

Quickly Organize Statistical Output for Review

Bradford J. Danner, inVentiv Health Clinical, Tennessee

ABSTRACT

Statisticians and programmers in the pharmaceutical industry are often required to include model based tables to satisfy the requirements of study protocols. Working for a client who does not explicitly require or expect the output created from SAS procedures as part of a submission has understandably created difficulty when reviewing such tables for accuracy. The programming team tends to focus on the end product with less thought as to checking the SAS procedural output. However, statisticians reviewing and approving the table will often require the statistical output to determine if the procedure was applied correctly. If prompted to produce the statistical output, a flood of information is typically consigned to a large text file, often with little or no organization which then impedes review. A recommended alternative employing judicious use of titles and options is proposed here to improve the readability and review of the statistical output for both programmers and statisticians.

INTRODUCTION

Clinical studies often require analysis of an endpoint using a statistical model. Some examples are repeated measures, logistic regression, survival or time to event analysis, and regression analysis with stepwise selection. Obviously SAS® is quite capable in performing these types of analysis, the statistician provides code to the programming team, programs are developed, submitted, and information required for outputs in a summary report are generated and presented. Possible information presented in a report table is odds ratios, least square means, and associated confidence intervals. Much of the effort of the programming team, and interest of the report reviewers, is focused on production, quality, and interpretation of the final table. However, as statisticians and programmers, during the generation of the actual report tables, we should also be focused on confirmation that the model is correctly established and assumptions are met, information often not included in the report table, but which SAS does generate with every run of a statistical modeling procedure, such as MIXED, LOGISTIC, GLM, or GLMMIX.

Working for a sponsor or partner that does not require delivery of statistical output associated with report tables, the programming team typically focuses on producing only the report output, and are directed to suppress all SAS statistical output. Since a statistical background is not always a requisite for SAS programmers, original and validation programmers are generally dependant upon statisticians to provide the correct code, and not focused on looking at the SAS statistical output. Statisticians checking the resulting table will often program an independent version, and get different results from the report table. The most effective manner in which to detect discrepancies is in the SAS statistical output of both versions.

Experience has shown that when statistical output is requested from the programming team, possible responses are 1) confusion as to what is requested or how to accomplish, 2) the SAS output saved to one very large text/LST file, or 3) the SAS output consigned to a large number of small text/files as a result of how the original program was designed. Unfortunately, the standard LST or text files produced by SAS can be awkward to navigate, especially if there are many model runs. In one recent example when statistical output was requested, the study had approximately 20 endpoints that were analyzed with a repeated measures model – so there were 20 tables with least square means and confidence intervals. The programming team first sent one text file, approximately 1000 pages long, and had neglected to provide a title or note to indicate which output was associated with which table! The revised version was 20 smaller files, which still did not contain information needed to associate with the tables themselves.

As an alternative, which requires minimal forethought and programming expertise, the standard PDF produced by SAS is another option available to statisticians and programmers. With the use of well placed titles, and perhaps a couple options, statistical output for several different endpoints/tables can be centralized in one file, with automatically generated navigation embedded. The proposed method illustrated here demonstrates an efficient alternative in which to review many statistical models by centralizing diagnostic output created by SAS procedures within one or few PDF files, rather than one or many text files requiring a user to scroll through its entirety.

METHODS AND RESULTS

There is a wealth of information available within the statistical output produced by SAS procedures which allows statisticians and programmers to determine if the model is correctly established, and can be instrumental in

identifying small differences between original and validation programmers, as well as tertiary statistical reviewers. An example of where statistical output review allows us to complete a report table will be illustrated using a mixed model with repeated measures and a covariate.

Consider a study with several biomarkers. Each was sampled three times throughout the study for each subject, of which there were nearly 2000, and Type III tests and p-values, least squares means, and confidence intervals were required for two tables in the clinical study report. The following mock shell examples illustrate the required tables.

EXAMPLE – MIXED MODEL REPEATED MEASURES TABLES

Assessment of Demographic Effects on Biomarkers in Longitudinal Model: MMRM Model				
Biomarkers	Effects	Degrees of Freedom	F Value	p-Value
Biomarker 1	Group	xx	x.xx	x.xxx
	Race	xx	x.xx	x.xxx
	Gender	xx	x.xx	x.xxx
	Age at randomization	xx	x.xx	x.xxx
	Visit	xx	x.xx	x.xxx
	Group*visit	xx	x.xx	x.xxx
Biomarker 2				
...				

Model-Estimated Geometric LSMeans: MMRM Model					
Dose Group (as Randomized)					
Biomarker	Sampling Time Point	Group 1		Group 2	
		Geometric LSMeans	(95% CI)	Geometric LSMeans	(95% CI)
Biomarker 1	Before Dose 1	xx.x	(xx.xx, xx.xx)	xx.x	(xx.xx, xx.xx)
	1 Month after Dose 2	xx.x	(xx.xx, xx.xx)	xx.x	(xx.xx, xx.xx)
	1 Month after Dose 3	xx.x	(xx.xx, xx.xx)	xx.x	(xx.xx, xx.xx)
Biomarker 2					
...					

