The SAS TABLE1 Macro

Mathew Pazaris, Ellen Hertzmark, and Donna Spiegelman

November 21, 2013

Abstract

The %table1 macro computes indirectly standardized rates, means, or proportions. The results are automatically prepared, by level of a given exposure variable, in a formatted MS Word table. The table is intended for use in publications with minimal additional formatting and/or preparation required. Table 1 of many papers is a breakdown of cohort characteristics by exposure categories. In most instances, it is necessary to age-standardize the means or proportions of other potential confounders before displaying them by exposure category.

Keywords: SAS, macro, standardization, table1, publication, cohort characteristics, age-standardization, confounders, exposure category

Contents

1 Description

The %table1 macro computes basic summary statistics, typically used in the Table 1 for publications. For continuous variables, the macro computes standardized means and standard deviations (SD). For categorical variables results are displayed as percentages. The macro computes the SD of the weighted data. This estimates the SD the group would have if it had had the same age distribution as the standard. In addition, basic summary statistics, including the sample size (n), mean, SD and standard error are displayed in a table in the SAS log-file. Here, the standard error of the standardized mean assumes that the weights relating to the means are constants, and that we do not take the between-age-group variance into account. This statistic however is for reference and is not included in the computed MS Word table.

2 Invocation and Details

In order to run this macro, your program must know where find it. You can tell SAS where to find macros by using the options

```
mautosource sasautos= <directories where macros are located>.
```

For example, at the Channing Lab, as described on chandoc.bwh.harvard.edu, under the Read Macros documentation. As an example, an options statement might be:

```
options nocenter ps=78 ls=125 replace formdlim='[';
mautosource
sasautos=('/usr/local/channing/sasautos',
   '/proj/nhsass/nhsas00/nhstools/sasautos);
```

This will allow you to use all the SAS read macros for the data sets (`/proj/nhsass/nhsas00/nhstools/sasautos`), as well as other Channing SAS macros, developed by our group, such as %PM, %INDIC3, %EXCLUDE, %LOGITR, and %MPHREG9.

The macro call is

```
\%table1(

data     = Dataset name. (required)
exposure = Exposure variable (required) - values must be consecutive integers
           starting at 1.
noexp    = Set to T if you do not want basic table characteristics given by
           level of an exposure (Default =F) (optional)
agegroup = Age or age-group variable for age-adjustment (required)
varlist  = List of variables. The order listed is the order they will
           be displayed in the table. (required)
header   = This option is for the SAS log header.  Not the Microsoft word
           document table.  (optional).
missing  = missing value for exposure (optional)
covar    = covariate list (optional)
printvar = This option has been removed and is no longer valid.

extstand = external standard age distribution (optional)
explab   = For the MS word table exposure levels are obtained from the
           data-set formats.  If formats are not used, up to 15 labels may
           used for the exposure level labels.  These labels are also used
           for the SAS log.  The defaults are:
           label0=level 0,
           label1=level 1, label2=level 2, label3=level 3,
           label4=level 4, label5=level 5,
           label6=level 6, label7=level 7, label8=level 8,
           label9=level 9, label10=level 10,
           label11=level 11, label12=level 12, label13=level 13,
           label14=level 14, label15=level 15, (optional)
notes    = Setting this option to notes tells the macro to print
           SAS Notes (Default = nonotes) (optional)

ageadj   = Setting this option to F tells the macro not to age-adjust any
           of the outcome variables.  (Default =T) (optional)

/*** MS Word table specific options WARING these options only apply to the
MS WORD table and do not effect the output in the SAS log.  The output in the
SAS log is for legacy and debuging purpouses only and should not be used for
final results ***/

nortf    = If set to T, this will prevent the macro from creating a MS
           Word table.  Use this option, if you are debugging your program
           and do not want a Word document created when the macro runs.
           Note: output will still be displayed in the SAS log-file
          (default=F) (optional),
noadj    = list of variables that should not be age-adjusted.  Age for
           example, should not be age adjusted and should be included in
           this list. (optional)
cat      = list of categorical variables.  All categorical variables should
           be included in this list. (optional)
rtftitle = title for the MS word table landscape - T tells the macro to print
```

```
            the MS Word table in landscape orientation.  The default is F for
            5 or less exposure levels, and T for greater than 5 levels. (optional)
fn        = Allows footnotes to be included in  the table Values should be
            separated by '@'   For example @variable1  footnote for variable 1
            @variable2 footnote for variable 2  - This would create a footnote
            'footnote for variable 1' for the variable 'variable1' and 'footnote
            for variable 2' for the variable 'variable2' (optional)
uselbl    = T if you would like to use label options instead of formatting
            and SAS labels to define rows and columns (optional),
file      = file name for the MS Word table.  Do not include ``.doc'', this
            will be appended automatically.) (required if nortf=F),
miscol    = In the MS Word tables, means, percentages and standard deviations are
            automatically calculated among non-missing.  Setting this option to T
            provides an additional column containing the percentage missing among
            the total. (Default = F) (optional)
poly      = List of ordinal variable that you would like to be automatically
            converted to polytomous categorical indicator variables by the macro.
            These variables MUST be labeled and formatted.  Only non-missing levels
            of the variable will be displayed in the table. The ordinal variable
            must also be included in the varlist, in order to identify the correct
            position for the set of variables in the table
            (optional)
polycat   = This is a list of polytomous categorical variables.  As opposed to the
            poly option, these variables should already be in the correct
            indicator variable format. For this indicator variable format, there must
            be a variable for each level of the polytomous variable set, even the
            reference level. Each variable, in the group, should be set to missing if
            the participant is missing the polytoumous variable. Here, a label and
            position in the table The variables are entered with the @ symbol,  no
            space, then a number indicating the position in the table, followed by a $
            symbol, then a space and the list of indicator variables.  For example,
            if we wanted to include indicators for age group, we might use the option
            ``polycat= @2 Age Group$ age1 age2 age3 age4 age5 age6''. This would provide
            6 age groups, labeled 'Age Group' and are included at the 2nd row of the
            table. (Note: when determining which row number for these variables, you
            must consider that variables included in the poly option
            increase the row count by the number of non-missing levels + 1).(optional)
fmtlib    = library containing the formats catalog for variables used in the table.
            (Default = work) (optional)
fmtcat    = format catalog containing formats for variables used in the table
            (default = formats) (optional)
multn     = setting this option to T results in labels for the column headers (levels
            of exposure) displaying both the
            number of observations, for the given exposure level, and the number of
            unique id records for that exposure. n/N format.  (default = F) (optional)
id        = unique identifier for each participant (default = id) (optional)
sep       = For continuous variables determines how the mean and standard deviation
            should be separated.  PAR = parenthesis PM = +/- (default = PAR) (optional)
dec       = Significant decimals printed in the MS Word Table.  Setting this option to
            1 will results in 1 significant digit, 2 would tell the macro to print 2
            significant digits etc... However, setting to 0 tells the macro to use the
            default algorithm.
);
```

