

SAS® Functions You May Have Been MISSING

Mira Shapiro, Bethesda, MD

ABSTRACT

Those of us who have been using SAS for more than a few years often rely on our tried- and-true techniques for standard operations like assessing missing values. Even though the old techniques still work, we often miss some of the “new” functionality added to SAS that would make our lives much easier. In effort to ascertain how many people skipped questions on a survey and, what percentage of people answered each question, I did a search of past conference papers and came across 2 functions that were introduced in SAS 9.2-- CMISS and NMISS. By using a combination of these functions and Proc Transpose, a full missing assessment can be done in a concise program. This paper will provide examples and explore the features of the “newer” functions NMISS and CMISS and compare them with the longer standing MISSING function.

INTRODUCTION

The first step in exploring a new data set includes a careful assessment of each variable’s fill rate and missing values. In the past, MISSING was the only function specifically available for evaluating missing values. In SAS 9.2, CMISS and NMISS functions were introduced simplifying the programming required to assess both numeric and character variable missing values.

As an experienced SAS programmer, I was unaware of CMISS and NMISS but happily discovered these functions in SAS 9.4. This discussion is focused on the MISSING, CMISS and NMISS functions and how they can be used to quickly and easily create a missing value report.

DATA and METHODS

The dataset used throughout this paper is SASHelp.heart and can be found in the SASHelp library. Some minor changes have been made in the output in the examples to demonstrate the capabilities of the functions to count missing values.

SAS® OnDemand for Academics, accessible without cost on the web, was used to run the SAS code and to create all of the results. The interface to SAS is SAS Studio 3.4. These resources are available to all users and can be used to re-run and explore all of the concepts introduced in the paper.

MISSING FUNCTIONS DESCRIBED (MISSING, NMISS, CMISS)

There are many ways to work with missing values in SAS, among them is DATA Step programming, procedure options and Proc SQL statements. This paper focuses specifically on how to use the MISSING, NMISS and CMISS functions effectively. SAS documentation and numerous user group papers give detailed descriptions of these functions. The following table summarizes the appropriate use and output for each of the functions and can be used for a quick reference when embarking on a missing value assessment.

Table 1 Missing Value Functions

| Function | Numeric | Character | Results |
|----------|-------------|-------------|--|
| MISSING | YES | YES | Numeric:missing(.) value returns 1 Character: blank value returns 1 single parameter only |
| NMISS | YES | NO | Numeric Variable *one variable: missing (.) value returns 1, valid returns 0 *multiple: adds 1 for each missing; returns sum Character Variable * all values return 1, NOTES indicate invalid numeric data and data converted to numeric single or multiple parameters numeric only |
| CMISS | YES / MIXED | YES / MIXED | Numeric Variable *one variable: missing (.) value returns 1, valid returns 0 *multiple: adds 1 for each missing; returns sum Character Variable *one variable: blank value returns 1, valid returns 0 *multiple: adds 1 for each missing; returns sum Mixed Character and Numeric *returns the sum of the blank character variables and the missing (.) numeric variables single or multiple parameters numeric, character or mixed |

Table 1: Summary of missing value function usage and results.

EXAMPLES

Example 1: MISSING Function

The MISSING function takes just one parameter and returns a 1 for a blank field for a character variable and a zero otherwise. For numeric variables it returns a 1 for a missing (.) and zero otherwise.

* Program: MISSING Function;

```
Data Missing_function;
Set sashelp.heart;
char_test = missing(chol_status);
num_test = missing(ageatdeath);
run;

* Character variable results;
Proc print data=missing_function;
Var chol_status char_test;
run;

*Numeric variable results;
Proc print data=missing_function;
Var ageatdeath num_test;
run;
```

Results Example 1 : Character Variable

| Obs | Chol_Status | char_test |
|-----|-------------|-----------|
| 1 | | 1 |
| 2 | Desirable | 0 |
| 3 | High | 0 |
| 4 | High | 0 |

Results Example 1: Character variable with MISSING function.

