UniLogistic: A SAS Macro for Descriptive and Univariable Logistic Regression Analyses

Navneet K. Dhand
The University of Sydney

Abstract

Descriptive and univariable logistic regression analyses are essential before constructing multivariable models, but are very time consuming, particularly if a large number of explanatory variables are to be evaluated. A macro UniLogistic is described in this paper that conducts descriptive and univariable logistic regression analyses (binomial, ordinal or nominal, as required) in SAS and presents results in formatted tables in Excel and graphics in PDF files. Implementation of the macro is illustrated in this paper using example datasets from statistics and epidemiology textbooks.

Keywords: binomial, ordinal, multinomial, cumulative logit, generalized logit, graphical summaries, summary statistics, SAS.

1. Introduction

Descriptive and univariable analyses are recommended before building multivariable logistic regression models because they provide information about distributions of variables and crude associations of explanatory variables with the outcome (Vittinghoff, Glidden, Shiboski, and McCulloch 2005). These analyses also enable identification of data entry errors, missing values and outliers, and thus aid in cleaning the dataset before building multivariable models. The importance of these analyses cannot be overemphasized, but they are often ignored or rushed through due to considerable time required in conducting them (Dohoo, Martin, and Stryhn 2004).

A SAS macro UniLogistic is described in this paper that conducts almost all the descriptive and univariable logistic regression analyses (binomial, ordinal or multinomial) required to make informed decisions before building multivariable models. The macro uses the SAS statistical program (SAS Institute Inc. 2003a) for conducting analyses but creates preformatted tables of results in Excel (Microsoft Corporation 2003) that can be easily adapted for journal
articles or research reports. Graphical summaries are saved in portable document files (PDF, Adobe Systems Incorporated 2010) in the directory specified by the user. All these results are produced very quickly, but more importantly, the code for implementing this macro is very straightforward and consists of just four arguments in its simplest form.

The macro works in the Microsoft Windows environment and requires SAS 9.1 or above (at least the SAS/BASE and the SAS/STAT components, SAS Institute Inc. 2003c), Acrobat reader (Adobe Systems Incorporated 2010) and Excel (Microsoft Corporation 2003) programs to function. Note that the PDF files are automatically created by the macro and the user does not need to have Acrobat or any other PDF creator installed on the computer.

In this paper, I will first describe the macro – its arguments, implementation and the output it produces – and then present examples demonstrating use of the macro for conducting binomial, ordinal and multinomial logistic regression analyses.

2. The macro arguments

UniLogistic macro is included with this article in the file UniLogistic.sas. The arguments taken by the macro are summarized in Tables 1 and 2.

Only the first four arguments are essential for conducting binomial logistic regression analyses (Table 1), i.e., the dataset name (DSN), the name of the outcome variable (OUTCOME), the names of the categorical (CATVAR) and continuous explanatory variables (CONTVAR). For conducting ordinal and multinomial logistic regression analyses, the user will need to specify an additional MODELTYPE argument that takes values ordinal and multinomial, respectively, for fitting cumulative and generalized logit models. All the remaining arguments are optional or have some default values.

XLFFILE and DIR arguments take the names of the Excel file and the directory, respectively,

| Arguments | Description | Notes |
|-----------|-------------|-------|
| DSN       | Name of the SAS dataset | Convert datasets in other formats to SAS format before implementing the macro. |
| OUTCOME   | Name of the outcome variable | Outcome can be binary, ordinal or multinomial. Note that an outcome in ‘event/trial’ format can also be specified. |
| CATVAR    | A list of all categorical explanatory variables | Do not include the outcome variable name or separate variable names by commas. |
| CONTVAR   | A list of all quantitative explanatory variables | Do not separate variable names by commas. |
| MODELTYPE | Type of the model to be fitted when there are > 2 categories of the outcome | The option ordinal fits cumulative logit whereas the option multinomial fits generalized logit model. This argument is not required for fitting binomial logistic regression models |

