Table1Macro

# DATA CHECK FEATURES

A whole array of parameter and data checking steps are carried out to make sure the successful execution of macro.

1. Check all parameter assignments, if not valid, re-assign default value to corresponding parameter or stop macro and ask for new parameter assignment
2. Source data set must exist and not be empty
3. All variables listed in variable list parameters, including calist, coplist and cononplist, must be in the source data set
4. Duplicated variable names in any variable list, including calist, coplist and cononplist, are removed
5. Variables listed in continuous variable lists, coplist and cononplist, must be numeric
6. If specified, group variable must be in the data set and is removed from summary variable list, including calist, coplist and cononplist
7. If calist parameter is not empty, check eratio parameter and make sure its value ranges between 0 and 1, this parameter provides the threshold of whether Chi-square or Fisher’s Exact test is used to test association between categorical variables
8. If file path is included in the savefilename parameter and operating system is Microsoft Windows, check if the folder specified exists, if not, result file is saved in SAS current folder
9. Duplicated statistic requested in copmain, copsupplement, cononpmain and cononpsupplement parameters are removed
10. If any variable are all missing, no statistic will be calculated and missing information is reported even though caincludemissing, copincludemissing or cononpincludemissing is set as “no”
11. Check if statistic requested in copmain, copsupplement, cononpmain and cononpsupplement is valid. This marco uses proc means to calculate statistic for continuous variables. Only statistics supported by proc means is allowed
12. If more than 1 statistic is specified in copmain or cononpmain, the first one is picked to be reported in final table
13. If the same statistic is specified in copsupplement or cononpsupplement as copmain or cononpmain, it’ll not calculated
14. If more than 1 unique statistic is in the copsupplement or cononpsupplement and the first one is the same as the one in copmain or cononpmain, the second one in copsupplement or cononpsupplement is calculated
15. Duplicated variables in outputorder parameter are removed. The position of duplicated variable in final report is decided by first time it’s listed in outputorder parameter

# TEST DATA

A modified cars data set shipped with SAS package in SAShelp library is used to demonstrate all examples. For this complete data set, some missing data points are created to illustrate the macro features requiring missing data.

**data** car1;

set cars;

call streaminit(**12345**);

where drivetrain in ("Front" "Rear"); if drivetrain = "Front" then do;

gender = rand("uniform") > **0.5**;

if rand("uniform") > **0.9** then call missing(gender);

end;

else gender = rand("uniform") > **0.65**; temprandom = rand("uniform");

select;

when(temprandom > **0.6**) carmake = **1**;

when(temprandom > **0.3**) carmake = **2**;

otherwise carmake = **0**;

end;

if rand("uniform") > **0.95** then call missing(mpg\_city);

if rand("uniform") > **0.98** then call missing(mpg\_highway);

