

# SAS® Macros for Quick Summary Reporting and Uppercasing of All Variables

Craig Balog, Cleveland Clinic, Cleveland, OH

## ABSTRACT

This paper presents two simple and basic macros that can be used by any SAS® programmer. The first macro, %FREQCHK, shows the frequency for all categorical variables in a given data set using the SAS FREQ procedure. The second macro, %UPPER, simply uppercases all variables in one or more SAS data sets. Both macros are simple to use and are a tremendous help and timesaver when programming.

## INTRODUCTION

When trying to understand your data, have you ever wanted to do a PROC FREQ for only the categorical variables in a data set without having to type all of your variable names? If you simply do a PROC FREQ of your entire data set, this may very well result in a large amount of unwanted output since you may have both unique identifiers and continuous data in your data set. This paper presents a macro that does a PROC FREQ for only the categorical variables.

Additionally, have you ever had one or more SAS data sets where you wanted the variables to be uppercased? Maybe you have looked at the output from the SAS CONTENTS procedure and found it hard to figure out which variables were in your data set because some variables were uppercase, others lowercase, and still others in a mixed case format. In the output from PROC CONTENTS uppercase letters precede any lowercase letters when alphabetizing; for example, "Z" comes before "a" alphabetically. This paper also presents a simple macro that solves this problem of alphabetizing, by uppercasing all the variables in one or more data sets, and therefore making the contents appear cleaner while preserving all other attributes of the original data set.

## THE %FREQCHK MACRO

The macro %FREQCHK lists the frequencies using PROC FREQ for all variables in a data set that are categorical. In this macro, a variable is defined as categorical when the number of categories for a variable are less than or equal to a cutoff specified in the macro call. The number of categories is determined by using the NODUPKEY option on the SORT procedure for every variable in the input data set, and then using the largest value of \_N\_ in the data set output from PROC SORT. This macro is ideal for determining the values for each categorical variable in your data set (especially character variables). Calling the %FREQCHK macro along with the MEANS procedure, is a quick and simple way to understand your data. This macro is simple to use and only has three parameters which are the following:

- 1) **ds** - The name of the input data set
- 2) **cutoff=10** - The specified cutoff (which is defaulted to 10) that defines whether or not a variable is categorical.
- 3) **titleyn=1** - If this option is set to 1 then the name of the data set is specified in the title. Otherwise, if 1 is not specified then no title is printed. This option is defaulted to 1.

### Example:

Suppose you have a data set *demog*. If you want to see the values of the categorical variables in *demog* and do not wish to see the title of the data set then call to the %FREQCHK macro in the following manner:

```
%freqchk (ds=demog,cutoff=5,titleyn=)
```

This will give the following output:

The FREQ Procedure

Female				
FEMALE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	343	68.60	343	68.60
Yes	157	31.40	500	100.00

Race

RACE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
White	438	87.60	438	87.60
Black	46	9.20	484	96.80
Asian	1	0.20	485	97.00
Other	15	3.00	500	100.00

Notice in the above example that the parameter **cutoff** was specified as 5. All variables that have 5 or less categories are presented. If this parameter is changed from 5 to 3 and the **titleyn** parameter is set to 1, the call is as follows:

```
%freqchk (ds=demog,cutoff=3,titleyn=1)
```

This will give the following output:

demog

The FREQ Procedure

Female				
FEMALE	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	343	68.60	343	68.60
Yes	157	31.40	500	100.00

Notice that the variable RACE is not included because **cutoff** was specified as 3. Also notice that the data set name *demog* appears in the title.

#### CODE FOR THE %FREQCHK MACRO

```
%macro freqchk (ds=,cutoff=10,titleyn=1);
  proc contents data=&ds out=freqchk1 noprint;
  run;

  data _null_;
    set freqchk1 (keep=name) end=last;
    call symput('var' || trim(left(_n_)),trim(left(name)));
    if last then call symput('count',trim(left(_n_)));
  run;

  %do i=1 %to &count;
    proc sort data=&ds out=freqchk2 (keep=&&var&i) nodupkey;
      by &&var&i;
    run;
    data _null_;
      set freqchk2 end=last;
      if last then do;
        if _n_ <= &cutoff then flag=1;
        else flag=0;
        call symput("flag&i",trim(left(flag)));
      end;
    run;

    proc datasets library=work;
      delete freqchk2;
    quit;
  %end;
  %let okflag=0;
  %do i=1 %to &count;
    %if &&flag&i=1 %then %let okflag=1;
  %end;

  %if &titleyn=1 %then %str(title "&ds");
  %else %str(title);
```

```

    %if &okflag=1 %then %do;
        proc freq data=&ds;
            table %do i=1 %to &count;
                %if &&flag&i=1 %then %do;
                    &&var&i
                %end;
            %end;;
        %end;

    proc datasets library=work;
        delete freqchk1;
    quit;

%mend freqchk;

