

%CHCKFRQS – A Macro Application for Generating Frequencies for QC and Simple Reports

John Iwaniszek, MSc, Stat-Tech Services, LLC, Chapel Hill, NC
Natalie Walker, Stat-Tech Services, LLC, Chapel Hill, NC
Laura Lovette, Stat-Tech Services, LLC, Chapel Hill, NC

ABSTRACT

There are occasions when data must be summarized quickly and efficiently, with minimal formatting, and displayed in a format that is easily interpreted. The macro described below was written with that goal in mind. It is used most often in QC where table results must be reproduced and QC results compared to target tables. The most basic frequencies can be generated using SAS® PROC FREQ, but the default output is usually organized in a manner that is loaded with excess information. A SAS® macro, %CHCKFRQS, was written that produces frequencies of categorical variables alone or stratified by one or more BY variables. By default it bases the denominator on the input data set, but it also has an option allowing the entry of a special denominator file (useful of multi-record dataset like adverse event or other multi-response data). There are several display options allowing formatting percents, suppressing display of percents or totals, formatting the analysis variable, and suppressing all output. Output options include the ability to send the data to an output data set for further processing.

INTRODUCTION

There are tasks in clinical programming when data must be summarized quickly and presented with minimal formatting in a format that is easily interpreted. One such task involves producing parallel summaries to be used in verifying the results of clinical tables. In this situation, the parallel summaries (QC output) must be similar enough in format to the tables being verified that the correspondence of the QC output to the target table is obvious and provides a similar flow to that of the summaries in the target table. The degree that the QC output corresponds to the target table must be balanced with the amount of time it takes to format the output.

%CHCKFRQS was designed as a tool to create frequencies, and cross tabulations and display the results in simple uncluttered form. It is used most often as a QC tool, but is also useful whenever simple frequencies are required for investigating data and the relationships between variables.

Some features of %CHCKFRQS include optional display of group totals, percents, forced categories based on the summary variable's format, zero-fill, listing records with missing values on the summary variable, application of a denominator independent of the input data set, and output of the frequency results (and suppression of the default listing output) for use in other parts of the summary program.

The following examples demonstrate many of the major features of %CHCKFFRQS. These examples include:

- Plain one-way frequency with percent and total suppressed
- Plain one-way frequency showing percent with total suppressed
- Plain one-way frequency showing percent and total
- Plain one-way frequency showing percent and total printed from the output data set with PROC PRINT labels turned off.
- Two way frequency with percents and totals
- Two way frequency with percents and totals demonstrating CATFORCE and ZEROFILL
- Two-way frequency multi-record data with denominator file specified.

EXAMPLES

Example 1 - Plain one-way frequency with percent and total suppressed

The following example is a simple one-way frequency with no percents or total counts displayed.

The following call to %CHCKFRQS produces simple frequencies of RACE. Note that no percents are displayed because the default setting for the macro parameter PERCFORM has no value (missing). The parameter SHOWTOT is set to N to suppress display of the total.

```
title2 "Example 1: Plain one-way frequency with percent and total suppressed";
%chckfrqs( inset=sasdata.adsl, invar=race, fmt=$race., stdtitle=Example,
           titlxt=Race Categories, showtot=n);
```

Example 1: Plain one-way frequency with percent and total suppressed

Example 1 => Frequency of RACE (Race) Race Categories

Race	Frequency Count
Black	7
White	9
Hispanic	2
Indian	2

Example 2 - Plain one-way frequency showing percent with total suppressed

The following example is the same frequency of RACE as in example 1, but the macro parameter PERCFORM has been set to 5.1 to display percents with one digit to the right of the decimal place.

```

title2 "Example 2: Plain one-way frequency showing percent with total suppressed";
%chckfrqs( inset=sasdata.adsl, invar=race, percform=5.1, fmt=$race.,
          stdtitle=Example, titltext=Race Categories Example 2, showtot=n);

```

Example 2: Plain one-way frequency showing percent with total suppressed

Example 2 => Frequency of RACE (Race) Race Categories

Race	Frequency	
	Count	Percent
Black	7	35.0
White	9	45.0
Hispanic	2	10.0
Indian	2	10.0

Example 3 - Plain one-way frequency showing percent and total

This next example is the same frequency as in the previous two examples, but the total and percents are now fully displayed.

```

title2 "Example 3: Plain one-way frequency showing percent and total";
%chckfrqs( inset=sasdata.adsl, invar=race, percform=5.1, fmt=$race.,
          stdtitle=Example, titltext=Race Categories Example 3, showtot=y);

```

Example 3: Plain one-way frequency showing percent and total

Example 3 => Frequency of RACE (Race) Race Categories

Race	Frequency		
	Count	Total	Percent
Black	7	20	35.0
White	9	20	45.0
Hispanic	2	20	10.0
Indian	2	20	10.0

Example 4 - Plain one-way frequency showing percent and total printed from the output data set with PROC PRINT labels turned off.