Table 1: Essential arguments for conducting descriptive and univariable binomial, ordinal or multinomial logistic regression analyses using UniLogistic macro.
Table 2: Optional arguments for conducting descriptive and univariable binomial, ordinal or multinomial logistic regression analyses using UniLogistic macro. Note that although the arguments XLFILE and DIR are optional, it is strongly recommended to specify values for these arguments to avoid accidentally overwriting any previously stored file with the same name.

| Arguments | Description | Notes |
|-----------|-------------|-------|
| XLFILE    | Name of the Excel file for storing results | UniLogistic Results is used as a default name if the file name is not specified. The same name is used as a prefix for PDF files. |
| DIR       | Name of the directory to save Excel and PDF files | Default is C:\Temp if no directory name is specified. |
| REF       | Reference level of class variables | The option first is default but can be changed to last. |
| ORDER     | Sorting order for categories of the outcome variable | The option descending (default) models the probability of the highest outcome category whereas ascending models the probability of the lowest outcome category. |
| STATISTIC | Correlation statistic for testing association between explanatory variables | 1: Spearman correlation (default); 2: Pearson correlation; 3: Kendall’s tau-b; 4: Stuart’s tau-c; 5: Lambda asymmetric C|R; 6: Lambda asymmetric R|C; 7: Lambda symmetric; 8: Uncertainty coefficient C|R; 9: Uncertainty coefficient R|C; 10: Uncertainty coefficient symmetric (Consult SAS documentation for the procedure FREQ for further details, SAS Institute Inc. 2003b). |
| CHISTATISTIC | $\chi^2$ statistic for testing association between categorical variables | 1: $\chi^2$ (default); 2: Likelihood-ratio $\chi^2$; 3: Mantel-Haenszel $\chi^2$. |
| OFFICE    | Version of Excel (Microsoft Corporation 2003) installed on the computer | Only required for Excel 2007. The default is 2003. |

3. Implementation and the output

The user calls UniLogistic macro by inputting the arguments discussed above (Tables 1 and 2). Before conducting the actual analyses, the macro performs some error checks to discover potential problems. If an error is detected (say, if the user has not specified the
dataset name), the macro will cancel execution and print a an error message in the SAS log. If all specifications are in order, the macro connects to Excel (Microsoft Corporation 2003) using the method of Roper (2000), creates and saves an Excel file with a desired number of sheets to store the results, and turns-off Excel error messages to ensure trouble-free access from SAS (SAS Institute Inc. 2003a).

After these preliminary actions, the macro conducts descriptive and univariable logistic regression analyses, saves the desired results in SAS datasets and then exports the saved results to Excel and PDF files. Before quitting, the macro deletes all the datasets created during the implementation to avoid cluttering, and to prevent errors in case the macro is implemented again.

First, we will discuss the procedure followed by the macro to conduct analyses and to create tables, and then the approach used to create and save graphics.

3.1. Conduct analyses in SAS and create tables in Excel

Create datasets

A nested macro UniBasic is defined within UniLogistic macro to conduct logistic regression analyses using the SAS Logistic procedure (SAS Institute Inc. 2003c). UniBasic macro is implemented for each categorical and quantitative variable individually by scanning the list of variables specified by the user. The results are saved in six SAS datasets for each variable using the Output Delivery System (ODS) and include the likelihood-ratio $\chi^2$ results, parameter estimates, odds ratios, goodness-of-fit results, contingency tables for categorical variables, summary statistics for quantitative variables, and the information for convergence. This nested macro also creates a dataset to store the number of missing observations for each variable using the SAS structured query language (SQL) procedure.

Datasets are also created for each pair of variables to evaluate associations between all possible combinations of explanatory variables using the FREQ procedure in SAS (SAS Institute Inc. 2003b). These datasets are used to create a table of associations between all explanatory variables to evaluate collinearity (Spearman rank correlation coefficient, $\chi^2$, $p$ value or other parameters as requested). Additionally, Pearson correlation coefficients are calculated for each combination of quantitative variables using the CORR procedure in SAS (SAS Institute Inc. 2003b).

Combine datasets

The datasets containing the likelihood-ratio $\chi^2$ results, the goodness-of-fit results and the convergence information for all explanatory variables are concatenated and then merged (using a set of SQL procedures) to create a combined dataset containing all these results. Similar combined datasets are created for (1) parameter estimates and odds ratio results, (2) contingency tables for categorical variables, (3) summary statistics for quantitative variables, (4) association results for all pairs of variables, and (5) correlation results for all pairs of quantitative variables. See the macro code available on the journal website for the details.