A mixed-effect model repeated measure (MMRM) with the fixed effects of race, gender, group, visit, group by visit interaction, and age at randomization was employed, with an unstructured covariance matrix to model within-subject correlation, and log-transformed values of the biomarker results as the response variable. In order to obtain the necessary statistics, the following code was employed:

```
PROC MIXED DATA=MODEL_SET METHOD=ML;
  BY ASSAYCD;
  CLASS PATIENT GROUP RACE GENDER VISIT;
  MODEL RV = GROUP RACE GENDER AGE_AT_RANDOMIZATION VISIT GROUP*VISIT / NOINT;
  REPEATED / TYPE=UN SUBJECT=PATIENT R RCORR;
  LSMEANS GROUP*VISIT / CL;
  ODS OUTPUT TESTS3=TABLE1 LSMEANS=TABLE2; RUN;
```

While independently programming the table results, slight discrepancies were encountered, and to efficiently resolve the differences a review of the statistical output was warranted. Upon request to the programming team to save a copy of the SAS statistical output, a text file was quickly produced with a simple addition to the program around the PROC MIXED code:

```
PROC PRINTTO PRINT="STAT_OUTPUT1.LST" NEW; RUN;
PROC PRINTTO PRINT;
```

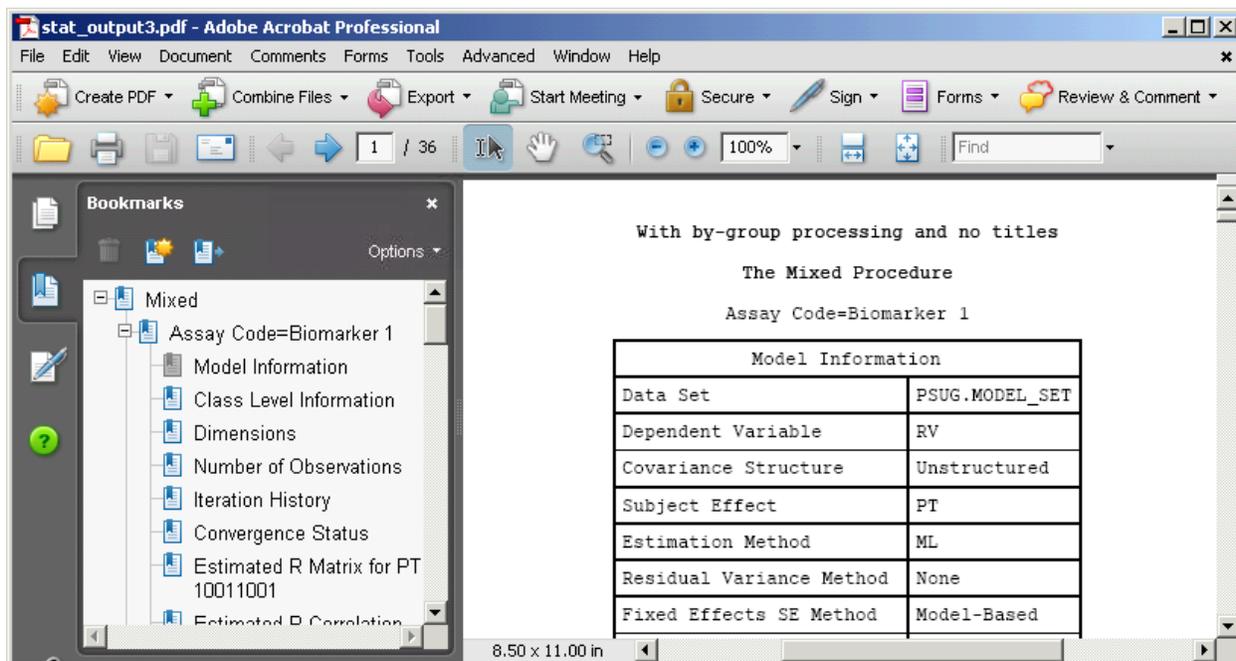
Reviewing the file, which was approximately 100 pages in this case, it was found the program was designed to process the data without the use of by-group processing, so while each of the 4 biomarkers was included, there was

no indication of which biomarker was being presented within the output. After some minor adjustment to use a BY statement within the PROC MIXED as the code above illustrates, the standard output then presented indication of the biomarker of interest, though was still nearly 100 pages long, and was very cumbersome to scroll through the file.