**run**;

# MACRO PARAMETERS INTRODUCTION

**Table 1 Macro Parameters**

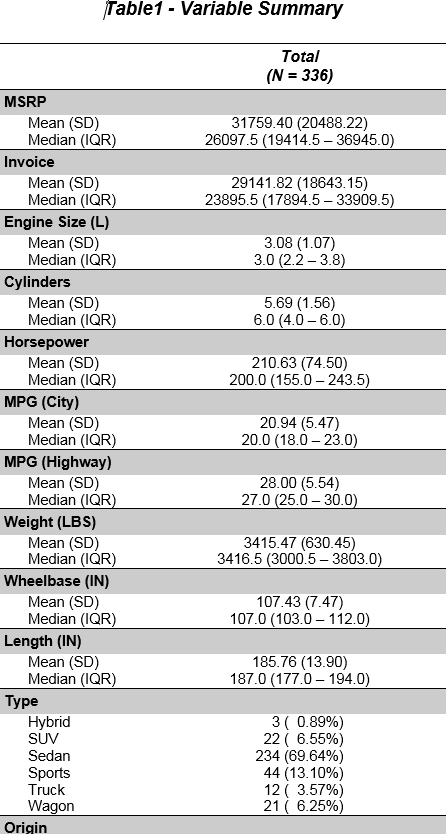
|  |  |
| --- | --- |
| **Parameter** | **Function** |
| dsn | Source data set name |
| group | Group variable name, if provided, report is stratified on this variable, p values are calculated using this variable as group indicator |
| calist | Categorical variable list, delimited by space |
| coplist | Continuous variable list requesting p values calculated using parametric test, including T-test and ANOVA, delimited by space |
| cononplist | Continuous variable list requesting p values calculated using non-parametric test, including Wilcoxon Rank Sum and Kruskal Wallis test, delimited by space |
| outputorder | Output order of variable in final report |
| continuousorder | Order of 2 types of continuous variable   1. - variables in coplist show first in final report 2. - variables in cononplist show first in final report |
| outdsn | Data set name storing information for final report |
| eratio | Ratio of number of cells with expected frequency less than 5, used for deciding type of test on variable in calist. If calculated ratio is greater than the value specified for this parameter, Fisher’s Exact test is performed for p value, otherwise, Chi-square test is conducted for p value. Value of this parameter must range between 0 and 1. By default, it’s set at 0.2. Setting it at 0 for only Fisher’s Exact test or 1 for only Chi-square test. |
| percenttype | Type of percentage reported for categorical variables Row - row percentage  Column - column percentage  Overall column only has column percentage |
| copmain | Main statistic for continuous variable in coplist, by default, it’s mean |
| copsupplement | Supplemental statistic for continuous variable in coplist, by default, it’s standard deviation |
| cononpmain | Main statistic for continuous variable in cononplist, by default, it’s median |
| cononpsupplement | Supplemental statistic for continuous variable in cononplist, by default, it’s inter-quarter range |
| caexcludelist | List of categorical variables not requesting p value calculation |
| copexcludelist | List of continuous variables in coplist not requesting p value calculation |
| cononpexcludelist | List of continuous variables in cononplist not requesting p value calculation |
| caincludemissing | Include missing information for categorical variables |
| copincludemissing | Include missing information for continuous variables in coplist |
| cononpincludemissing | Include missing information for continuous variables in cononplist |
| missingtop | Yes - missing information is put on top row of report for each variable  No - missing information is put on bottom row of report for each variable |

|  |  |
| --- | --- |
|  | By default, it’s yes |
| notest | If yes, P values are not calculated for all variables |
| cadec | Number of decimal on percentages calculated for categorical variable, by default, it’s 2 |
| copdec | Number of decimal on statistics calculated for continuous variable in coplist, by default, it’s 2 |
| cononpdec | Number of decimal on statistics calculated for continuous variable in cononplist, by default, it’s 1 |
| overall | If yes, add overall column to final report, by default, it’s no |
| paddingchar | Type of padding char used to line up numbers in the report 0 - no padding char   1. – 0, any font can be used for final report 2. – space, only typewriter style font can be used for best visualization By default, its 1 |
| reportinSAS | If yes, a copy of report shown on SAS, by default, it’s no |
| variableshading | If yes, background color of the cell with variable name/label is grey, by default, it’s yes |
| labelvariable | If yes, label categorical variable if Fisher’s Exact test is conducted for p value or continuous variable when T-test is conducted for p value and Satterthwaite method is used for calculating degree of freedom. By default, it’s no |
| suppresswarning | If yes, suppress all warning message from popping up, only shows in log window. By default, it’s yes. |
| createRTF | If yes, create a RTF file with final report. By default, it’s yes. |
| compacttable | If yes, final report is in compact style. By default, it’s yes. |
| tabletitle | Specify report table title |
| savefilename | Specify name and location of RTF file |
| showgroup | If yes, show group variable label/name in final report, by default, it’s yes |
| orderbyformat | If yes, order categories of categorical variable with their formatted value, by default, it’s no |
| pvaluetop | For continuous variable, put p value on top row of their report, by default, it’s no |
| orientation | Set up the file orientation of final report, by default, it’s portrait for report without a group variable and landscape for report with a group variable provided |
| missingpercent | If yes, provide percentage calculation for categorical variable missing information row, by default, it’s no |
| missinglabel | Row header for categorical variable missing row, by default, it’s Missing |
| NAlabel | Label for non-computable continuous variable statistic, by default, it’s NA |
| Showtotalcount | Specify if the total number of observation is shown in the final table |