Export datasets to Excel

Selected results from the combined datasets are then exported to the Excel file using the
Dynamic Data Exchange (DDE) facility in SAS (Vyverman 2001, 2002). This facility enables customized export of data from SAS to a specified cell (or a range of cells) of an Excel file. The macro not only uses this facility to export data but also to export titles, headers and notes/messages for the user. To ensure that the results are exported to the desired location, the macro counts the number of observations in each combined dataset using the SQL procedure, prior to exporting results.

Format results in Excel

The results exported to Excel are then formatted to make them reader-friendly and to ensure that only a minimal effort is required to reformat the tables for publication. This involves changing the font for headers to Arial 12 size bold, applying borders at the top and the bottom of the table, rounding most numbers to two and the p values to three decimal places, and setting the columns width to fit contents. Although rounding of numbers can be accomplished directly in SAS, performing this operation in Excel allows the user to change the formatting at a later stage.

Customize analyses

For calculation of parameter estimates, by default, the macro makes the first ordered category of a categorical variable as a reference category. However, the last ordered category can be made as a reference category by using ref = last as an additional argument (Table 2).

UniLogistic macro models the probabilities for the highest outcome category, by default. This is same as explicitly specifying the option descending in the SAS PROC Logistic statement. For example, for a binary outcome variable in 1 (event) and 0 (non event) format, the macro models the probability of 1 (event). This also means that for an outcome variable with levels 1 and 2, the macro will model the probability of 2. For an ordinal outcome variable, the macro models the probability of a higher score and for a multinomial outcome variable, the macro makes the first category as a reference. However, the user can change these orders by specifying the order = ascending option in the macro implementation code (see Table 2). Similarly, by default, only the Spearman rank correlation coefficient estimate, $\chi^2$ and p value of the association between each pair are exported but these default settings can be changed using the STATISTIC and CHISTATISTIC options described in Table 2.

3.2. Graphical summaries

The macro creates four types of graphical summaries to aid in visual assessment of the associations and the distributions of the variables. A histogram for each quantitative variable and a box-and-whisker plot for each quantitative variable grouped by the outcome are created using UNIVARIATE and BOXPLOT procedures in SAS, respectively. Formatting is done using GOPTIONS statement in SAS and the graphics are saved in PDF files to the user specified directory using the SAS FILENAME statement.

In addition, two bar charts are created for each categorical variable using the SAS GCHART procedure, a simple bar chart describing the frequency distribution of the variable and a grouped bar chart by the categories of the outcome variable. Similar to above, these charts are saved in PDF files in the specified directory. To enable accessibility from just one location, hyperlinks for the PDF files are created in the Excel file in which the results are stored.
4. Examples

4.1. Binomial logistic regression

The dataset

Implementation of the UniLogistic macro is demonstrated using an example dataset NOCARDIA, described in detail in Dohoo et al. (2004). This dataset is from a case-control study conducted to investigate an outbreak of Nocardia mastitis (Ferns, Dohoo, and Donald 1991), in particular, to evaluate the association of various dry cow therapies with casecont, the case or control status of a herd (1/0).

We will assume that the data are in SAS format and are stored in the C:\myfolder directory. The dataset, UniLogistic macro and its implementation code are available as supplementary materials from the journal website.

Implementation

Specify the %INCLUDE statement to indicate the location of the macro file, input values for various arguments as shown in the code below (also see Tables 1 and 2) and invoke the macro. The macro will create a new Excel file, save it as Mastitis Results.xls in the directory C:\myfolder and create all the tables of results and graphs/plots.