```

## THE %UPPER MACRO

The second macro, %UPPER, simply uppercases every variable in one or more data sets. This macro only has one parameter which is the name of the data set (or data sets) which are to be uppercased. The macro works on both temporary and permanent data sets. It is also worth mentioning that all attributes including labels, formats, informats, variable type, variable length, and position remain the same after being uppercased.

### Example:

Suppose you have a data set *demog* with the following variables and attributes as described by PROC CONTENTS:

2	AGE	Num	8	0		Age
4	PT_WGT	Num	8	16	8.2	Weight (kg)
3	PT_hgt	Num	8	8	8.2	Height (cm)
6	RACE	Num	8	32	RACE.	Race
5	female	Num	8	24	YN.	Female
1	subject	Char	25	40		Patient

In addition, suppose you also have a data set *save.endpt* with the following variables and attributes:

3	MI	Num	8	8	YN.	MI
1	SUBJECT	Char	9	24		Patient
2	death	Num	8	0	YN.	Death
4	stroke	Num	8	16	YN.	Stroke

To uppercase all of the variables for these two data sets, you can either do this with

1. One macro call:

```
%upper (demog save.endpt)
```

2. or two macro calls:

```

%upper (demog)
%upper (save.endpt)

```

After invoking the %UPPER macro for both data sets, the output for *demog* is the following:

2	AGE	Num	8	0		Age
5	FEMALE	Num	8	24	YN.	Female
3	PT_HGT	Num	8	8	8.2	Height (cm)
4	PT_WGT	Num	8	16	8.2	Weight (kg)
6	RACE	Num	8	32	RACE.	Race
1	SUBJECT	Char	25	40		Patient

and the output for *save.endpt* is as follows:

2	DEATH	Num	8	0	YN.	Death
3	MI	Num	8	8	YN.	MI
4	STROKE	Num	8	16	YN.	Stroke
1	SUBJECT	Char	9	24		Patient

Notice that the only thing that has changed is that the variable names have been uppercased. All other attributes are the same.

#### CODE FOR THE %UPPER MACRO

```
%macro upper (ds);

    %local count i;
    %let count=0;
    %do %until(&&dsnm&count=);
        %let count=%eval(&count+1);
        %let dsnm&count=%scan(&ds,&count,%str( ));
    %end;
    %let count=%eval(&count-1);

    %do i=1 %to &count;

        proc contents data=&&dsnm&i out=upper1 (keep=name varnum) noprint;
        run;
        proc sort data=upper1;
            by varnum;
        run;
        data _null_;
            set upper1 end=last;
            call symput("upvar&i._"||trim(left(_n_)), trim(left(upcase(name))));
            if last then call symput("upcount&i",trim(left(_n_)));
        run;
        data &&dsnm&i;
            retain %do j=1 %to &&upcount&i;
                &&upvar&i._&j
            %end;;
            set &&dsnm&i;
        run;
        proc datasets library=work nolist nodetails;
            delete upper1;
        quit;

    %end;
%mend upper;
```

#### CONCLUSION

The %FREQCHK and %UPPER macros are both simple and very useful tools for the programmer. The %FREQCHK macro is great for determining the values of your categorical data so that you can further process it, and when you are finished processing your data, you can use the %UPPER macro to uppercase the variables in your output data set or data sets. These two macros would both make great additions to your macro library.

#### ACKNOWLEDGMENTS

I would like to thank my colleagues April Bebee-Sanders, Danielle Brennan, Amy Hsu, and Kathy Wolski for their review and critique of this paper. Their input has been invaluable and is very much appreciated.

#### CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Craig Balog  
Cleveland Clinic  
Cardiovascular Coordinating Center / JJ65  
Cleveland, OH 44195  
Work Phone: (216) 444-9681  
Fax: (216) 444-9520  
E-mail: balogc@ccf.org

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.