# EXAMPLES

## CREATE VARIABLES SUMMARY WITHOUT GROUP VARIABLE AND P VALUES

|  |  |  |  |
| --- | --- | --- | --- |
| %Table1Macro(dsn | | = | car1, |
| calist | | = | type origin gender carmake, |
| coplist | = | msrp invoice enginesize cylinders horsepower mpg\_city  mpg\_highway weight wheelbase length, | |
| cononplist | = | msrp invoice enginesize cylinders horsepower mpg\_city mpg\_highway weight wheelbase length); | |

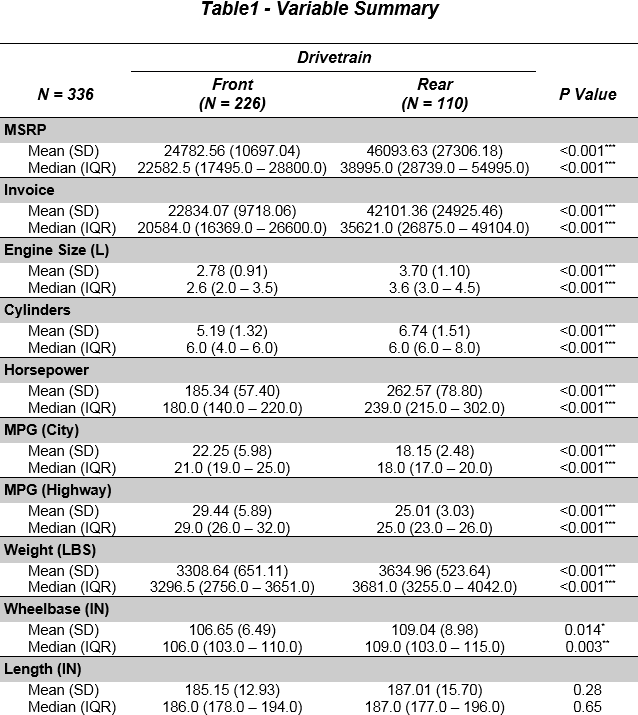


**Figure 1 Result of Example 1**

## CREATE VARIABLES SUMMARY WITH GROUP VARIABLE AND P VALUES

|  |  |  |
| --- | --- | --- |
| %Table1Macro(dsn | = | car1, |
| calist | = | type origin gender carmake, |
| coplist | = | msrp invoice enginesize cylinders horsepower mpg\_city mpg\_highway weight wheelbase length, |

|  |  |  |
| --- | --- | --- |
| cononplist | = | msrp invoice enginesize cylinders horsepower mpg\_city  mpg\_highway weight wheelbase length, |
| group | = | Drivetrain); |



**Figure 2 Result of Example 2**

## SPECIFY OUTPUT ORDER OF VARIABLES

|  |  |  |  |
| --- | --- | --- | --- |
| %Table1Macro(dsn | | = | car1, |
| calist | | = | type origin gender carmake, |
| coplist | | = | msrp invoice enginesize cylinders horsepower mpg\_city mpg\_highway weight wheelbase length, |
| cononplist | = | msrp invoice enginesize cylinders horsepower mpg\_city  mpg\_highway weight wheelbase length, |
| group | = | Drivetrain, |
| ***outputorder*** | = | gender carmake mpg\_city mpg\_highway type***);*** |

