# Working Draft: A Table 1 Macro that Actually Produces Publication-Ready Results: %Table1nDone

Martha Wetzel, MSPH

# Abstract

Academic papers in the medical field typically include a table summarizing the demographic characteristics of the population and/or treatment groups—the ubiquitous Table 1. These tables often require an overall summary, as well as a between-group comparison. Furthermore, these tables summarize mixed data types, including both continuous and categorical data, requiring different statistical tests and SAS procedures. Without automation, analysts can spend hours per project calculating and arranging the results into the correct format, often having to redo the entire table when an investigator realizes there is an error with the included population. The %Table1nDone macro was created to reduce analyst time spent on Table 1s.

This paper presents a new Table 1 macro that calculates summary statics overall and by group, performs statistical testing as required, and produces an RTF file containing the final summary. The %Table1nDone macro expands on existing Table 1 macros by 1) streamlining variable input via an Excel file, 2) creating a table with both overall and by-group summary data, 3) producing an RTF table in the format expected by many journals, and 4) saving permanent data sets of key information for analyst review.

Logic is built into the macro to select the appropriate statistical test based on user-supplied factors such as variable type (e.g., categorical, continuous) and data factors (e.g., number of comparison groups, cell-size counts). Publication-ready output contains results formatted as N (%), mean (standard deviation), or median (quartile 1, quartile 3), depending on the type of data. The macro furthermore produces a report for the analyst to review for unexpected values in the data. This paper provides an overview of the macro’s capabilities, a description of the use and required parameters, an explanation of the statistical tests included, examples of output, and the macro code.

# Overview

Academic papers so often begin with a table containing descriptive statistics about the population studied that such tables are often referred to as simple “Table 1s.” Depending on field/journal, bivariate analyses comparing the population characteristics between key groups can also be included in the Table 1. The purpose of this macro is to streamline and standardize the production of these tables. To that end, this macro outputs statistics from univariate and/or bivariate analysis in a publication-ready RTF document. Secondarily, this macro is intended to provide the analyst with additional information, typically excluded from a Table 1, necessary for thorough understanding of the data.

Several existing macros produce descriptive tables. This macro expands on the existing macros by 1) streamlining variable input, 2) creating a table with both overall and by-group summary data, 3) producing an RTF document set up in a manner expected by many journals, and 4) saving permanent data sets of key information for analyst review.

Three pieces of output are created by this macro. The macro’s main product is the publication-ready RTF table, an example of which is shown below as Table 1. This table can be customized to display an overall statistics column and/or columns showing the statistics by group. Additional options allow the inclusion of columns showing the response counts, the metrics used, and, if the by-group option is turned on, p-values for group comparisons. The second piece of output is a temporary report showing descriptive statistics for the continuous variables, including means, medians, and minimums/maximums. This table is created as a check for the analyst to make sure that the data do not contain, for example, any extreme outliers that had been previously unnoticed. Finally, the macro saves SAS data sets containing detailed additional statistics. In particular, for the continuous variables, the saved data set includes the number missing observations, minimums/maximums, mode, and the results from the Shapiro-Wilk normality test and other performed statistical tests.

All examples in this paper draw from the sashelp.heart data set.

**Table 1: Sample Output**

| **Variable** | **Level** | **Overall N=5209** | **Female N=2873** | **Male N=2336** | **P-Value** | **Metrics** |
| --- | --- | --- | --- | --- | --- | --- |
| Age at Study Start |  | 43 (37, 51) | 43 (37, 51) | 44 (37, 51) | 0.922 | Median (Q1, Q3) |
|  | | | | | | |
| Vital Status | Alive | 3218 (61.78%) | 1977 (68.81%) | 1241 (53.13%) | **<.001** | N (%) |
| Dead | 1991 (38.22%) | 896 (31.19%) | 1095 (46.88%) | N (%) |
|  | | | | | | |

# Use

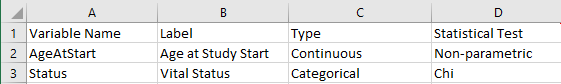
The steps for use of this macro are outlined below.

## Data Preparation

The input data should be at the individual level (e.g., if summarizing patient demographics, each line should represent one patient). Continuous variables must be numeric type and categorical variables must be character type.

## Variable List

The key to this macro setting up an Excel driver. The user adds the list of requested variables to the Excel driver, along with the desired label in the output and the requested tests. An example of a correctly completed driver is shown below in Figure 1.



**Figure 1: Excel Driver**

It is recommended that the driver template be used for setting up the driver to ensure there are no errors in the Type and Statistical Test columns. The driver template can be accessed at [insert link]. If the driver template is unavailable, a driver can be created on a spreadsheet tab named “Variables.” Valid values for each column are described below.

* Variable Name: Contains the names of all variables to be analyzed. All variables must be present in the input data set.
* Label: Enter the label to be used in the output for each variable. Note that this label will overwrite existing variable labels.
* Type: Valid values are “Continuous” and “Categorical.”
* Statistical Test: Valid values for continuous variables are “Parametric”, “Non-parametric,” and “Auto-select.” Valid values for categorical are “Chi”, “Fisher’s”, and “Auto-select.” A description of the statistical tests associated with these choices is provided in the section Statistical Details.

## Macro Call

The keyword parameters are defined below.

### Required Parameters

* DATASET: Name of the input SAS data set
* DRIVER: File path and name for Excel driver file
* OUTPATH: File pathway for RTF output
* FNAME: File name for the RTF output
* CLASSVAR: Name of the grouping variable. This is required if bivariate analyses are requested

### Optional Parameters

* OVERALL: Y/N include the overall descriptive statistics. Default = Y
* BYCLASS: Y/N perform bivariate analysis using the class variable specified in the CLASSVAR option
* OUTLIB: Name of library to save out SAS data sets to. Default = Work
* ROUNDTO: Decimal place to round results to, in the format used by the round function. Default = 0.01
* DISPLAY\_PVAL: Y/N display p-values in RTF report. Only valid with BYCLASS = Y. Default = Y
* DISPLAY\_METRIC: Y/N display metric names in RTF report. Default = Y
* DISPLAY\_N: Y/N display variable response counts in RTF report. Default = Y

### Example Call

The following code can be used to run the macro on the sashelp.heart data set.

%***TABLE1NDONE***(

DATASET = sashelp.heart, /\* Input data set \*/

DRIVER = C:\Heart Analysis\Descriptives Driver\_Heart.xlsx,

OUTPATH= C:\Heart Analysis, /\* File pathway for RTF output \*/

FNAME= Table **1** Heart, /\* File name for RTF output \*/

OVERALL = Y,

BYCLASS = Y,

CLASSVAR = Sex,

DISPLAY\_PVAL = Y, /\* Display p-values in RTF report \*/

DISPLAY\_METRIC = Y, /\* Display metric names in RTF report \*/

DISPLAY\_N = N /\* Display variable N's in RTF report \*/

);

## Log and Output Review

[Explain common log warnings and review saved output data sets]

# Statistical Testing

[Provide the macro logic for selecting the appropriate statistical test]