A Simple SAS Macro Technique to Build a Report Data Set
Christianna S. Williams, Yale University, New Haven, CT

ABSTRACT
Often one needs to reproduce the same statistical analysis for many samples or for many variables. However, frequently the desired report requires only a small portion of the standard output for a given set of analyses. Manually extracting the needed information to produce a presentable report is tedious and error prone. Further, this effort must be duplicated if the input data changes. With the Output Delivery System (ODS) virtually all procedure output is now available in the form of SAS® data sets and is thus amenable to further manipulation. This brief paper presents a fairly general method to extract specified information from these output data sets and use it to build a data set that can then be presented using one of the SAS reporting procedures. A simple illustration using statistical output from PROC FREQ and reported using PROC PRINT will be described.

INTRODUCTION
One of the wonderful things about SAS® is the enormous array of statistical methods that it offers. One of the not so wonderful things has been the enormous quantity of output that each procedure provides by default. Once the conscientious analyst has reviewed this output to make sure that the analysis is appropriate, only selected portions may be of interest to report to the client. Therefore, it would be useful to automate the extraction of the desired information to produce a report. This is particularly true for repetitive analyses or cases when the input data is likely to change. SAS Macro and ODS to the rescue! SAS Macro allows one to conduct a specific analysis repeatedly, with one or more parameters changing each time. ODS allows one to select desired portions of the output from each repetition of the analysis. DATA STEP code can then be used to manipulate these pieces into the desired format for reporting, and the selected information from each analysis can be appended to a growing “Report Data Set”. When the analyses are complete, the Report is written. This basic technique can be adapted to generate concise, understandable reports for many different types of statistical analyses. In my experience the time required to develop a program using this technique is more than compensated by the time saved by avoiding more laborious methods of producing presentable reports.

STRATEGY
Below are the basic steps in implementing this technique. A concrete example and the resulting output follow:
1) Develop the SAS code (one or more PROCs) that will run the desired analysis for one repetition.
2) Determine the output objects that are needed for inclusion in the desired report, and write the ODS statements to obtain these elements in one or more SAS Data Sets. In my experience, the best way to do this is the following:
   a) Determine the names of the output objects that are produced by the code developed in Step 1 by invoking “ODS TRACE ON”.
   b) Use the ODS OUTPUT statement to generate SAS data sets from the objects containing the desired data.
   c) Use PROC PRINT and/or CONTENTS to examine these data sets to determine the names and formats of variables containing the information of interest.
3) If information is needed from more than one output data set and/or some additional manipulation is needed to format the information as desired, write DATA STEP code to achieve this. Make sure to maintain sufficient information to distinguish each repetition of the analyses.
4) Modify the code developed in Steps 1 to 3 above, converting references that will change from one repetition to the next to macro variables, and enclose the code in a macro.
5) Add a step within the macro that will add the new information from each repetition of the analysis to the Report Data Set.
6) Call the macro once for each repetition of the analysis.
7) When finished, use one of the SAS reporting procedures (e.g. PROC PRINT) or a DATA _NULL_ step to write the report, which now contains only the desired portions of the statistical output for all repetitions of the analysis.

EXAMPLE
In the data used for the example below, we were examining factors that may be associated with the presence of chronic dizziness in older adults, and in the early stages of the analysis I wished to cross-tabulate a large number of potential predictor variables with the presence of dizziness. I wanted to concisely report the number and percentage with and without dizziness for each value of the predictor variables and the overall chi-square statistic and p-value for the associations.

The following PROC FREQ code will do the desired analysis for a single independent variable, in this case use of beta blockers to treat hypertension. Both variables have only two levels, with a 1 meaning the characteristic is present and a 2 that it is absent.

```
PROC FREQ DATA = dizzy;
   TABLES betablok*chrondiz /NOPERCENT NOCOL
   CHISQ;
```

Part of the output produced is shown below.

