE*.The thresholds for quality suppressions are specified by setting a variable called ‘limit\_sup’, and the thresholds for quality warnings are specified by setting a variable called ‘limit\_warn’. The program code in the macro *THRESHOLD* should be in terms of all possible domain variables that are related to the release rules. For example, if the release rules are based on geography, then the macro should include logic for all the geographic variables available on the data file. If the data file contains province and CMA, then the macro should include logic for both variables, so that the rules are applied correctly when estimates are computed by province, and also when estimates are computed by CMA.

For users using CV-based Release Rules

The automatic application of the SSMD 2021 release rules can be skipped by including the following statement in the main program: %let ReleaseRules=NO;. This statement should appear after the %include statement that inputs the general estimation macros. (It should appear afterwards because the program code for the estimation macros resets the macro variable ‘ReleaseRules’ back to its default value of YES.) The programs, SAS\_estimation\_macros.sas and SUDAAN\_estimation\_macros.sas, contain macros called CV\_pct, CV\_mean or CV, that can be used to apply the CV-based rules. The macro CV\_pct computes the CV for proportions and estimated counts using the standard error before applying the rules; this macro should therefore be called before any rounding is performed.