As stated earlier, %CHCKFRQS allows suppression of the default output from the macro and can output the summary to a SAS data set for further processing or alternate display modalities. This next example is based on the one-way frequency summary used in the previous examples and shows the output file produced by %CHCKFRQS as printed using PROC PRINT, as well as the contents of the data set as revealed by a simple PROC CONTENTS run.

```

title2 "Example 4: Plain one-way frequency showing percent and total printed from
the";
title3 "output data set with PROC PRINT labels turned off";
%chckfrqs( inset=sasdata.adsl, invar=race, percform=5.1, fmt=$race.,
          stdtitle=Example, titltext=Race Categories, nolist=y,
          outfile=dcount, showtot=y);

```

```
proc print data=dcount;
run;
```

Example 4: Plain one-way frequency showing percent and total printed from the output data set with PROC PRINT labels turned off

Obs	RACE	COUNT	Percent	Total
1	B	7	35	20
2	C	9	45	20
3	H	2	10	20
4	I	2	10	20

Contents of data set

Num	Variable	Type	Len	Pos	Label
2	COUNT	Num	8	0	Frequency Count
3	Percent	Num	8	8	
1	RACE	Char	1	24	Race
4	Total	Num	8	16	

Example 5 – Two way frequency with percents and totals

The following example is a two-way frequency with percent and total display option activated. Note that the by-group total only appears on the final summary line within a by-group.

```
title2 "Example 5: Plain two-way frequency showing percents and totals";
%chckfrqs( inset=adsl, invar=race, byvar=trtp, percform=5.1, fmt=$race.,
stdtitle=Example, ttitle=Race Categories, showtot=y);
```

Example 5: Plain two-way frequency showing percents and totals

Example 5 => Frequency of RACE (Race) Race Categories

Planned Treatment Group	Race	Frequency Count	Total	Percent
Treatment A	Black	4	10	40.0
	White	3		30.0
	Hispanic	1		10.0
	Indian	2		20.0
Treatment B	Black	3	10	30.0
	White	6		60.0
	Hispanic	1		10.0

Example 6 – Two way frequency with percents and totals demonstrating CATFORCE and ZEROFILL

The following example demonstrates how %CHCKFRQS can be directed to display categories not present in the summarized data. It uses the display format for the INVAR

variable to construct a matrix of possible categories, and fills any categories not present in the summary data set with zero-counts. This example uses the same data as was summarized in Example 5, with the additional feature of forced categories (CATFORCE=Y) and zero-fill (ZEROFILL=Y).

```
title2 "Example 6: Plain two-way frequency showing percents and totals";
title3 "demonstrating CATFORCE and ZEROFILL";
%chckfrqs( inset=adsl, invar=race, byvar=trtp, percform=5.1, fmt=$race.,
           stdtitle=Example, titlext=Race Categories, showtot=y, catforce=Y,
           zerofill=Y);
```

Example 6: Plain two-way frequency showing percents and totals
demonstrating CATFORCE and ZEROFILL

Example 6 => Frequency of RACE (Race) Race Categories

Planned Treatment Group	Race	Frequency Count	Total	Percent
Treatment A	Other	0		0.0
	Black	4		40.0
	White	3		30.0
	Hispanic	1		10.0
	Indian	2	10	20.0
Treatment B	Other	0		0.0
	Black	3		30.0
	White	6		60.0
	Hispanic	1		10.0
	Indian	0	10	0.0

Example 7 – Two-way frequency multi-record data with denominator file specified

The following example introduces the %CHCKFRQS feature that allows use of a denominator derived independently of the data set from which the frequencies are derived. The denominator comes from a denominator file that has one record per unit of analysis (subjects or patients, for example).

The data for this example are records similar to a medical history section of a CRF where each subject may have indicated that they have none, one, or more of a set of medical findings. Those subjects who have no medical history findings have no records in the input data set.