<table>
<thead>
<tr>
<th>BETABLOK (beta-adrenergic blockers)</th>
<th>CHRONDIZ (Chronic Dizziness)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td></td>
</tr>
<tr>
<td>Row Pct</td>
<td>1</td>
</tr>
<tr>
<td>---------</td>
<td>---</td>
</tr>
<tr>
<td>1</td>
<td>59</td>
</tr>
<tr>
<td>1</td>
<td>31.22</td>
</tr>
<tr>
<td>2</td>
<td>202</td>
</tr>
<tr>
<td>2</td>
<td>22.49</td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>Total</td>
<td>261</td>
</tr>
</tbody>
</table>

Statistics for Table of BETABLOK by CHRONDIZ

<table>
<thead>
<tr>
<th>Statistic</th>
<th>DF</th>
<th>Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>1</td>
<td>5.512</td>
<td>0.011</td>
</tr>
<tr>
<td>Likelihood Ratio Chi-Square</td>
<td>1</td>
<td>5.214</td>
<td>0.013</td>
</tr>
<tr>
<td>Continuity Adj. Chi-Square</td>
<td>1</td>
<td>5.401</td>
<td>0.014</td>
</tr>
<tr>
<td>Mantel-Haenszel Chi-Square</td>
<td>1</td>
<td>5.505</td>
<td>0.011</td>
</tr>
<tr>
<td>Phi Coefficient</td>
<td></td>
<td>0.0774</td>
<td></td>
</tr>
</tbody>
</table>

By adding the statement

```
ODS TRACE ON;
```
and re-running, the information (edited to save space) shown below is written to the log.

Output Added:
Name: CrossTabFreqs
Label: Cross-Tabular Freq Table

Output Added:
Name: ChiSq
Label: Chi-Square Tests

Output Added:
Name: Fisher'sExact
Label: Fisher's Exact Test

The next step is to insert the following ODS statement before the PROC FREQ code:

```
ODS OUTPUT CrossTabFreqs=xtab ChiSq=chisq;
```

A PRINT of these data sets is shown in Figure 1. All the desired information is contained in four observations of the XTAB data set.
and a single observation of the CHISQ data set. Convert the four observations in XTAB to two (one for each level of BETABLOK) and merge this with the desired row from the CHISQ data on the value of Table. To allow these observations to be readily concatenated with those for other predictors, rename the BETABLOK variable to something generic, without losing track of the fact that these statistics are for this particular predictor. This is very straightforward within a macro. Optionally assign a format that will label the values of the predictor in the Report Data Set, which format will be passed to the macro as a parameter along with the name of the predictor variable. (Note that the code defining the formats would be outside the macro). Append the information for each repetition to the Report data set.

The macro written for this example, a few calls to it, the final PROC PRINT, and its output are shown below:

```sas
%MACRO tabit(indep=, fmt=)
    ODSCROSS TABLES &indep*chrondiz/nopercent nocol;
    DATA xtab2 KEEP= &indep predvar predvalf freqyes pctyes freqno pctno;
    SET xtab KEEP= &indep predvar predvalf frequency rowpercent table
    WHERE=(chrondiz ne . AND &indep ne .)
    BY &indep;
    LENGTH predvar $12;
    predvar = &indep;
    predvalf = put(&indepvar,fmt.);
    RETAIN freqyes freqno pctyes pctno;
    IF chrondiz = 1 THEN DO;
        freqyes=frequency; pctyes=rowpercent; END;
    ELSE IF chrondiz = 2 THEN DO;
        freqno=frequency; pctno=rowpercent; END;
    IF LAST.&indepvar THEN OUTPUT;
RUN;
DATA addit DROP=statistic table;
MERGE xtab2 chisq(WHERE=(statistic='Chi-Square'))
    BY table;
    IF NOT FIRST.table THEN DO;
        value = .; df = .; prob = .; END;
    RUN;
PROC APPEND BASE=Report DATA=addit; RUN;
%MEND tabit;
%tabit(indepvar=betablok, fmt=no2f.);
%tabit(indepvar=alc3, fmt=alc3f.);
%tabit(indepvar=dep_nm, fmt=no2f.);
... etc...
PROC FORMAT;
VALUE $varf 'betablok'='On B-blockers' ...
PROC PRINT DATA=report;
BY predvar notsorted; ID predvar;
VAR predvalf freqyes pctyes freqno pctno
    chi_square df prob;
LABEL predvar = 'Characteristic' ...
FORMAT pctyes pctno 6.2 predvar $varf.
    prob chi_square 7.4;
RUN;
```