3 Example 1 - Basic Macro call

The example is from the data-set from Feskanich (2008) Here we are interested in age standardized characteristics of the nurses who were postmenopausal at 1988 and looking at 5 levels of years spent working in rotating night shifts.

To start, we would like the column headings to reflect an appropriate description of the respective levels of shift work. We provide two options for column labels. The first option is to use the "explab" option. This option is used to manually assign a label to each level of shift work. Alternatively, we could format the dataset prior to inclusion in table 1 and the macro will automatically use the format values. We believe that formatting the data-set is a more robust options, as it allows us to reuse the same formats in multiple SAS procedures and helps to keep track of the variable coding throughout the analysis. So, we will start with a proc format statement and then assign the format the shift work variable shft. to.

```
proc format;
  value shftf
  1='never'
  2='1-2 yrs'
  3='3-9 yrs'
  4='10-19 yrs'
  5='20+ yrs';
data alldat;
  set alldat;
  format shft shftf.;
run;
```

Next, we would like to make sure the row labels reflect an appropriate description of the respective characteristics. By default, the variable name will be used as a row label, however, variable names do not often represent the variable meaning accurately. So, to better represent the variable meanings and to avoid additional table formatting of the MS Word table latter, we will assign labels to the population characteristic variables.

```
data alldat;
   set alldat;
label shift  = 'Years of rotating nightshift work'
      agegrp = 'Age, Groups'
      ageyr  = 'Age, years'
      alco   = 'Alcohol, g/day'
      bmi    = 'Body mass index, kg/mBYTE(178)'
      mets   = 'Activity, met-h/week'
      caff   = 'Caffeine, mg/day'
      calc   = 'Calcium, mg/day'
      chld   = 'Parity'
      cpmh   = 'HRT user'
      csmk   = 'Current smoker'
      null   = 'Nulliparous'
      ost    = 'Osteoporosis diagnosis'
      prot   = 'Protein, g/day'
      retn   = 'Retinol, BYTE(181)g/day'
      thz    = 'Thiazide diuretic user'
      vitd   = 'Vitamin D, BYTE(181)g/day';
 run;
```

Above you may notice some odd text, such as BYTE(178). BYTE is a UNIX function to represent ASCII characters using numeric representations. The table 1 macro has been developed to allow special characters in row and column headings, titles and footnotes, using the BYTE function. BYTE(178) represents a superscript 2 and BYTE(181) provides the mu symbol.

The full list of numeric extended-ASCII representations can be found online at various websites, such as

`http://www.idevelopment.info/data/Programming/ascii_table/PROGRAMMING_ascii_table.shtml`

The byte function is recommended whenever possible, however, not all special characters can be represented by the BYTE function. As an example, super script 1 and 2 may be used; however, there is no ASCII representation for superscripts 3 or higher. So, to allow further document enhancement, we have written the macro to also allow use of the SAS ODS control words as described in this SAS documentation:

`http://support.sas.com/rnd/base/ods/templateFAQ/Template_rtf.html#super`

In order to use SAS ODS control word, a control words an ODS escape character must be assigned, as described in the documentation. If you would like to use ODS control words, please note that for the %table1 macro, we have pre-assigned the caret character ^as the ODS escape character.

