Passing Statistical Results from SAS® Procedures to Word Documents

Kem Phillips, Pfizer Inc., New York, NY

Abstract

Results from SAS procedures can be passed to Word documents by assigning a unique name to each result and using Word’s macro language MacroBasic to transfer the names and results to the document. The names are constructed from variable names and statistical notations by processing special command files into SAS programs. The results of some procedures are obtained from the Output Delivery System. Table formatting is handled in Word.

Introduction

Manuscripts and tables are often revised because data have been corrected, or because better table formats have been developed. Ideally, changes in data would be automatically reflected in manuscripts, and restructuring tables would not require reprogramming or recomputation of results. This paper describes a way to achieve part of this goal.

There are three essential aspects of this system. First, each value from a SAS procedure is given a unique name, and a SAS file of the names and values is created. Second, the names are put into the manuscript where the numbers are to appear. Third, the numbers are transferred.

To name the results, a set of “generating” programs converts command files into SAS programs. These programs invoke SAS procedures, construct names for each statistical result and merge them with a master file. Microsoft Word’s macro language WordBasic is used to transfer names and values to the Word manuscript.

With this system, table formatting and numbering are handled in Word, and do not require SAS programming. Changes in the statistical results are transferred to Word almost automatically, and table layouts are changed in Word. In addition, Word features such as special formatting and automatic table numbering, listing and referencing can be exploited.

Naming Statistical Results

Statistics are numbers calculated from data. Means, frequency counts and p-values are typical statistics. Each is determined by features such as the variable analyzed, the statistical construct (mean, standard deviation, p-value) and the statistical model. Concatenating variable names and SAS notations provides a meaningful name for any statistic. For example, in a clinical trial, the mean age for females treated with placebo might be called

\[ \text{MEAN\_AGE\_SEX=FEMALE\_DRUG=PLACEBO} \]

and the p-value obtained from a Chi-square test of the table TREAT\_SEX could be

\[ \text{PVAL\_TREAT\_SEX\_CHISQ} \]

Calculating Named Statistical Results: Using the Output Delivery System

Once statistics have been named, they must be attached to the numerical results. With the Output Delivery System (ODS) (SAS/STAT® Software: Changes and Enhancements, Release 6.07), all of the results of a procedure can be put into SAS data sets. For example, in a clinical trial of a drug used to treat depression, PROC MIXED is used for an analysis of variance of a depression score (DEPRESS) by sex with the model

\[ \text{MODEL DEPRESS = DRUG SITE DRUG*SITE;} \]

The analysis of variance results are put into a temporary SAS file, AVTABLE, with the statement

\[ \text{MAKE 'TESTS' OUT=AVTABLE;} \]

AVTABLE might look like Table 1. Table 2 contains names for the 24 numbers in Table 1.

Table 2: Names Assigned to Values in Table 1

<table>
<thead>
<tr>
<th></th>
<th>DDF_DEPRESS_DRUG_MAINMOD SEX=FEMALE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DDF_DEPRESS_DRUG_MAINMOD SEX=FEMALE</td>
</tr>
<tr>
<td>2</td>
<td>DDF_DEPRESS_DRUG_MAINMOD SEX=FEMALE</td>
</tr>
<tr>
<td>24</td>
<td>P_F_DEPRESS_DRUG*SITE_MAINMOD SEX=MALE</td>
</tr>
</tbody>
</table>

At present, ODS is available only with PROC MIXED and PROC GENMOD, but values can be put into SAS files using other OUTPUT data sets, as well as PROC PRINTTO. However, the uniformity of ODS files simplifies and integrates the system.