%include 'C:\myfolder\unilogistic.sas';

%unilogistic (dsn = mastitis, outcome = casecont, catvar = dbarn dout dcprep dneo dclox doth, contvar = numcow prod bscd dcpct, xlfle = Mastitis Results, dir = C:\myfolder)

The output

Example output file for the analyses conducted using the NOCARDIA dataset (Dohoo et al. 2004) is available as a supplementary material on the journal website. Sheet 1 of the Excel file displays two tables: (1) contingency tables of categorical variables with the outcome; and (2) summary statistics for quantitative variables by the outcome (Figure 1 and 2). Note that the titles and headers of tables were created by the macro. The macro also applied the formatting including the borders, rounding of the summary statistics to one decimal place, and fitting the column widths to their contents.

Also note that the links are provided in this sheet to view graphical summaries. These links will only work if the user implements the macro on their own computer. However, the example graphics can be downloaded from the journal website for inspection.

Sheet 2 contains a table of likelihood-ratio \( \chi^2 \) statistics and \( p \) values for the unconditional associations of explanatory variables with the outcome (Figure 3). In addition, it contains two
Figure 1: Contingency tables created by UniLogistic macro based on the NOCARDIA dataset (Dohoo et al. 2004).

Figure 2: Summary statistics created by UniLogistic macro based on the NOCARDIA dataset (Dohoo et al. 2004).
Figure 3: A table of univariable logistic regression results created by UniLogistic macro to evaluate the unconditional association of explanatory variables with the outcome variable casecont based on the NOCARDIA dataset (Dohoo et al. 2004).

| Parameters | Likelihood ratio Chi-Square | DF | P-Value | Wald chi-Square (DF) | Model Converged? |
|------------|-----------------------------|----|---------|---------------------|-----------------|
| DOUT       | 5.22                        | 2  | 0.074   | 0.00 (G)            | No              |
| DCPREP     | 8.79                        | 3  | 0.002   | 0.00 (G)            | No              |
| DNEO       | 28.57                       | 1  | <0.001  | 1.00 (G)            | Yes             |
| DCLOX      | 9.11                        | 1  | 0.003   | 0.00 (G)            | Yes             |
| DOCTH      | 2.41                        | 1  | 0.121   | 0.00 (G)            | Yes             |
| NUMCOW     | 4.33                        | 1  | 0.035   | 66.65 (G)           | Yes             |
| PROD       | 4.29                        | 1  | 0.039   | 76.86 (G)           | Yes             |
| BSAC       | 0.47                        | 1  | 0.412   | 144.66 (101)        | Yes             |
| DCPC      | 11.57                       | 1  | 0.001   | 21.36 (17)          | Yes             |

Figure 4: Tables listing variables sorted by their unconditional p values and by their number of missing values, respectively.

lists of variables (Figure 4), one sorted by p values (to help select variables for multivariable analysis), and the other by the number of missing values (to verify if any variable has missing observations). Note that the table of likelihood-ratio $\chi^2$ results contains a column displaying information about model convergence. Also note that the variables for which the model did not converge were automatically omitted from the list of variables sorted by p values.

Sheet 3 contains a table of parameter estimates, standard errors, odds ratios and their 95% profile likelihood confidence intervals based on the unconditional associations of all explanatory variables with the outcome (Figure 5).

Sheet 4 contains the results of associations between all explanatory variables (Figure 6) and Pearson correlation coefficients between quantitative variables to help evaluate collinearity. See the example output file Mastitis Results.xls for details.

4.2. Ordinal and multinomial logistic regression

The procedures for conducting ordinal and multinomial logistic regression analyses for polytomous outcomes are similar to that of binomial logistic regression, and are therefore being discussed only briefly. There is only one additional argument MODELTYP which takes values ordinal and multinomial, respectively, for fitting a cumulative logit and generalized logit models. The generalized logit model is simply an extension of the binomial logistic model for a polytomous outcome variable and compares log odds for each category of the outcome
Figure 5: The table of parameter estimates and odds ratios created by the UniLogistic macro for the explanatory variables in the NOCARDIA dataset (Dohoo et al. 2004).