Now to run the macro:

```
-----------------------------------------------------------------
%table1(data=alldat,
agegroup=agegrp,
exposure=shift,
varlist=ageyr bmi mets csmk cpmh thz ost null chld calc
vitd prot retn alco caff,
noadj = ageyr,
cat= csmk cpmh thz ost null,
rtftitle=Age-standardized characteristics of the study
population of postmenopausal nurses at 1988 baseline
by number of years spent working in rotating night shifts,
landscape=F,
fn=@mets Metabolic equivalents from recreational and leisure-
time activities. @cpmh Postmenopausal hormone replacement
therapy. @chld Number of children among parous women.,
file = testf1, uselbl=F);

-----------------------------------------------------------------
```

- Here we have assigned the input dataset to alldat.

- The option agegroup lets us assign a standardization variable. Typically the macro is used for age-standardized tables, so it is called "agegroup", although you could standardize to a variable other than age.

- The option "varlist" is for the table rows. The order variables are entered here is the way they will appear in the table. All variables to be included in the table 1 rows must be included in the "varlist".

- We would like to include age in year "ageyr" however, it does not make sense to age-standardize age, so we have included it in the "noadj" option. "noadj" is simply a list of variables that should not be age standardized, however, they still must be listed in the "varlist' option.

- The "cat" option a list of categorical variables. The default is for continuous variables, and continuous variables are represented with their mean and SD. However, some variable are binary categorical and typically represented as a percentage. Note: Categorical variables should always be entered as binary indicator variables if there are multiple levels, unless polytomous categorical variables are used. In the case of polytomous variables, please see the polytomous variables section.

- The option rtftitle, allows us to assign a title to the table. Titles are automatically proceeded by "Table 1" in bold print.

- By default, if there are greater than 5 levels of the exposure variable, the table will be printed in landscape. This option however, can be manually over-ridden with the `landscape` option.

- The `fn` option allows footnotes to be included for outcome variables. The @ symbol is used to specify a new footnote and should be immediately followed with the variable name, a space, and then the footnote. Here we have assigned footnotes to the variables `(mets)`, `(cpmh)` and `(chld)`. Superscript letters are automatically assigned to the variable so, they should not be assigned manually to the variable labels.

- The `file` option is the name of the file that the table should be saved to. A .doc extension will automatically be appended.

- The `uselbl` option tells the macro whether or not to use the explab labels for the exposure variable level labeling, instead of the variable formats. By default the variable formats are used i.e. uselbl=F. However, if a variable format is missing, the value from the explab label option, for the given level of exposure will still be used (see explab option in section 2 for more information). explab labels are assigned through the macro variable options label1 through label15. If no label is assigned, the default label is "level n" where n is the exposure level. So, for the first level of the exposure, it would be "level 1", for the second "level 2" and so on, up to a maximum "level 15". Note: there is no maximum if you are instead using variable formats.

For the MS Word table, we provide percentages for categorical variables. Additionally, we have added formatting, such as title, footnotes and the levels of the exposure variable and outcome variables have been labeled based on our formats and labels respectively.

Table 1 Age-standardized characteristics of the study population of postmenopausal nurses at 1988 baseline by number of years spent working in rotating night shifts

	Years of rotating nightshift work				
	never (n=9043)	1-2\ yrs (n=4731)	3-9\ yrs (n=5022)	10-19\ yrs (n=1906)	20+\ yrs (n=1324)
Age, years*	56.5(5.2)	56.0(5.1)	56.7(5.0)	56.8(5.0)	58.0(4.8)
Body mass index, kg/m^2	24.1(5.7)	24.0(5.5)	24.4(5.8)	24.7(6.4)	25.1(6.7)
Activity, met-h/week[a]	12.5(14.4)	13.5(14.7)	13.9(15.1)	14.2(16.8)	14.5(16.3)
Current smoker, %	21	20	22	26	25
HRT user[b], %	37	38	36	34	29
Thiazide diuretic user, %	15	14	15	17	17
Osteoporosis diagnosis, %	6	5	6	7	6
Nulliparous, %	5	6	7	5	6
Parity[c]	3.4(1.6)	3.4(1.6)	3.3(1.6)	3.4(1.6)	3.4(1.8)
Calcium, mg/day	915(351)	918(341)	918(342)	909(326)	898(340)
Vitamin D, μg/day	8.2(5.5)	8.1(5.4)	8.2(5.1)	8.2(5.1)	8.3(5.5)
Protein, g/day	74.1(11.5)	74.3(11.3)	75.0(11.4)	74.4(11.7)	74.2(11.9)
Retinol, μg/day	1368(1270)	1396(1379)	1403(1322)	1430(1273)	1461(1433)
Alcohol, g/day	6.6(10.0)	6.4(9.7)	6.4(9.8)	5.8(9.5)	5.4(9.4)
Caffeine, mg/day	353(218)	351(216)	364(219)	387(227)	391(242)