Table 1: Results of PROC MIXED MAKE Statement

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>DDF</th>
<th>NDF</th>
<th>F</th>
<th>P_F</th>
<th>SEX</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRUG</td>
<td>60</td>
<td>1</td>
<td>16.60165</td>
<td>0.00101</td>
<td>FEMALE</td>
</tr>
<tr>
<td>SITE</td>
<td>60</td>
<td>3</td>
<td>0.7784261</td>
<td>0.51015</td>
<td>FEMALE</td>
</tr>
<tr>
<td>DRUG*SITE</td>
<td>60</td>
<td>3</td>
<td>0.505431</td>
<td>0.69044</td>
<td>FEMALE</td>
</tr>
<tr>
<td>DRUG</td>
<td>80</td>
<td>1</td>
<td>4.321293</td>
<td>0.04123</td>
<td>MALE</td>
</tr>
<tr>
<td>SITE</td>
<td>80</td>
<td>3</td>
<td>2.331677</td>
<td>0.08144</td>
<td>MALE</td>
</tr>
<tr>
<td>DRUG*SITE</td>
<td>80</td>
<td>3</td>
<td>0.535419</td>
<td>0.88055</td>
<td>MALE</td>
</tr>
</tbody>
</table>
Programs for Generating Names and Values

The purpose of this system is to simplify writing documents; it would be defeated if complicated programs to name the results had to be written each time. However, names can be generated from ODS files by SAS programs constructed from special command files. For example, the file in Table 3 can be processed into a SAS program that runs PROC MIXED using the model in the file EFFICACY.GLM, adds a MAKE command to invoke ODS, creates a SAS file containing the names and values for the TEST statistics, and merges the results with a permanent file.

Table 3: Example of a Command File

| LIBNAME | CLNTRIAL C:\SAS\DEPRESS | (File of named statistical results) |
| STTFILE | CLNTRIAL.STTFILE | |
| DATAFILE | CLNTRIAL.DEPRSSDT | (Input data set) |
| SELECT | EVALUABL = 'YES' | (Data subset to be analyzed) |
| BY | SEX | |
| VAR | DEPRESS | (Variable to be analyzed) |
| TEST | F NDF DDF P_F | (Statistics to be saved) |
| MODEL | MAINMOD | (Model name for this analysis) |
| MIXED | C:\SAS\DEPRESS\EFFICACY.GLM | (File containing PROC MIXED program) |

Getting Statistics Names into Word

Statistical results from SAS are to be linked to Word through the statistics names. These names may be very long, and, in addition, the names must not be visible in the document, so that Word's hidden font must be used. Typing these long names into a document and selecting the hidden font would be far more work than simply typing the numbers, even several times. However, names are generated by the programs described in the last section, and WordBasic (Microsoft Word Developer's Kit) can transfer names to a document.

Using WordBasic, two methods for transferring names were developed, a "point-and-click" method and a numbering scheme. For point-and-click, the user clicks on the position in the manuscript, then moves to the name list and then hits a macro button. A first version of this macro proved clumsy. In that version the document and the list of names were displayed simultaneously on the screen, but only a large, high-resolution screen could display enough of both documents to make the process efficient. A more successful version is activated by positioning the document cursor, clicking on a button to change the display to the list of names, scrolling through to find the proper name and clicking the button again to return to the manuscript and transfer the name.

A second method was developed where numbers are associated with the statistics names, and these numbers typed into the document from a printed list using a convention to set the numbers off from the document text. For example, in the list above, (24) would refer to

\[ \text{P_F DEPRESS DRUG SITE MAINMOD SEX=MALE} \]

These numbers are typed into the manuscript and then a WordBasic macro substitutes the names. The process is facilitated by building small lists of names for each table.

These macros insert into the Word document expressions like

\[ \text{<<F\_DEPRESS\_DRUG\_MAINMOD\_SEX=MALE} \]

\[ (=\text{round}(2) \% \#0.0) \]

The brackets, <<, mark the expression for later substitution of the value of the statistics, which will later be placed after round(). The expression \( (=\text{round}(2) \% \#0.0) \) is a Word field, rather than a text string. It rounds the number to 2 places and formats it to show 2 decimal places. Ultimately, only the rounded value will be visible; the field and statistic name are made invisible using the "hidden" font. The number of decimal places is determined in the generating programs but can be changed in the document itself.

