**Summary Table – Univariate Association with a Categorical Outcome**

**Macro Name:** UNI\_CAT

**Created Date/Author:** Feb. 2012/Yuan Liu

**Last Update Date/Person**: Oct 06, 2015/Yuan Liu

**Other Contributors**: Dana Nickleach

**Current Version**: V30

**Working Environment:** SAS 9.3 English version

**Contact**: Dr. Yuan Liu [yliu31@emory.edu](mailto:yliu31@emory.edu)

**Purpose:** To conduct a univariate analysis for a categorical outcome with a list of covariates, individually. For categorical covariates, a contingency table along with the Chi-square test (parametric p-value) or Fisher’s exact test (non-parametric p-value) can be produced. For numerical covariates, the sample size, mean and median along with ANOVA test (parametric p-value) or Kruskal-Wallis test (non-parametric p-value) can be produced.

**Notes:** 1) The order of variables in the summary table is the same as the input order. For the best results, you may want to put the demographic variables together and also clinical characteristics variables together; 2) The biostatistician may need to help investigator to decide which statistics (parametric or non-parametric p-value) is more appropriate for the data.

**Parameters:**

|  |  |  |
| --- | --- | --- |
| **Macro variable** | **Description** | **Required** |
| DATASET | The name of the data set to be analyzed. | Yes |
| OUTCOME | Categorical variable to be associated with CLIST and NLIST variables. More than one variable can be listed separated by empty space. However, these variables appear in the table header and too many variables will cause the table to wrap due to the document page width limitations, producing undesirable results. Each variable name must not be more than 30 characters long. | Yes |
| CLIST | List of categorical variables, separated by empty space. | Yes |
| NLIST | List of numerical variables, separated by empty space. | Yes |
| NONPAR | Specify a value of F, T, or A to indicate whether to conduct non-parametric tests. If the value is T then both parametric and non-parametric tests will be conducted. If the value is F then only parametric tests will be conducted. A value of A means that for categorical variables, the appropriate test statistic, non-parametric or parametric, will be automatically chosen based on whether the chi-square test is invalid, but for numerical covariates only the parametric test will be calculated. Option A is only available for SAS V9.3 or later. The default value is F. | No |
| SPREAD | Set to T to also report standard deviation, min, and max for numerical variables. The default value is F. | No |
| BY | A separate analysis will be conducted for each value of the variable specified here. | No |
| MHC | Set to T to report p-values from Mantel-Haenszel chi-square tests instead of Pearson chi-square tests. The default value is F. | No |
| DOC | Set to F to suppress creation of the RTF file. The default value is T. | No |
| OUTPATH | File path for output table to be stored. | Yes |
| FNAME | File name for output table. | Yes |
| ROWPERCENT | Set to F to report column percentages instead of row percentages from the contingency table. The default value is T. | No |
| ORIENTATION | Value of PORTRAIT or LANDSCAPE to indicate the page layout of the report. The default value is PORTRAIT. | No |
| WEIGHT | Weight variable to use in a WEIGHT statement. Weights will not be normalized by the macro. The reported N will be the sum of the weights. This option will not work with NONPAR = T or A. Leave blank if not weighting. | No |
| MATCHID | If your data is from a matched sample, indicate the id variable that links matched pairs. If not, then leave this blank. If you specify a MATCHID then McNemar's test for two level categorical variables and Bowker's test of symmetry for more than two level categorical variables will be conducted instead of a chi-square test. A paired t-test as opposed to ANOVA for numerical variables will be conducted. This option is currently not set up to conduct non-parametric tests. This option is also only appropriate for 1-1 matching. Note that the data set should be in the format of one observation per subject, not one observation per match. The data will be transformed as needed in order to conduct the necessary tests. The OUTCOME variable should correspond to the variable that identifies the repeated measurement. For example, if a patient had one measurement on their right foot and one on their left then the OUTCOME variable should be foot. The OUTCOME variable should have two categories. | No |
| DEBUG | Set to T if running in debug mode. Work datasets will not be deleted in debug mode. This is useful if you are editing the code or want to further manipulate the resulting data sets. The default value is F. | No |