Values are means(SD) or percentages and are standardized to the age distribution of the study population.
[a] Metabolic equivalents from recreational and leisure- time activities.
[b] Postmenopausal hormone replacement therapy.
[c] Number of children among parous women.
* Value is not age adjusted

4 Example 2 - No Exposure

If we want to run the macro without an exposure, in order to see table 1 statistics for the dataset overall, we can use the `noexp` option as follows Note: when doing so, the data will no be age-standardized:

```
---------------------------------------------------------------
%table1(data=alldat,
varlist=ageyr bmi mets csmk cpmh thz ost null chld calc
vitd prot retn alco caff,
cat= csmk cpmh thz ost null,
rtftitle=Characteristics of the study
population of postmenopausal nurses at 1988 baseline
by number of years spent working in rotating night shifts,
landscape=F,
fn=@mets Metabolic equivalents from recreational and leisuretime
activities. @cpmh Postmenopausal hormone replacement
therapy. @chld Number of children among parous women.,
file = testf2, uselbl=F, noexp=t);
---------------------------------------------------------------
```

Table 1 Characteristics of the study population of postmenopausal nurses at 1988 baseline

	(n=22026)
Age, years	56.5(5.1)
Body mass index, kg/m^2	24.3(5.8)
Activity, met-h/week[a]	13.3(15.0)
Current smoker, %	21
HRT user[b], %	36
Thiazide diuretic user, %	15
Osteoporosis diagnosis, %	6
Nulliparous, %	6
Parity[c]	3.4(1.6)
Calcium, mg/day	915(344)
Vitamin D, μg/day	8.2(5.3)
Protein, g/day	74.4(11.5)
Retinol, μg/day	1393(1315)
Alcohol, g/day	6.4(9.8)
Caffeine, mg/day	360(220)

Values are means(SD) or percentages.

[a] Metabolic equivalents from recreational and leisuretime activities.

[b] Postmenopausal hormone replacement therapy.

[c] Number of children among parous women.

5 Example 3 - Calculating number of participants and number of observations

If we have repeated measures data, with multiple records per participant, We may want to display both the number of records and the number of participants. Or, we may want to use the macro to confirm that there is only one record per participant. To do so, we can use the `multn` option as follows. In this example, there should only be one record per participant, so N and n should be the same. We are happy to see that this is the case.

```
%table1(data=alldat,
agegroup=agegrp,
exposure=shift,
varlist=ageyr bmi mets csmk cpmh thz ost null chld calc
vitd prot retn alco caff,
noadj = ageyr,
cat= csmk cpmh thz ost null,
rtftitle=Age-standardized characteristics of the study
population of postmenopausal nurses at 1988 baseline
by number of years spent working in rotating night shifts,
landscape=F,
fn=@mets Metabolic equivalents from recreational and leisure-
time activities. @cpmh Postmenopausal hormone replacement
therapy. @chld Number of children among parous women.,
file = testf3, uselbl=F, multn=t);
```

Table 1 Age-standardized characteristics of the study population of postmenopausal nurses at 1988 baseline by number of years spent working in rotating night shifts

	Years of rotating nightshift work				
	never (n\ N = 9043 \ 9043)	1-2\ yrs (n\ N = 4731 \ 4731)	3-9\ yrs (n\ N = 5022 \ 5022)	10-19\ yrs (n\ N = 1906 \ 1906)	20+\ yrs (n\ N = 1324 \ 1324)
Age, years*	56.5(5.2)	56.0(5.1)	56.7(5.0)	56.8(5.0)	58.0(4.8)
Body mass index, kg/m²	24.1(5.7)	24.0(5.5)	24.4(5.8)	24.7(6.4)	25.1(6.7)
Activity, met-h/week[a]	12.5(14.4)	13.5(14.7)	13.9(15.1)	14.2(16.8)	14.5(16.3)
Current smoker, %	21	20	22	26	25
HRT user[b], %	37	38	36	34	29
Thiazide diuretic user, %	15	14	15	17	17
Osteoporosis diagnosis, %	6	5	6	7	6
Nulliparous, %	5	6	7	5	6
Parity[c]	3.4(1.6)	3.4(1.6)	3.3(1.6)	3.4(1.6)	3.4(1.8)
Calcium, mg/day	915(351)	918(341)	918(342)	909(326)	898(340)
Vitamin D, μg/day	8.2(5.5)	8.1(5.4)	8.2(5.1)	8.2(5.1)	8.3(5.5)
Protein, g/day	74.1(11.5)	74.3(11.3)	75.0(11.4)	74.4(11.7)	74.2(11.9)
Retinol, μg/day	1368(1270)	1396(1379)	1403(1322)	1430(1273)	1461(1433)
Alcohol, g/day	6.6(10.0)	6.4(9.7)	6.4(9.8)	5.8(9.5)	5.4(9.4)
Caffeine, mg/day	353(218)	351(216)	364(219)	387(227)	391(242)