CONCLUSIONS
Even before the advent of ODS, I found this general method of building a report data set to be widely applicable for repetitive statistical analyses including various regression methods and non-parametric analyses. Now that every bit of output is accessible in SAS data sets, it is a still more powerful tool.

CONTACT INFORMATION
Your comments and questions are welcome. Contact the author at Christianna.Williams@yale.edu.

SAS is a registered trademark or trademark of SAS Institute Inc. in the USA and other countries. indicates USA registration.

**FIGURE 1.**

<table>
<thead>
<tr>
<th>Table</th>
<th>BETABLOK <em>CHRONDIZ</em> <em>TYPE</em> <em>TABLE</em></th>
<th>Frequency</th>
<th>Row Percent</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>BETABLOK_by_CHRONDIZ</td>
<td>1 1 11 1</td>
<td>59</td>
<td>31.2169</td>
<td>.</td>
</tr>
<tr>
<td>BETABLOK_by_CHRONDIZ</td>
<td>1 2 11 1</td>
<td>130</td>
<td>68.7831</td>
<td>.</td>
</tr>
<tr>
<td>BETABLOK_by_CHRONDIZ</td>
<td>1 . 10 1</td>
<td>189</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>BETABLOK_by_CHRONDIZ</td>
<td>2 1 11 1</td>
<td>202</td>
<td>22.4944</td>
<td>.</td>
</tr>
<tr>
<td>BETABLOK_by_CHRONDIZ</td>
<td>2 2 11 1</td>
<td>696</td>
<td>77.5056</td>
<td>.</td>
</tr>
<tr>
<td>BETABLOK_by_CHRONDIZ</td>
<td>2 . 10 1</td>
<td>898</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>BETABLOK_by_CHRONDIZ</td>
<td>1 . 01 1</td>
<td>261</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>BETABLOK_by_CHRONDIZ</td>
<td>. 2 01 1</td>
<td>826</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>BETABLOK_by_CHRONDIZ</td>
<td>. . 00 1</td>
<td>1087</td>
<td>.</td>
<td>0</td>
</tr>
</tbody>
</table>

**FIGURE 2.**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
<th>dizzy</th>
<th>dizzy</th>
<th>dizzy</th>
<th>dizzy</th>
<th>square</th>
<th>DF</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>On B-blockers</td>
<td>Yes</td>
<td>59</td>
<td>31.22</td>
<td>130</td>
<td>68.78</td>
<td>6.5107</td>
<td>1</td>
<td>0.0107</td>
</tr>
<tr>
<td>No</td>
<td>202</td>
<td>22.49</td>
<td>696</td>
<td>77.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol use</td>
<td>None</td>
<td>172</td>
<td>26.92</td>
<td>467</td>
<td>73.08</td>
<td>7.7903</td>
<td>2</td>
<td>0.0203</td>
</tr>
<tr>
<td>1-10 oz./month</td>
<td>60</td>
<td>20.00</td>
<td>240</td>
<td>80.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;10 oz./month</td>
<td>20</td>
<td>18.02</td>
<td>91</td>
<td>81.98</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depressive symptoms</td>
<td>Yes</td>
<td>91</td>
<td>41.74</td>
<td>127</td>
<td>58.26</td>
<td>46.5356</td>
<td>1</td>
<td>0.0000</td>
</tr>
<tr>
<td>No</td>
<td>140</td>
<td>19.15</td>
<td>591</td>
<td>80.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>