Getting Values into Word

Once names are in a Word document, another WordBasic macro reads the manuscript searching for "<<", determines the name, finds its match in the data file, and transfers the statistic value to the \( (=\text{round}) \) field. Thus the value is updated without removing the statistic name. The result might be

\[ \text{<<F\_DEPRESS\_DRUG\_MAINMOD\_SEX=MALE} \]

\[ (=\text{round}(4.321293,2) \% \#0.0) \]

Since the field specifies 2 decimal places, only 4.32 will show in the document.

Using Dynamic Data Exchange

This system now uses intermediate ASCII files between SAS and Word. However, Bott (Bott, 1995) discusses linking SAS and Word using Dynamic Data Exchange (DDE). Using DDE would speed the transfer of names and values, and would eliminate unnecessary files.

A Model for Report Generation

All SAS System-using organizations produce reports containing the results of SAS runs. These organizations can be classified according to the degree of order and efficiency with which reports
are written. The extremes are total order and total chaos. When
total order prevails, reports are programmed once and run
automatically. With total chaos, anything can change at any time
and there is no chance of planning. In either case, the system
described in this paper provides no benefits. However, when tables
can be planned, but with some changes in formats and data
expected, this system can increase productivity. Here is a model of
this process, framed in the context of a clinical trial run by a
pharmaceutical company.

Since a clinical trial proceeds according to a protocol within a clinical
program, there are usually a number of standard tables. Other
tables are not entirely set since they depend on the characteristics of
the data. In addition, amendments to the data may result from late
reporting or detecting errors in the course of analysis.

In any case, a preliminary set of tables is drafted and set up in
Word. It defines the set of statistics to be calculated. Preliminary
data analyses may suggest changes to the tables. When the
analyses are firmly set, command files are constructed, debugged
and run to produce a file of statistical results. The statistics names
are then moved to the Word tables. The results are transferred in an
nearly automatic last step. The manuscript text may also include
statistics names, so that any data changes are reflected in the
manuscript. Note that the Word tables serve two purposes,
planning and execution.

Now, if the data are changed, the existing programs are rerun and
the results substituted. Rearranging tables is done in Word, and
revising table numbers and table lists is done automatically. If new
statistical results are needed, the command files are revised and
processed, the new names transferred, and the values substituted.

Conclusion

The efficiency of the system depends on how fast and accurately the
statistics names can be transferred to the Word manuscript. A
large, high-resolution screen would make searching for names
easier, but the process is accelerated by simply editing existing
tables and creating subsets of names for individual tables so that a
smaller data base is searched.

Moving numbers into the document is very slow in the existing
system since two documents must be repeatedly activated, but this
is essentially a "batch" process which can run independently, and
can probably be improved by technical means such as DDE. Other
enhancements can be added, such as extensive error checking of
the command files, possibly using a data dictionary.

Not all tables are easily created using this system. It requires that
tables be formatted in advance, but the categories defining a table
may not be known. For example, a table of counts of various
adverse events in a clinical trial will depend on which events occur.

This prototype system demonstrates results from SAS procedures
can be efficiently linked to Word documents.

References

Software: Changes and Enhancements, Release 6.07, Cary, NC:

Redmond, WA.

3. M. Brit. “Using Dynamic Data Exchange (DDE) to Pass Data to
Microsoft Word Documents from within the SAS System”. To be
presented at the 1996 SUGI Conference, Chicago, IL.

SAS and SAS/STAT are registered trademarks of SAS Institute Inc.
in the USA and other countries. Microsoft is a registered trademark
of Microsoft Corporation. © indicates USA registration.

Author's address: Pfizer, Inc., Dept of Clinical and Scientific Affairs
2194, 235 E. 42nd St., New York, NY 10017. Email:
kphilip@chelsea.ios.com