Values are means(SD) or percentages and are standardized to the age distribution of the study population.
[a] Metabolic equivalents from recreational and leisure- time activities.
[b] Postmenopausal hormone replacement therapy.
[c] Number of children among parous women.
* Value is not age adjusted

6 Example 4 - polytomous categorical variables using the poly option (recommended)

Here instead of using the continuous age variable `ageyr`, we would like to include categories of age, derived from the ordinal age-group variable `agegrp`. Here we format and label age group and then include it in the `poly` option. Notice that there are 9 levels of agegroup formats, but only 6 levels in the table. This is because there is no one older than the 65-69 age group. The poly option will only include non-missing categories in the table. We also must include it in the `varlist`, this is required, so that the macro knows which row the age group variables should be placed in the table. Here, we have included agegrp in the first row. Finally, because we do not want to standardize `agegrp`, we must add it to the `noadj` option.

```
proc format;
  value agegrpf
  1='< 45 yrs'
  2='45-49 yrs'
  3='50-54 yrs'
  4='55-59 yrs'
  5='60-64 yrs'
  6='65-69 yrs'
  7='70-74 yrs'
  8='75-79 yrs'
  9='80+ yrs';
```

```
run;

data alldat;
set alldat;
label agegrp = 'Age Group';
format agegrp agegrpf.;
run;

%table1(data=alldat,
exposure=shift,
agegroup=agegrp,
varlist=agegrp bmi mets csmk cpmh thz ost null chld calc
vitd prot retn alco caff,
noadj = agegrp,
cat= csmk cpmh thz ost null,
rtftitle=Age-standardized characteristics of the study
population of postmenopausal nurses at 1988 baseline
by number of years spent working in rotating night shifts,
landscape=F,
fn=@mets Metabolic equivalents from recreational and leisuretime
activities. @cpmh Postmenopausal hormone replacement
therapy. @chld Number of children among parous women.,
file = testf4, uselbl=F, poly=agegrp);
```

Table 1 Age-standardized characteristics of the study population of postmenopausal nurses at 1988 baseline by number of years spent working in rotating night shifts

	Years of rotating nightshift work				
	never (n=9043)	1-2\ yrs (n=4731)	3-9\ yrs (n=5022)	10-19\ yrs (n=1906)	20+\ yrs (n=1324)
Age Group*					
- 45 yrs, %	2	2	1	2	1
- 45-49 yrs, %	9	10	9	8	5
- 50-50 yrs, %	26	29	25	24	20
- 55-59 yrs, %	36	35	37	38	37
- 60-64 yrs, %	23	21	24	24	32
- 65-69 yrs, %	4	3	4	4	6
Body mass index, kg/m^2	24.1(5.7)	24.0(5.5)	24.4(5.8)	24.7(6.4)	25.1(6.7)
Activity, met-h/week[a]	12.5(14.4)	13.5(14.7)	13.9(15.1)	14.2(16.8)	14.5(16.3)
Current smoker, %	21	20	22	26	25
HRT user[b], %	37	38	36	34	29
Thiazide diuretic user, %	15	14	15	17	17
Osteoporosis diagnosis, %	6	5	6	7	6
Nulliparous, %	5	6	7	5	6
Parity[c]	3.4(1.6)	3.4(1.6)	3.3(1.6)	3.4(1.6)	3.4(1.8)
Calcium, mg/day	915(351)	918(341)	918(342)	909(326)	898(340)
Vitamin D, μg/day	8.2(5.5)	8.1(5.4)	8.2(5.1)	8.2(5.1)	8.3(5.5)
Protein, g/day	74.1(11.5)	74.3(11.3)	75.0(11.4)	74.4(11.7)	74.2(11.9)
Retinol, μg/day	1368(1270)	1396(1379)	1403(1322)	1430(1273)	1461(1433)
Alcohol, g/day	6.6(10.0)	6.4(9.7)	6.4(9.8)	5.8(9.5)	5.4(9.4)
Caffeine, mg/day	353(218)	351(216)	364(219)	387(227)	391(242)

Values are means(SD) or percentages and are standardized to the age distribution of the study population.
Values of polytomous variables may not sum to 100% due to rounding
[a] Metabolic equivalents from recreational and leisuretime activities.
[b] Postmenopausal hormone replacement therapy.
[c] Number of children among parous women.
* Value is not age adjusted

7 Example 5 - Polytomous categorical variables using the polycat option (alternative)

The alternative to using the `poly` option, is to use the `polycat` option. The advantage of the `polycat` option over the `poly` option, is that it offers more control. However, it is more difficult to implement. Here, again, instead of using the continuous age variable `ageyr`, we would like to include categories of age, derived from the ordinal age-group variable `agegrp`.