The denominator is supplied by a file that works as a master list of subjects. In this data set every subject is accounted for and has a treatment group assignment. The denominator file is introduced using the DENOMFILE parameter and linked to the frequencies by the variables populating the BYVAR parameter.

```
title2 "Example 7: Frequency of TRT by medical history body system";
title3 "with denominator file specified";

%chckfrqs( inset=mh, invar=bodsys, byvar=trtp, percform=5.1, fmt=mhbodsys.,
           stdtitle=Example, titlext=Trt by Medical History Body System,
           showtot=Y, denomfile=adsl, catforce=y, zerofill=y);
```

Example 7: Frequency of TRT by medical history body system
with denominator file specified

Example 7 => Frequency of BODSYS () Trt by Medical History Body System

Planned Treatment Group	BODSYS	Frequency Count	Total	Percent
Treatment A	Ear, Nose and Throat	2		20.0
	Eye	0		0.0
	Respiratory	3		30.0
	Cardiovascular	2		20.0
	Gastrointestinal	3		30.0
	Hepatobiliary and Pancreas	4	10	40.0
Treatment B	Ear, Nose and Throat	2		20.0
	Eye	2		20.0
	Respiratory	1		10.0
	Cardiovascular	1		10.0
	Gastrointestinal	2		20.0
	Hepatobiliary and Pancreas	3	10	30.0

CONCLUSION

%CHCKFRQS is a powerful tool for producing a variety of useful summaries. It is intended to be used in quickly and efficiently generating output that can be used for examining data, generating N's and totals for data review, and for Quality Control purposes where the output is used to independently verify camera ready-summary tables. The macro incorporates many features that give it a great deal of flexibility in how the results are displayed. It also provides a means to introduce a denominator data set so that multi-record-per-subject data may be summarized with the percents based on the correct unit of analysis.

It is clear that what %CHCKFRQS accomplishes can be done using other more standard features of the SAS system, and that certain aspects of the macro could be made more efficient (particularly the way the denominator data set is handled – conceivably, the denominator could be derived from the multi-record input data set in those situations where it is known that all subjects are represented in the input data set). One drawback to the macro is that as written it is difficult to add a sub-setting feature, analogous to “WHERE” processing, however, this when necessary is easily handled in a data step. Also, there is no feature allowing a format or list of formats to be applied to the BYVARS so any by-variables must be formatted in previous data step.

But despite its shortcomings, %CHCKFRQS is a powerhouse of a summary tool that finds its way into a variety of applications from data review, Quality Control, and main table production.

%CHCKFRQS MACRO PARAMETERS

Parameter=Default	Purpose / Options
inset=	Name of input data set
invar=	Name of variable on input data set to be summarized
byvar=	Any by variables used to stratify INVAR
percform=	Display format for percents. If left null, no percents will be displayed
fmt=	Format for INVAR. If invar is formatted on data set, then formatted values will be displayed
stdtitle=QC Check	This is a label attached to the default title produced by %CHCKFRQS
titlxt=	Extended text for default title
prntmiss=n	Print list of records with missing values on INVAR
nolist=n	Suppress default output
outfile=	Name of output data set containing default output
idvar=	Name of variable to serve as print ID variable
shotitle=y	Indicates whether default title will be displayed
showtot=y	Controls display of within group totals. If frequency table is one-way then totals appear next to each category frequency. If one or more variables are indicated in the BYVAR parameter, then totals appear for the inner-most group.
denomfile=	Name of optional denominator file
order=ALPHA	Sort order of INVAR on final display
incmiss=y	Include observations with missing values in denominator
catforce=n	Include (force) all format categories in final display
zerofill=n	if CATFORCE=Y then ZEROFILL=Y will fill forced categories with zero counts.

CONTACT INFORMATION

John Iwaniszek
Director of Programming and Study Services
Stat-Tech Services, LLC
Chapel Hill, NC 27514
919 929 5015
JIwaniszek@StatTechServices.com
WWW.StatTechServices.com

Natalie Walker
Statistical Programming Manager
Stat-Tech Services, LLC
Chapel Hill, NC 27514
919 929 5015
NWalker@StatTechServices.com
WWW.StatTechServices.com

Laura Lovette
Statistical Programmer I
Stat-Tech Services, LLC
Chapel Hill, NC 27514
919 929 5015
LLovette@StatTechServices.com
WWW.StatTechServices.com

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

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