**Usage Example:**

**DATA** analysis;

input id os\_censor Sex $ Age duration os progress $ trt $;

LABEL os = 'Overall Survival (months)'

progress = 'Progression'

trt = 'Treatment'

duration = 'Duration of Radiation';

DATALINES;

1 1 M 40 44 20 No B

2 1 F 45 46 16 Yes A

3 1 F 40 32 20 No B

4 1 F 47 32 23 No B

5 0 M 41 25 22 No B

6 1 M 54 35 13 No B

7 1 M 48 50 9 Yes A

8 1 M 36 33 12 Yes B

9 0 F 49 51 8 Yes A

10 1 M 49 52 10 Yes A

11 1 M 44 35 12 No A

12 1 M 49 50 8 Yes A

13 1 M 44 44 14 Yes A

14 1 M 50 31 10 Yes A

15 1 M 53 40 15 No B

16 0 M 52 29 20 Yes B

17 1 F 46 45 5 Yes A

18 1 F 37 44 11 Yes A

19 1 M 49 46 13 No B

20 1 M 42 31 11 No A

;

TITLE 'Table 2 Univariate Association with Progression';

%UNI\_CAT(dataset = analysis,

outcome = progress,

clist = sex trt,

nlist = age duration,

nonpar = T,

rowpercent = F,

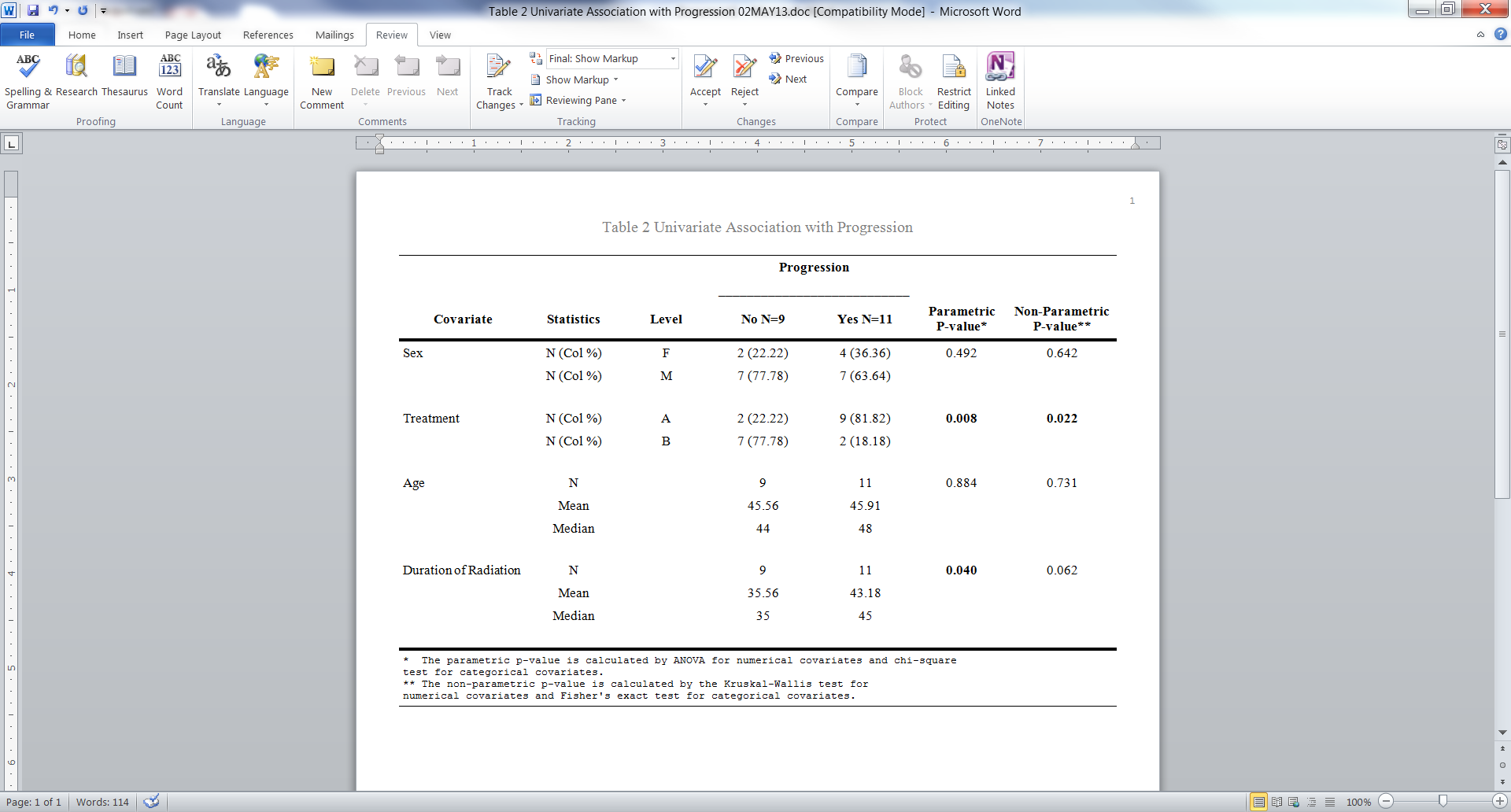
orientation = portrait,

outpath = C:\Documents and Settings\User\My Documents\,

fname = Table 2 Univariate Association with Progression);

TITLE;

**Summary Table Example:**



**Permission:**

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:   
  
The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.  
  
THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

**Log of Updates:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **By** | **Description** | **Version** |
| 6/19/12 | Dana Nickleach (dnickle@emory.edu) | Changed lengths from 100 to 96 to avoid a warning in proc report. Dropped missing values in proc transpose to avoid a warning. Got rid of leading space in the RTF file name. | V2 |
| 6/21/12 | Dana Nickleach (dnickle@emory.edu) | Added weight and debug parameters. Changed Proc Datasets so that the pre-existing files in the work library will not be deleted | V3 |
| 6/25/12 | Dana Nickleach (dnickle@emory.edu) | Changed length of Text1 – Text4 so that the notes would not be cutoff | V4 |
| 7/30/12 | Dana Nickleach (dnickle@emory.edu) | Initialized level variable to avoid warning. Reset options at the end of the macro to the options in use when the macro is called. This is mainly to return the orientation to the original setting when the macro is finished. | V5 |
| 8/7/12 | Dana Nickleach (dnickle@emory.edu) | Changed order of proc options so that the data set will be deleted in the end. Prevented analysis data set from being overwritten in PROC SORT. | V6 |
| 8/9/12 | Dana Nickleach (dnickle@emory.edu) | Increased length of covariate to 256, the maximum possible length of a variable label so that labels are not cut off in the table and added the SPANROWS option to accommodate longer labels. | V7 |
| 8/13/12 | Dana Nickleach (dnickle@emory.edu) | Added NONPAR=A option and also added a warning in the log if you are running NONPAR=F and the chi-square test is invalid. Version 9.3 is required for both of these. Changed the footnotes in NONPAR=A or T to be adjusted based on whether there are numerical and categorical covariates present. | V8 |
| 9/5/12 | Dana Nickleach (dnickle@emory.edu) | Increased the length of the outcome variable label so that characters won’t be cut off in the report. | V9 |
| 9/10/12 | Dana Nickleach (dnickle@emory.edu) | Made an adjustment so that a continuous and categorical variable with the same label can be displayed in the table. Dropped total rows from freq table to avoid missing value note. Changed parametric footnote to say Kruskal-Wallis instead of Wilcoxon. More than one outcome variables is now allowed. | V10 |
| 9/12/12 | Dana Nickleach (dnickle@emory.edu) | Fixed error when NLIST is empty. | V10 |
| 9/19/12 | Dana Nickleach (dnickle@emory.edu) | Made changes so that long variable names should not cause problems. This became an issue due to modifications in V10. | V11 |
| 10/10/12 | Dana Nickleach (dnickle@emory.edu) | Added spread parameter and fixed numeric conversion warnings. | V12 |
| 10/15/12 | Dana Nickleach (dnickle@emory.edu) | Added check for outcome variables that are too long. Suppressed plots to avoid warning when using weights. Prevented case sensitivity of T/F parameters. Adjusted footnotes in nonparametric case according to numeric or character variables. | V13 |