Result Example 1: Numeric Variable

| Obs | AgeAtDeath | num_test |
|-----|------------|----------|
| 1 | 55 | 0 |
| 2 | 57 | 0 |
| 3 | . | 1 |
| 4 | . | 1 |

Results Example 1: Numeric variable with MISSING function.

Example 2: NMISS Function: Single Numeric Variables or Multiple Numeric Values

The NMISS function operates only on numeric variables and takes either a single parameter or multiple parameters separated by a comma. If one numeric variable is used a 1 is returned for a missing value and a 0 for a non-missing value. When multiple numeric variables are used, NMISS returns the sum that is composed of a 1 for each of variables that are missing.

***Program: NMISS Function**

```

Data nmiss_function;
Set sashelp.heart;
*Single variable;
nmiss_test = nmiss(ageatdeath);

*Multiple variables;
multi_nmiss_test = nmiss(ageatdeath,ageatstart,height,weight);
run;

* Single numeric variable results;
Proc print data=nmiss_function;
var ageatdeath nmiss_test;
run;

*Multiple numeric variable results
proc print data=nmiss_function;
var multi_nmiss_test ageatdeath ageatstart height weight;
run;

```

Results Example 2 : Single Numeric Variable

| Obs | AgeAtDeath | nmiss_test |
|-----|------------|------------|
| 1 | 55 | 0 |
| 2 | 57 | 0 |
| 3 | . | 1 |

Results Example 2: NMISS function with single numeric parameter

Results Example 2 : Multiple Numeric Variables

| Obs | multi_nmiss_test | AgeAtDeath | AgeAtStart | Height | Weight |
|-----|------------------|------------|------------|--------|--------|
| 1 | 0 | 55 | 29 | 62.50 | 140 |
| 2 | 0 | 57 | 41 | 59.75 | 194 |
| 3 | 1 | . | 57 | 62.25 | 132 |
| 4 | 2 | . | . | 65.75 | 158 |

Results Example 2: NMISS function with multiple numeric parameters

Example 3: CMISS Function : Single Numeric or Character Variables or Multiple Numeric, Character or Combination

The CMISS function operates on both numeric and character variables either with a single or multiple parameters. CMISS will also count the missing values using parameters that include both numeric and character fields.

*Program: CMISS Function

```

Data multi_cmiss_function;
Set sashelp.heart;
multi_num_cmiss_test = cmiss(ageatdeath,ageatstart,height,weight);
multi_char_cmiss_test = cmiss(status, deathcause, sex);
multi_mixed_cmiss_test = cmiss(ageatdeath,ageatstart,status, deathcause);
run;
proc print data=multi_cmiss_function (obs=5);
var multi_num_cmiss_test ageatdeath ageatstart height weight;
run;
proc print data=multi_cmiss_function (obs=5);
var multi_char_cmiss_test status deathcause sex;
run;
proc print data=multi_cmiss_function (obs=5);
var multi_mixed_cmiss_test ageatdeath ageatstart status deathcause;
run;

```

Results Example 3 : Multiple Numeric Variables

| Obs | multi_num_cmiss_test | AgeAtDeath | AgeAtStart | Height | Weight |
|-----|----------------------|------------|------------|--------|--------|
| 1 | 0 | 55 | 29 | 62.50 | 140 |
| 2 | 0 | 57 | 41 | 59.75 | 194 |
| 3 | 1 | . | 57 | 62.25 | 132 |
| 4 | 1 | . | 39 | 65.75 | 158 |
| 5 | 2 | . | . | 66.00 | 156 |

Results Example 3: CMISS function with multiple numeric variables

Results Example 3 : Multiple Character Variables

| Obs | multi_char_cmiss_test | Status | DeathCause | Sex |
|-----|-----------------------|--------|------------|--------|
| 1 | 0 | Dead | Other | Female |
| 2 | 0 | Dead | Cancer | Female |
| 3 | 1 | Alive | | Female |
| 4 | 2 | Alive | | |
| 5 | 1 | Alive | | Male |