In order to use the `polycat` option, we must first set up categorical age group variables. Here, we will use the indic3 macro to do so. To do so, fist, we must set `usemiss` to 0, because, the table 1 macro should calculate percentages out of non-missing. So, we want all indicator variables to be missing if agegrp is missing (note: in this example, there are not any observations with missing however, we discuss missing for demonstration purposes).

Next, although the `indic3` macro has a label option, it encapsulates the labels in quotation marks. For our purposes, this is undesirable, so we will label the variables ourselves.

For the `indic3` macro, the reference level `reflev` is always named with an "r" suffix instead of a number. In this case, we have assigned `reflev` = 1, so the first level is called `ager`, not `age1`.

Finally, the macro will not automatically limit polytomous variables to non-missing categories when using the `polycat` option, so we must do so manually. Therefore, we only include `age1-age6`.

To run the macro with the `polycat` option, we must specify the beginning of the variable set with the '@' sign then a row position indicator, here the row is 2, so the set will be included as the second row. No polycat variables should be included in `varlist` option, however, we need to include the entire list of indicators in the `noadj` option, if we would like them to be age-adjusted. When polycat variables are included in `noadj` they must be entered in the same order as in the `polycat` option.

```
data alldat;
set alldat;
\%indic3(vbl=agegrp, prefix=age, reflev=1, min=1, max=9, usemiss=0);

label ager='< 45 yrs'
      age2='45-49 yrs'
      age3='50-54 yrs'
      age4='55-59 yrs'
      age5='60-64 yrs'
      age6='65-69 yrs';
run;



%table1(data=alldat,
agegroup=agegrp,
exposure=shift,
varlist=bmi mets csmk cpmh thz ost null chld calc
vitd prot retn alco caff,
noadj =ager age2 age3 age4 age5 age6,
cat= csmk cpmh thz ost null,
rtftitle=Age-standardized characteristics of the study
population of postmenopausal nurses at 1988 baseline
by number of years spent working in rotating night shifts,
landscape=F,
fn=@mets Metabolic equivalents from recreational and leisure time
activities. @cpmh Postmenopausal hormone replacement
therapy. @chld Number of children among parous women.,
file = testf5, polycat= @2 Age Group$ ager age2
age3 age4 age5 age6);
```

Table 1 Age-standardized characteristics of the study population of postmenopausal nurses at 1988 baseline by number of years spent working in rotating night shifts

	Years of rotating nightshift work				
	never (n=9043)	1-2\ yrs (n=4731)	3-9\ yrs (n=5022)	10-19\ yrs (n=1906)	20+\ yrs (n=1324)
Body mass index, kg/m^2	24.1(5.7)	24.0(5.5)	24.4(5.8)	24.7(6.4)	25.1(6.7)
Age Group*					
- < 45 yrs, %	2	2	1	2	1
- 45-49 yrs, %	9	10	9	8	5
- 50-54 yrs, %	26	29	25	24	20
- 55-59 yrs, %	36	35	37	38	37
- 60-64 yrs, %	23	21	24	24	32
- 65-69 yrs, %	4	3	4	4	6
Activity, met-h/week[a]	12.5(14.4)	13.5(14.7)	13.9(15.1)	14.2(16.8)	14.5(16.3)
Current smoker, %	21	20	22	26	25
HRT user[b], %	37	38	36	34	29
Thiazide diuretic user, %	15	14	15	17	17
Osteoporosis diagnosis, %	6	5	6	7	6
Nulliparous, %	5	6	7	5	6
Parity[c]	3.4(1.6)	3.4(1.6)	3.3(1.6)	3.4(1.6)	3.4(1.8)
Calcium, mg/day	915(351)	918(341)	918(342)	909(326)	898(340)
Vitamin D, μg/day	8.2(5.5)	8.1(5.4)	8.2(5.1)	8.2(5.1)	8.3(5.5)
Protein, g/day	74.1(11.5)	74.3(11.3)	75.0(11.4)	74.4(11.7)	74.2(11.9)
Retinol, μg/day	1368(1270)	1396(1379)	1403(1322)	1430(1273)	1461(1433)
Alcohol, g/day	6.6(10.0)	6.4(9.7)	6.4(9.8)	5.8(9.5)	5.4(9.4)
Caffeine, mg/day	353(218)	351(216)	364(219)	387(227)	391(242)

Values are means(SD) or percentages and are standardized to the age distribution of the study population.
Values of polytomous variables may not sum to 100% due to rounding

[a] Metabolic equivalents from recreational and leisure time activities.

[b] Postmenopausal hormone replacement therapy.

[c] Number of children among parous women.

* Value is not age adjusted

8 A note about row numbering when using the polycat and poly options together

As seen above, determining the row placement of variables in the `poly` option is determined by placement in `varlist`. Row placement in for variables in the `polycat` option is determined by a number following the @ symbol. We should note however, that there is one other important difference. It is assumed that if someone wants to use `poly` then they will use `poly` consistently and if they use `polycat` then they will use `polycat` consistently. Although, we do allow the use of both `poly` and `polycat` together, however, there may be some unexpected results.