1. **REPORT VARIABLE MISSING INFORMATION**

|  |  |  |
| --- | --- | --- |
| %Table1Macro(dsn | = | car1, |
| calist | = | type origin gender carmake, |
| coplist | = | msrp invoice enginesize cylinders horsepower mpg\_city mpg\_highway weight wheelbase length, |
| cononplist | = | msrp invoice enginesize cylinders horsepower mpg\_city mpg\_highway weight wheelbase length, |
| group | = | Drivetrain, |
| outputorder | = | gender carmake mpg\_city mpg\_highway type, |
| caincludemissing | = | yes, |
| copincludemissing | = | yes, |
| cononpincludemissing | = | yes); |

## REQUEST STATISTIC FOR CONTINUOUS VARIABLES OTHER THAN DEFAULT ONE

|  |  |  |
| --- | --- | --- |
| %Table1Macro(dsn | = | car1, |
| calist | = | type origin gender carmake, |
| coplist | = | msrp invoice enginesize cylinders horsepower mpg\_city mpg\_highway weight wheelbase length, |
| cononplist | = | msrp invoice enginesize cylinders horsepower mpg\_city mpg\_highway weight wheelbase length, |
| group | = | Drivetrain, |
| outputorder | = | gender carmake mpg\_city mpg\_highway type, |
| copmain | = | mean, |
| copsupplement | = | clm, |
| cononpmain | = | median, |
| cononpsupplement | = | range); |

1. **REQUEST ROW PERCENTAGE FOR CATEGORICAL VARIABLE INSTEAD OF COLUMN PERCENTAGE**

|  |  |  |
| --- | --- | --- |
| %Table1Macro(dsn | = | car1, |
| calist | = | type origin gender carmake, |
| coplist | = | msrp invoice enginesize cylinders horsepower mpg\_city mpg\_highway weight wheelbase length, |
| cononplist | = | msrp invoice enginesize cylinders horsepower mpg\_city mpg\_highway weight wheelbase length, |
| group | = | Drivetrain, |
| outputorder | = | gender carmake mpg\_city mpg\_highway type, |
| percenttype | = | row); |

## CHANGE OUTPUT DECIMALS FOR CATEGORICAL PERCENTAGES AND STATISTICS CALCULATED FOR CONTINUOUS VARIABLES

|  |  |  |  |
| --- | --- | --- | --- |
| ***%Table1Macro(dsn*** | | = | car1, |
| ***calist*** | | = | type origin gender carmake, |
| ***coplist*** | | = | msrp invoice enginesize cylinders horsepower mpg\_city mpg\_highway weight wheelbase length, |
| ***cononplist*** | = | msrp invoice enginesize cylinders horsepower mpg\_city  mpg\_highway weight wheelbase length, |
| ***group*** | = | Drivetrain, |
| ***outputorder*** | = | gender carmake mpg\_city mpg\_highway type, |
| ***cadec*** | = | 0, |
| ***copdec*** | = | 1, |
| ***cononpdec*** | = | 2***);*** |

1. **CHANGE ERATIO CUTOFF OF CHOOSING FISHER’S EXACT TEST TO TO CALCULATE P VALUES FOR CATEGORICAL VARIABLES**

|  |  |  |
| --- | --- | --- |
| ***%Table1Macro(dsn*** | = | car1, |
| ***calist*** | = | type origin gender carmake, |
| ***coplist*** | = | msrp invoice enginesize cylinders horsepower mpg\_city mpg\_highway weight wheelbase length, |
| ***cononplist*** | = | msrp invoice enginesize cylinders horsepower mpg\_city mpg\_highway weight wheelbase length, |
| ***group*** | = | Drivetrain, |
| ***outputorder*** | = | gender carmake mpg\_city mpg\_highway type, |
| ***eratio*** | = | 0.5***);*** |

## SPECIFY VARIABLES FOR WHICH P VALUE IS NOT CALCULATED

|  |  |  |
| --- | --- | --- |
| %Table1Macro(dsn | = | car1, |
| calist | = | type origin gender carmake, |
| coplist | = | msrp invoice enginesize cylinders horsepower mpg\_city mpg\_highway weight wheelbase length, |
| cononplist | = | msrp invoice enginesize cylinders horsepower mpg\_city  mpg\_highway weight wheelbase length, |
| group | = | Drivetrain, |
| outputorder | = | gender carmake mpg\_city mpg\_highway type, |
| caexcludelist | = | gender carmake, |
| copexcludelist | = | mpg\_city, |
| cononpexcludelist | = | mpg\_highway); |