| 10/22/12 | Dana Nickleach (dnickle@emory.edu) | Fixed proc report ODS listing error and changed “group” to “order” to prevent note. | V14 |
| 11/12/12 | Sungjin Kim  (skim61@emory.edu) | Made a correction: ‘calcuated’ to ‘calculated’ in the footnote. | V15 |
| 11/15/12 | Dana Nickleach (dnickle@emory.edu) | Fixed bug preventing character values of “-“ from printing in the table. Changed order of min, max, and std in report to match the order in the %DESCRIPTIVES macro. | V16 |
| 1/3/13 | Dana Nickleach (dnickle@emory.edu) | Added BY parameter. | V17 |
| 1/11/13 | Dana Nickleach (dnickle@emory.edu) | Modified so that 0 will be printed instead of blanks when one value of the BY variables has an empty column. | V17 |
| 1/15/13 | Dana Nickleach (dnickle@emory.edu) | Fixed bug in last modification. | V17 |
| 1/16/13 | Dana Nickleach (dnickle@emory.edu) | Fixed scenario when a p-value cannot be calculated so that the row still displays and an NA is printed. | V17 |
| 1/17/13 | Dana Nickleach (dnickle@emory.edu) | Modified so that 0 and NA will be printed instead of blanks for the numerical summary when one value of the BY variable has an empty column. | V17 |
| 1/17/13 | Dana Nickleach (dnickle@emory.edu) | Added MHC parameter and condensed two proc reports into one for ease of coding. | V18 |
| 3/13/13 | Dana Nickleach (dnickle@emory.edu) | Increased variable label length so that it is not cut off for numerical variables. | V19 |
| 3/15/13 | Dana Nickleach (dnickle@emory.edu) | Declared macro variables as local so that they don’t interfere with other macros calling this macro, made modifications to suppress some of the output, and added DOC parameter. | V20 |
| 3/26/13 | Dana Nickleach (dnickle@emory.edu) | Fixed bug caused by missing values in the outcome variable and also a bug in formatted category labels. | V21 |
| 4/19/13 | Dana Nickleach (dnickle@emory.edu) | Added check to make sure outcomes are also not in CLIST as this will cause errors. | V22 |
| 4/22/13 | Dana Nickleach (dnickle@emory.edu) | Fixed problem causing numeric variables to appear in incorrect order in the report when they have the same name as categorical variables. | V23 |
| 4/29/13 | Dana Nickleach (dnickle@emory.edu) | Suppressed intermediate output. | V24 |
| 5/3/13 | Dana Nickleach (dnickle@emory.edu) | Fixed so that it works with NLIST, but without CLIST variables. | V25 |
| 5/6/13 | Dana Nickleach (dnickle@emory.edu) | Added check to make sure each variable in CLIST is at least 2 levels and that the BY variable is not also in CLIST, otherwise it will cause errors. | V26 |
| 9/19/13 | Dana Nickleach (dnickle@emory.edu) | Changed BY statement to a CLASS statement in PROC MEANS to avoid errors caused when the outcome variable has multiple values assigned to the same format. | V27 |
| 10/9/13 | Dana Nickleach (dnickle@emory.edu) | Fixed bug caused by commas in outcome variable values. | V28 |
| 5/27/14 | Dana Nickleach (dnickle@emory.edu) | Added MATCHID parameter. | V29 |
| 10/06/15 | Yuan Liu | Fix errors when using DOC=F | V30 |