When a `polycat` variable is included after a `polycat` variable, the preceding group of indicators is treated as 1 variable. So, the first `polycat` variable could be at @2 and the second at @3. However, when a `poly` variable is included before a `polycat` variable, the group of indicator variables for the `poly` variable is NOT treated as 1 variable, with regards to row numbering, it is treated as the number of categories. So, if we have a `poly` option variable set at the 2nd row, and we want a `poltcat` variable to follow it, then the `polycat` variable can not be @3, but must instead be @(3 + 6). So, to position a `polycat` variable after the set of agegrp variables, the `polycat` variable must be @4, because 3+6 = 9. This is because there were 6 non-missing categories for `agegrp`.

9 Example 6 - Missing Column

The table 1 macro calculates percentages out of non-missing values. Sometimes however, we would like to know what percentage of observations, for a given variable are missing. We can do so by using the `miscol` option.

```
-------------------------------------------------------------

%table1(data=alldat,
agegroup=agegrp,
exposure=shift,
varlist=bmi mets csmk cpmh thz ost null chld calc
vitd prot retn alco caff,
noadj =ager age2 age3 age4 age5 age6,
cat= csmk cpmh thz ost null,
rtftitle=Age-standardized characteristics of the study
population of postmenopausal nurses at 1988 baseline
by number of years spent working in rotating night shifts,
landscape=F,
fn=@mets Metabolic equivalents from recreational and leisure time
activities. @cpmh Postmenopausal hormone replacement
therapy. @chld Number of children among parous women.,
file = testf6, polycat= @2 Age Group$ ager age2
age3 age4 age5 age6, miscol=t);

-------------------------------------------------------------
```

Table 1 Age-standardized characteristics of the study population of postmenopausal nurses at 1988 baseline by number of years spent working in rotating night shifts

	Years of rotating nightshift work					
	never (n=9043)	1-2\ yrs (n=4731)	3-9\ yrs (n=5022)	10-19\ yrs (n=1906)	20+\ yrs (n=1324)	Missing, %
Body mass index, kg/m²	24.1(5.7)	24.0(5.5)	24.4(5.8)	24.7(6.4)	25.1(6.7)	7.6
Age Group*						0.0
- < 45 yrs, %	2	2	1	2	1	
- 45-49 yrs, %	9	10	9	8	5	
- 50-54 yrs, %	26	29	25	24	20	
- 55-59 yrs, %	36	35	37	38	37	
- 60-64 yrs, %	23	21	24	24	32	
- 65-69 yrs, %	4	3	4	4	6	
Activity, met-h/week[a]	12.5(14.4)	13.5(14.7)	13.9(15.1)	14.2(16.8)	14.5(16.3)	0.2
Current smoker, %	21	20	22	26	25	0.0
HRT user[b], %	37	38	36	34	29	5.1
Thiazide diuretic user, %	15	14	15	17	17	0.0
Osteoporosis diagnosis, %	6	5	6	7	6	0.0
Nulliparous, %	5	6	7	5	6	1.8
Parity[c]	3.4(1.6)	3.4(1.6)	3.3(1.6)	3.4(1.6)	3.4(1.8)	7.2
Calcium, mg/day	915(351)	918(341)	918(342)	909(326)	898(340)	5.2
Vitamin D, μg/day	8.2(5.5)	8.1(5.4)	8.2(5.1)	8.2(5.1)	8.3(5.5)	5.2
Protein, g/day	74.1(11.5)	74.3(11.3)	75.0(11.4)	74.4(11.7)	74.2(11.9)	5.2
Retinol, μg/day	1368(1270)	1396(1379)	1403(1322)	1430(1273)	1461(1433)	5.2
Alcohol, g/day	6.6(10.0)	6.4(9.7)	6.4(9.8)	5.8(9.5)	5.4(9.4)	5.2
Caffeine, mg/day	353(218)	351(216)	364(219)	387(227)	391(242)	5.2

Values are means(SD) or percentages and are standardized to the age distribution of the study population. Values of polytomous variables may not sum to 100% due to rounding

[a] Metabolic equivalents from recreational and leisure time activities.

[b] Postmenopausal hormone replacement therapy.

[c] Number of children among parous women.

* Value is not age adjusted

10 Example 7 - Missing values and PM option

In Some cases some levels of an exposure may not be applicable for some variables, such as in the following example. Here we are looking at an NHS II Vitamin D case-control study and have set the exposure to be cases or controls. In this example, variables related to diagnosis are only applicable to cases, and not controls. Therefore the data is missing for the controls. In this case, the macro prints blank space when there are empty cells.