1. **REQUEST NO P VALUE CALCULATION FOR ALL VARIABLES**

|  |  |  |
| --- | --- | --- |
| %Table1Macro(dsn | = | car1, |
| calist | = | type origin gender carmake, |
| coplist | = | msrp invoice enginesize cylinders horsepower mpg\_city mpg\_highway weight wheelbase length, |
| cononplist | = | msrp invoice enginesize cylinders horsepower mpg\_city mpg\_highway weight wheelbase length, |
| group | = | Drivetrain, |
| outputorder | = | gender carmake mpg\_city mpg\_highway type, |
| notest | = | yes); |

## INCLUDE THE OVERALL COLUMN IN THE FINAL REPORT

|  |  |  |
| --- | --- | --- |
| %Table1Macro(dsn | = | car1, |
| calist | = | type origin gender carmake, |
| coplist | = | msrp invoice enginesize cylinders horsepower mpg\_city mpg\_highway weight wheelbase length, |

|  |  |  |
| --- | --- | --- |
| cononplist | = | msrp invoice enginesize cylinders horsepower mpg\_city  mpg\_highway weight wheelbase length, |
| group | = | Drivetrain, |
| outputorder | = | gender carmake mpg\_city mpg\_highway type, |
| overall | = | yes); |

1. **REMOVE SHADING ON VARIABLE NAME LINE**

|  |  |  |
| --- | --- | --- |
| %Table1Macro(dsn | = | car1, |
| calist | = | type origin gender carmake, |
| coplist | = | msrp invoice enginesize cylinders horsepower mpg\_city mpg\_highway weight wheelbase length, |
| cononplist | = | msrp invoice enginesize cylinders horsepower mpg\_city mpg\_highway weight wheelbase length, |
| group | = | Drivetrain, |
| outputorder | = | gender carmake mpg\_city mpg\_highway type, |
| variableshading | = | no); |

## REQUEST SPACE AS THE PADDING CHAR TO LINE UP FILE RESULTS

|  |  |  |
| --- | --- | --- |
| %Table1Macro(dsn | = | car1, |
| calist | = | type origin gender carmake, |
| coplist | = | msrp invoice enginesize cylinders horsepower mpg\_city mpg\_highway weight wheelbase length, |
| cononplist | = | msrp invoice enginesize cylinders horsepower mpg\_city mpg\_highway weight wheelbase length, |
| group | = | Drivetrain, |
| outputorder | = | gender carmake mpg\_city mpg\_highway type, |
| paddingchar | = | 2); |

1. **REQUEST A COPY OF OUTPUT SHOWING ON SAS INTERNAL WEB BROWSER**

|  |  |  |
| --- | --- | --- |
| %Table1Macro(dsn | = | car1, |
| calist | = | type origin gender carmake, |
| coplist | = | msrp invoice enginesize cylinders horsepower mpg\_city mpg\_highway weight wheelbase length, |
| cononplist | = | msrp invoice enginesize cylinders horsepower mpg\_city mpg\_highway weight wheelbase length, |
| group | = | Drivetrain, |
| outputorder | = | gender carmake mpg\_city mpg\_highway type, |
| reportinSAS | = | yes); |