| A  | B   | C   | D    | E    | F    | G    | H    |
|----|-----|-----|------|------|------|------|------|
| 1  | Parameter Estimates and odds ratios for the unconditional association between explanatory variables and the outcome (CASECONT) |
| 2  | Parameters | Categories | Ref b | SE(b) | Odds ratios | LCL (OR) | UCL (OR) |
| 3  | INTERCEPT | 0.02 | 0.20 |
| 4  | DOUT | 2 | 1 | -14.22 | 7.00 | 70.70 | <0.001 | 0.90 |
| 5  | DOUT | 3 | 1 | 1.09 | 1.17 | 2.94 | 0.35 | 66.54 |
| 6  | INTERCEPT | 0.00 | 1.41 |
| 7  | DOCPREP | 2 | 1 | 0.41 | 1.68 | 1.50 | 0.04 | 57.03 |
| 8  | DOCPREP | 3 | 1 | 0.11 | 1.43 | 1.11 | 0.04 | 26.62 |
| 9  | DOCPREP | 4 | 1 | -14.20 | 456.60 | <0.001 | 1.46 |
| 10 | DOUT | 1 | 0 | -1.76 | 0.48 |
| 11 | DNEO | 1 | 0 | 2.43 | 0.64 | 11.37 | 4.21 | 36.69 |
| 12 | INTERCEPT | 0.27 | 0.22 |
| 13 | DOLOX | 1 | 0 | -1.14 | 0.88 | 0.32 | 0.12 | 0.75 |
| 14 | INTERCEPT | 0.34 | 0.28 |
| 15 | DOIT | 1 | 0 | -0.80 | 0.39 | 0.56 | 0.25 | 1.17 |
| 16 | INTERCEPT | -0.71 | 0.41 |
| 17 | NUMCOW | 0.01 | 0.01 | 1.02 | 1.00 | 1.03 |
| 18 | INTERCEPT | 2.64 | 1.32 |
| 19 | PROD | 0.11 | 0.06 | 1.12 | 1.01 | 1.25 |
| 20 | INTERCEPT | 0.27 | 0.41 |
| 21 | BSCC | 0.00 | 0.00 | 1.00 | 1.00 | 1.00 |
| 22 | INTERCEPT | -1.46 | 0.62 |
| 23 | DPC2T | 0.02 | 0.01 | 1.02 | 1.01 | 1.03 |

Figure 6: An excerpt of the table displaying results of association between all explanatory variables with each other in the NOCARDIA dataset (Dohoo et al. 2004).

| A  | B       | C       | D       | E       | F       |
|----|---------|---------|---------|---------|---------|
| 1  | Associations between the explanatory variables in the MASTITIS dataset |
| 2  | Parameters | Spearman Correlation | Asymptotic SE | Chi-Square | DF | P-value |
| 3  | TABLE DOUT * DOCPREP | 0.01 | 0.08 | 1.02 | 8 | 0.986 |
| 4  | TABLE DOUT * DNEO | 0.02 | 0.09 | 3.56 | 2 | 0.169 |
| 5  | TABLE DOUT * DOLOX | 0.02 | 0.10 | 4.11 | 2 | 0.126 |
| 6  | TABLE DOUT * DOIT | 0.01 | 0.10 | 1.21 | 2 | 0.254 |
| 7  | TABLE DOUT * NUMCOW | 0.29 | 0.07 | 71.00 | 88 | 0.006 |
| 8  | TABLE DOUT * PROD | -0.04 | 0.08 | 118.39 | 112 | 0.296 |
| 9  | TABLE DOUT * BSCC | -0.12 | 0.08 | 209.09 | 264 | 0.564 |
| 10 | TABLE DOUT * DOCP2T | -0.01 | 0.09 | 28.53 | 36 | 0.376 |
| 11 | TABLE DOCPREP * DNEO | -0.38 | 0.06 | 15.45 | 3 | 0.001 |
| 12 | TABLE DOCPREP * DOLOX | -0.11 | 0.08 | 2.82 | 3 | 0.420 |
| 13 | TABLE DOCPREP * DOIT | -0.02 | 0.11 | 7.43 | 3 | 0.060 |
| 14 | TABLE DOCPREP * NUMCOW | -0.17 | 0.08 | 123.31 | 132 | 0.548 |
| 15 | TABLE DOCPREP * PROD | -0.09 | 0.12 | 192.50 | 188 | 0.005 |
| 16 | TABLE DOCPREP * BSCC | -0.04 | 0.11 | 321.00 | 386 | 0.267 |
| 17 | TABLE DOCPREP * DOCP2T | -0.35 | 0.10 | 145.51 | 54 | 0.001 |

to the reference level. In contrast, the outcome is assumed to have a natural ordering in the cumulative logit model and the odds ratios are considered to be independent of the categories of the outcome variable.
UniLogistic: Descriptive and Univariable Logistic Regression Analyses in SAS