Results Example 3: CMISS function with multiple character variables

Results Example 3 : Multiple Mixed Numeric and Character Variables

| Obs | multi_mixed_cmiss_test | AgeAtDeath | AgeAtStart | Status | DeathCause |
|-----|------------------------|------------|------------|--------|------------|
| 1 | 0 | 55 | 29 | Dead | Other |
| 2 | 0 | 57 | 41 | Dead | Cancer |
| 3 | 2 | . | 57 | Alive | |
| 4 | 2 | . | 39 | Alive | |
| 5 | 2 | . | 42 | Alive | |

Results Example 3: CMISS function with mixed multiple numeric and character variables

Errors, Warnings and Notes Oh My...Do not use NMISS with Character variables

As described about the NMISS function operates only on numeric variables. If a character variable parameter is supplied to the NMISS function, several warnings and notes will appear in the SAS Log. The first indications of a problem are the following notes:

NOTE: Character values have been converted to numeric values at the places given by: (Line):(Column).

NOTE: Invalid numeric data, Chol_Status='Desirable' , at line 59 column 21.

In addition a warning about excessive error messages appears. However, results are created and 1 is returned for each entry. The most important thing to remember here is to always check the log for errors, warnings and notes. In this case there are results but they do not accurately reflect the true nature of the data.

CMISS for Survey Assessment

One of the many uses for CMISS is for assessing survey data. One of the important measures that is used to assess whether the length of a survey is appropriate is the fraction of unanswered questions. In a DATA Step, the following code will count the number of unanswered questions for a participant:

```
Unanswered_questions=CMISS(all responses both character and numeric separated by commas);
```

This result divided by the total number of questions will provide the values needed to properly assess survey length.

In order to access the number of participants who have answered specific questions can be accomplished by using Proc Transpose and then using the CMISS function on the newly created variables.

CONCLUSIONS

Seasoned SAS programmers often are not always aware of new functions that are introduced in SAS. The CMISS and NMISS functions introduced in SAS 9.2 provide an elegant path to assessing missing values. CMISS in particular is a very useful tool in assessing survey question responses and survey length.

SOURCES OF ADDITIONAL INFORMATION

There is a wealth of information available on SAS missing value functions and all aspects of SAS. Some useful resources are listed below.

Table 2: Resources

| URL | Description |
|---|--|
| http://www.lexjansen.com/ | A website that provides access and a search engine for SAS conference papers from SAS Global Forum, Regional conferences and specialized SAS conferences such as PharmaSUG |
| http://www.sascommunity.org/wiki/Main_Page | A SAS community resource that serves as a portal to many sources of SAS information on the web. |
| https://support.sas.com/ | The SAS Customer Support website that contains a wealth of information including: troubleshooting, documentation and training. |

Table 2: Sources of additional information

APPENDIX

Appendix 1: Proc Contents: SASHELP.heart

| Alphabetic List of Variables and Attributes | | | | |
|---|----------------|------|-----|------------------------------|
| # | Variable | Type | Len | Label |
| 12 | AgeAtDeath | Num | 8 | Age at Death |
| 5 | AgeAtStart | Num | 8 | Age at Start |
| 3 | AgeCHDdiag | Num | 8 | Age CHD Diagnosed |
| 15 | BP_Status | Char | 7 | Blood Pressure Status |
| 14 | Chol_Status | Char | 10 | Cholesterol Status |
| 13 | Cholesterol | Num | 8 | |
| 2 | DeathCause | Char | 26 | Cause of Death |
| 8 | Diastolic | Num | 8 | |
| 6 | Height | Num | 8 | |
| 10 | MRW | Num | 8 | Metropolitan Relative Weight |
| 4 | Sex | Char | 6 | |
| 11 | Smoking | Num | 8 | |
| 17 | Smoking_Status | Char | 17 | Smoking Status |
| 1 | Status | Char | 5 | |
| 9 | Systolic | Num | 8 | |
| 7 | Weight | Num | 8 | |
| 16 | Weight_Status | Char | 11 | Weight Status |

CONTACT INFORMATION

Your comments and questions are valued and encouraged.

Contact the author at:

Name: Mira Shapiro

E-mail: mira.shapiro at gmail.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.