Also, in this example, we would like to separate the means and standard deviation with a plus/minus sign rather than parenthesis. This is done using the `SEP` option and setting it to "PM".

```
-------------------------------------------------------------
        %table1(data=bldall,
         agegroup=agecat,

         exposure=caco,
         varlist=agebld menarc bmi18 bmibld wtchngc everoc95 duroc nulli par afb brstfd
         prebld famhx bbdhx vitd lapsemo predx invas erprposinv erprneginv,
         noadj = agebld menarc bmi18 afb,
         cat= everoc95 nulli brstfeed mstatbld famhx bbdhx invas predx prebld erprposinv
         erprneginv brstfd,
         rtftitle=Age-standardized characteristics of the blood study population with Vit D,
         landscape=F,
         fn=@duroc Duration of OC use from 1995 qq among women with everoc95=yes
            @par children among women with children at 1995qq
            @afb Age at first birth among women with children indicated on 1995qq
            @brstfd Percent of breastfeeding among women with children on 1995qq
            @erprposinv ER+/PR+ among invasive breast cancer
            @erprneginv ER-/PR- among invasive breast cancer,
         file = vitdexmple, uselbl=F, sep=PM);

-------------------------------------------------------------
```

Table 1 Age-standardized characteristics of the blood study population with Vit D

	Case Status	
	Case (n=613)	Control (n=1218)
Age at blood collection*	45.0 ± 4.4	44.8 ± 4.4
Age at menarche*	12.4 ± 1.3	12.4 ± 1.4
BMI at age 18, kg/m2*	20.7 ± 2.9	21.1 ± 3.0
BMI at blod collection, kg/m2	25.5 ± 6.3	26.3 ± 7.1
Weight change since age 18	11.6 ± 12.1	12.9 ± 13.8
Ever used OCs , %	86	86
Duration of past OC use (continuous)[a]	47.5 ± 46.9	45.7 ± 45.5
Nulliparous , %	22	20
Parity[b]	2.2 ± 0.9	2.3 ± 0.9
Age at first birth*[c]	26.6 ± 4.6	26.2 ± 4.6
Ever breast fed[d] , %	79	80
Premenopausal at blood collection , %	77	76
Family history of breast cancer , %	17	10
History of benign breast disease , %	23	16
Plasma 25(OH)D level, ng/mL	25.4 ± 9.5	25.0 ± 9.6
Time from blood collection to diagnosis	57.6 ± 33.5	
Premenopausal at diagnosis , %	69	
Invasive , %	71	
ER+/PR+[e] , %	66	
ER-/PR-[f] , %	17	

Values are means ± SD or percentages and are standardized to the age distribution of the study population.

[a] Duration of OC use from 1995 qq among women with everoc95=yes

[b] children among women with children at 1995qq

[c] Age at first birth among women with children indicated on 1995qq

[d] Percent of breastfeeding among women with children on 1995qq

[e] ER+/PR+ among invasive breast cancer

[f] ER-/PR- among invasive breast cancer

* Value is not age adjusted

11 Errors

Most errors with the %table1 macro are likely due to missing required options in the macro statement itself. Depending on the options used, there are up to 4 potential errors to missing required options for the macro.

"ERROR in macro call: You did not provide an exposure variable"

The exposure variable is required under all circumstances, unless noexp=T is specified. There is a default for the exposure option, which is just "exposure", however, if no exposure variable is listed, and there is no exposure variable in the data-set, you will receive this error.

"ERROR in macro call: You did not provide a file name for the MS Word table"

By default the %table1 macro will create a MS Word document containing table 1. This will not occur if the "nortf" option is explicitly set to "T", however, if not set to "T" or set to something other than "T", then a file name is required in the "file" option. In this case, if there is no file listed, you will receive this error.

"ERROR in macro call: You did not provide a variable for age-adjustment "

The table 1 macro is intended to use with adjusted data, which is typically age-adjusted. You may adjusted, on some variable other than age, if desired, however, that variable must still be listed in the "agegroup" option. If the agegroup option is left blank and you will get this error. The exception, is if you set the "ageadj" option to something other than "T". By default the "ageadj" option is set to T, which tells the macro to adjust the data. Setting to something else, such as "F", tells the macro not to adjust the data, and therefore no agegroup variable is required.

"ERROR in macro call: You did not provide a list of variables"

All outcome variables must be listed in the "varlist" option. Even if they are included in the "cat" or "noadj" options, they are still required in the "Varlist" option. This is required, so that the macro knows the proper ordering for the variables in the MS Word table.

"ERROR in macro call: You need to provide more than one variable in varlist"

The "varlist" option must contain more than one variable. The table 1 macro is not intended to examine individual variables.

"ERROR in macro call: You have included nonexistent variable(s) "missing variable name here" in the table1 call"

This error will occur if you try to include variables that are not found in the dataset.

"ERROR in macro call: You have included duplicate variable(s) "duplicate variable name here" in the table1 call"

This error will occur if you try to include variables more than once in the varlist statement.

12 Credits

Written by Mathew Pazaris, Ellen Hertzmark, and Donna Spiegelman Adapted from a program written by Eric Rimm for the Channing Laboratory. Questions can be directed to Mathew Pazaris `stmjp@channing.harvard.edu`,

13 References

Nightshift work and fracture risk: the Nurses' Health Study. Feskanich D, Hankinson SE, Schernhammer ES. Osteoporos Int. 2009 Apr;20(4):537-42.