The dataset

We will use the low birth weight dataset (LOWBWT) for conducting ordinal and multinomial logistic regression analyses, assuming the outcome to be ordinal and nominal, respectively. This dataset is from a study conducted to investigate factors associated with low birth weight and is described in detail in Hosmer and Lemeshow (2000).

Implementation and the output

Implementation of the macro is similar to conducting binomial logistic regression analyses except the use of an additional MODELTYPE argument. The following code will evaluate unconditional association of six categorical variables (RACE, SMOKE, PTL, HT, UI, FTV) and one continuous variable (AGE) with the ordinal outcome (WTCAT) and save the results in an Excel file lbwt_results_ordinal in the directory C:\myfolder. See Tables 1 and 2 for further details.

The output produced will be similar to as discussed above for binomial logistic regression except the table of parameter estimates and odds ratios which will, of course, show three intercepts for each variable instead of just one in the binomial logistic regression (Figure 7). See the output file lbwt_results_ordinal available as a supplementary material on the journal website for further details.

%include 'C:\myfolder\unilogistic.sas';
%
unilogistic (dsn = weight1, outcome = wtcat, modeltype = ordinal, catvar = RACE SMOKE PTL HT UI FTV, contvar = age, xlf = lbwt_results_ordinal, dir = C:\myfolder)

Figure 7: An excerpt of the table displaying parameter estimates and odds ratios for unconditional association of variables in the LOWBWT dataset with the outcome WTCAT, assumed to be ordinal.

|   | A    | B  | C    | D     | E     | F     | G     | H     |
|---|------|----|------|-------|-------|-------|-------|-------|
| 1 | Parameter Estimates and odds ratios for the unconditional association between explanatory variables and the outcome (WTCAT) |
| 2 | Parameters | Categories | Ref b | SE(b) | Odds-ratios | LCL (OR) | UCL (OR) |
| 6 | INTERCEPT | 4 | -1.20 | 0.21 | .... | .... | .... |
| 6 | INTERCEPT | 3 | -0.34 | 0.20 | .... | .... | .... |
| 7 | INTERCEPT | 2 | 0.70  | 0.20 | .... | .... | .... |
| 8 | RACE    | 1 | 1.08  | 0.41 | 2.68  | 1.33 | 6.37 |
| 8 | RACE    | 3 | 0.64  | 0.29 | 1.91  | 1.06 | 3.33 |
| 8 | INTERCEPT | 4 | -1.12 | 0.20 | .... | .... | .... |
| 8 | INTERCEPT | 3 | -0.25 | 0.18 | .... | .... | .... |
| 8 | INTERCEPT | 2 | 0.87  | 0.19 | .... | .... | .... |
| 13| SMOKE   | 1 | 0.76  | 0.27 | 2.14  | 1.26 | 3.66 |
Figure 8: An excerpt of the table displaying parameter estimates and odds ratios for unconditional association of variables in the LOWBWT dataset with the outcome WTCAT, assumed to be nominal.

The code for building generalized logit models is the same as for ordinal logistic models except that the user needs to specify `multinomial` for the `MODELTYPE` argument:

```
modeltype = multinomial,
```

The output produced by the macro will also be similar to the one produced for building ordinal models (of course, the estimates will be different for different model types) except the table of parameter estimates and odds ratios, which will contain another column of the levels of outcome variable (Figure 8).

The dataset, the implementation codes and the detailed outputs produced by the macro for both ordinal and multinomial logistic regression analyses are available from the journal website.

### 5. Discussion

Significant time is consumed in properly conducting descriptive and univariable logistic regression analyses. This results in a tendency by some analysts to jump straight to multivariable analyses which invariably results in erroneous conclusions (Dohoo et al. 2004). The macro described in this paper saves time for conducting descriptive and univariable logistic regression analyses, and therefore, encourages analysts to follow a systematic approach to data analyses. This will enable better conclusions from studies, particularly with a large number of explanatory variables, a common scenario in epidemiologic research (Bagley, White, and Golomb 2001).