## LABEL CATEGORICAL VARIABLES IF FISHER’S EXACT TEST IS USED FOR THEIR P VALUE CALCULATION, OR CONTINUOUS VARIABLES IF SATTERTHWAITE METHOD IS CONDUCTED TO COMPUTE DEGREE OF FREEDOM WHEN USING T-TEST FOR THEIR P VALUE CALCULATION

|  |  |  |
| --- | --- | --- |
| ***%Table1Macro(dsn*** | = | car1, |
| ***calist*** | = | type origin gender carmake, |
| ***coplist*** | = | msrp invoice enginesize cylinders horsepower mpg\_city mpg\_highway weight wheelbase length, |
| ***cononplist*** | = | msrp invoice enginesize cylinders horsepower mpg\_city  mpg\_highway weight wheelbase length, |

|  |  |  |
| --- | --- | --- |
| ***group*** | = | Drivetrain, |
| ***outputorder*** | = | gender carmake mpg\_city mpg\_highway type, |
| ***labelvariable*** | = | yes***);*** |



# TIPS AND SPECIAL APPLICATION

In final report, variables with same label/name are shown under same section. Therefore, we can provide continuous variable a format, and summarized it in both continuous and categorical styles

**proc format**;

value invoicef low - <**25000** = "Less than 25,000"

**25000** - high = "At least 25,000";

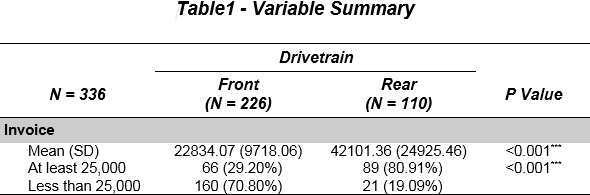
**run**;

**proc datasets** lib = work nolist; modify car1;

format invoice invoicef.;

**quit**;

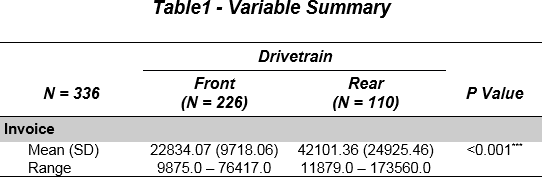
|  |  |  |
| --- | --- | --- |
| ***%Table1Macro(dsn*** | = | car1, |
| ***calist*** | = | invoice, |
| ***coplist*** | = | invoice, |
| ***group*** | = | Drivetrain***);*** |



**Figure 3 Result of Tip 1**

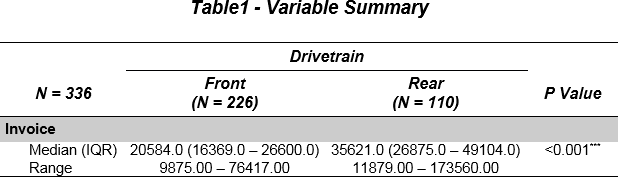
Report the third statistic for continuous variable

|  |  |  |
| --- | --- | --- |
| %Table1Macro(dsn | = | car1, |
| coplist | = | invoice, |
| cononplist | = | invoice, |
| cononpmain | = | range, |
| cononpsupplement | = | , |
| cononpexcludelist | = | invoice, |
| group | = | Drivetrain); |



**Figure 4 Result of Tip 2 (1)**

|  |  |  |
| --- | --- | --- |
| %Table1Macro(dsn | = | car1, |
| coplist | = | invoice, |
| cononplist | = | invoice, |
| copmain | = | range, |
| copsupplement | = | , |
| copexcludelist | = | invoice, |
| continuousorder | = | 2, |
| group | = | Drivetrain); |



**Figure 5 Result of Tip 2 (2)**

# CONCLUSION

1. Creating variable summary table is a time-consuming and error-prone process
2. This macro provides an easy way to automatically generate publication-ready variable summary table

# LIMITATION

1. Program is developed using SAS 9.4.3 and only fully tested in Windows operating system using Microsoft Word 2013
2. It might take very long time if Fisher’s Exact Test is performed to calculate p value for certain categorical variable. The p value calculation can be skipped by putting it in caexcludelist
3. Only up to 4 statistics can be reported for continuous variable
4. Total number of character in the name of all variable in source table should fulfill the formula: sum(length of all variable name) + number of all variable - 1 <= 65,534
5. Final result table background color can’t be freely changed if value of parameter paddingchar is set as 1