Time is required not only for conducting analyses, but also for manually creating tables of results for preparing journal articles or research reports based on the output from a statistical program. UniLogistic macro automatically prepares tables of results that can be easily adapted for various purposes, further saving researchers’ precious time.
UniLogistic: Descriptive and Univariable Logistic Regression Analyses in SAS

From the very outset, the aim was to create a program that is easier to implement by users because many of the potential users were likely to be graduate students. Therefore, in its simplest form, the user only need to specify four arguments to implement the macro, and all the analyses are conducted, results compiled, tables and graphs created in less than a minute on most personal computers. This user-friendly nature of this macro encouraged not only graduate but also many undergraduate students to use the macro for analyzing their research project data at our institute.

Of course, like any program, this macro has limitations. First, it does not have as many options as are available in a standard statistical package. Second, although peer reviewers have tested the macro and every care has been taken to remove any potential errors, it has not been probably tested under all possible conditions. To safeguard the user against erroneous results, however, the macro does not delete any major SAS output, so that the user can inspect the output in addition to inspecting the tabulated results in the Excel file. Third, the user must have basic knowledge and understanding of SAS (SAS Institute Inc. 2003a) to implement the macro. Finally, the macro does not perform multivariable analyses but gives a user almost all the information to start building multivariable models.

In addition to these limitations, some precautions are necessary during implementation. As the macro creates and saves Excel file itself for exporting results, the user must ensure that Excel program is closed before invoking the macro, and that any other Excel file of the same name as specified in the macro code must not be present in the nominated directory. Secondly, to function properly, the macro turns off Excel error messages. Therefore, the user must save any changes made in the Excel file after its creation because Excel will not prompt the user to save these changes. However, if the Excel program is closed and reopened, the original settings are restored, i.e., Excel will start prompting the user to save any changes made in the file. Thirdly, the hyperlinks created in the Excel file to access PDF files refer to the nominated directory and the file name. Therefore, the links will not work if the user tries to open the files via the hyperlinks on some other computer where the analyses were not conducted. In that case the user will need to physically open the files from the directory where the files are stored.

Other than these limitations and implementation issues, the macro functions smoothly and can save a lot of time for users besides encouraging them to follow a systematic approach to conduct analyses.

Acknowledgments

The author is grateful to Associate Professor Peter Thomson and Dr. Naomi Cogger for reviewing the macro and its implementation and providing constructive feedback. Thanks are also due to The SAS Users List (http://www.listserv.uga.edu/) for promptly responding to the author’s queries. The financial assistance provided by the Teaching Improvement and Equipment Scheme (TIES) of the University of Sydney for conducting the review is thankfully acknowledged.
References

Adobe Systems Incorporated (2010). Adobe Acrobat Reader 9, Version 9.3.2. URL http://www.adobe.com/.

Bagley SC, White H, Golomb BA (2001). “Logistic Regression in the Medical Literature: Standards for Use and Reporting, with Particular Attention to One Medical Domain.” Journal of Clinical Epidemiology, 54, 979–985.

Dohoo I, Martin W, Stryhn H (2004). Veterinary Epidemiologic Research. AVC Inc., Charlottetown, Prince Edward Island, Canada.

Ferns L, Dohoo I, Donald A (1991). “A Case-Control Study of Nocardia Mastitis in Nova Scotia Dairy Herds.” Canadian Veterinary Journal, 32, 673–677.

Hosmer DW, Lemeshow S (2000). Applied Logistic Regression. 2nd edition. John Wiley & Sons, New York.

Microsoft Corporation (2003). Microsoft Office Excel 2003. URL http://office.microsoft.com/excel.

Roper C (2000). “Intelligently Launching Microsoft Excel from SAS, Using SCL Functions Ported to Base SAS.” In Proceedings of the 25th Annual SAS Users Group International Conference, Paper 25, 97.

SAS Institute Inc (2003a). The SAS System, Version 9.1. Cary, NC. URL http://www.sas.com/.

SAS Institute Inc (2003b). SAS/BASE Software, Version 9.1. Cary, NC. URL http://www.sas.com/.

SAS Institute Inc (2003c). SAS/STAT Software, Version 9.1. Cary, NC. URL http://www.sas.com/.

Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE (2005). Regression Methods in Biostatistics. Springer-Verlag, New York.

Vyverman K (2001). “Using Dynamic Data Exchange to Export Your SAS Data to MS Excel – Against All ODS, Part I.” In Proceedings of the Twenty-Sixth Annual SAS Users Group International Conference, 26.

Vyverman K (2002). “Creating Custom Excel Workbooks from Base SAS with Dynamic Data Exchange: A Complete Walkthrough.” In Proceedings of the 27th Annual SAS Users Group International Conference, 27.
A. Supplementary files

The following supplementary files are available on the journal website:

A.1. The macro

- **UniLogistic.sas**: The macro file required for conducting all the analyses described in this paper. Please refer to [http://sydney.edu.au/vetscience/biostat/macros](http://sydney.edu.au/vetscience/biostat/macros) for any updated versions of the macro.

A.2. Binomial logistic regression

- **Implementation Code (BLR).sas**: Example code to conduct descriptive and univariable binomial logistic regression analyses for the outcome `casecont` based on the NOCARDIA dataset.
- **nocardia.xls**: The dataset used for demonstrating binomial logistic regression analyses in this paper.
- **Mastitis Results.xls**: The example output file created by the macro containing binomial logistic regression results for the NOCARDIA dataset.
- **Mastitis Results Bar Charts descriptive.pdf**: Bar charts produced by the macro for categorical variables in the NOCARDIA dataset.
- **Mastitis Results Bar charts with the outcome.pdf**: Grouped bar charts produced by the macro to evaluate association of categorical explanatory variables with the outcome (based on the NOCARDIA dataset).
- **Mastitis Results Histograms descriptive.pdf**: Histograms of continuous variables in the NOCARDIA dataset created by the macro.
- **Mastitis Results Box plots with the outcome.pdf**: Box-and-whisker plots of quantitative variables by the outcome created by the macro.

A.3. Ordinal and multinomial logistic regression

- **lowbwt.xls**: The dataset used for conducting ordinal and multinomial logistic regression analyses in this paper.
- **Implementation Code (OLR).sas**: Example code to conduct descriptive and univariable ordinal logistic regression analyses for the outcome WTCAT (created from the BWT variable in the dataset LBWT).
- **lbwt_results_ordinal.xls**: The example output file created by the macro containing ordinal logistic regression results for the outcome WTCAT in the LOWBWT dataset.
• **Implementation Code (MLR).sas**: Example code to conduct descriptive and univariate multinomial logistic regression analyses for the outcome WTCAT (assuming the outcome to be nominal) in the LOWBWT dataset.

• **lbwt_results_multinomial.xls**: The example output file created by the macro containing multinomial logistic regression results for the outcome WTCAT in the LOWBWT dataset (assuming the outcome to be nominal).

• **lbwt_results_ordinal Bar Charts descriptive.pdf**: Bar charts produced by the macro for categorical variables in the LOWBWT dataset.

• **lbwt_results_ordinal Bar charts with the outcome.pdf**: Grouped bar charts produced by the macro for the categorical explanatory variables in the LOWBWT dataset with the outcome WTCAT.

• **lbwt_results_ordinal Histograms descriptive.pdf**: Histograms of continuous variables in the LOWBWT dataset created by the macro.

• **lbwt_results_ordinal Box plots with the outcome.pdf**: Box-and-whisker plots of quantitative variables by the outcome created by the macro.

**Affiliation:**

Navneet K. Dhand  
Faculty of Veterinary Science  
The University of Sydney  
425, Werombi Road  
Camden 2570 NSW, Australia  
Telephone: +61/2/9351-1669  
Fax: +61/2/9351-1693  
E-mail: navneet.dhand@sydney.edu.au  
URL: [http://sydney.edu.au/vetscience/biostat/macros/](http://sydney.edu.au/vetscience/biostat/macros/)