Another possible way of storing this information for others to review, that is arguably just as easy, is to remove the use of the PRINTTO statement, and to replace it with the directive to send the output to a PDF destination:

```
ODS PDF FILE="STAT_OUTPUT2.PDF" STYLE=MINIMAL; RUN;
ODS PDF CLOSE;
```

The resulting file contains all the information the text output contained, however, file size and page number was instantly reduced to nearly a third of the original size. Rather than nearly 100 pages of output to scroll through, the file to review now has 36 pages, and the ability to navigate the document is provided for through automatically generated bookmarks for each of the modules produced by PROC MIXED (Display 1).



Display 1. Statistical Output in PDF Format

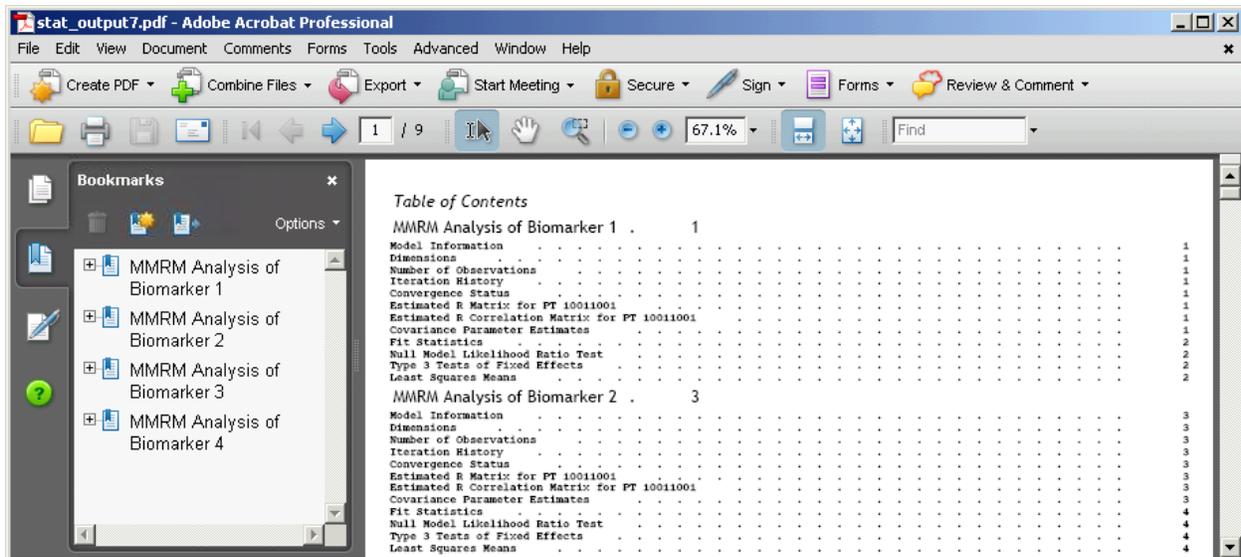
Going one step further, it may be possible to be even more efficient with the information included in the statistical output for review. For example, if it is determined that the datasets and variable assignments going into the model are identical, it may no longer be necessary to require a review of the class level information module. Using the ODS TRACE function, the class level module CLASSLEVELS was identified. By adding the following statement this information may be excluded from the next version:

```
ODS EXCLUDE CLASSLEVELS;
```

With this adjustment, the resulting statistical output file was reduced to 8 pages. Additional modules could have been excluded. Alternatively, inclusion of just the relevant modules necessary to compare with the final reports from PROC MIXED resulted in just 4 pages of statistical output with the following:

```
ODS SELECT TESTS3 LSMEANS;
```

Users may prefer to customize the file a bit more in order to make the review easier. Some options which may be effective are to 1) add a table of contents to the file itself on the first page using the CONTENTS=YES option in the ODS PDF statement, 2) controlling the naming of the automatically generated bookmarks using an ODS PROCLABEL statement prior to each run of the model, and 3) controlling the initial node expansion of the bookmark list to present only the first level (PDFTOC=1 option in ODS PDF statement), which was manually renamed using the ODS PROCLABEL function. Use of these three options is illustrated in Display 2.



Display 2. Augmented Statistical Output in PDF Format

CONCLUSION

The PDF file repository for SAS statistical output associated with clinical study report tables proposed here offers statisticians and programmers many advantages, especially in comparison to output produced for the SAS output window using a PROC PRINTTO. First of all, it is easy! Just turn off the standard output destination, and turn on the PDF with a file location. A second advantage, it is efficient to review and navigate, in comparison to one or several files that require the user to scroll through for useful information. Automatically embedded linkage within the PDF file allows the reviewer to quickly find useful information, and ignore what may be of less importance for a given investigation. A third advantage, the PDF file is extremely portable and easily shared with others.

An important step in the production and validation of a table reporting model-based statistics should always include a review of the statistical output associated with the model that produced the results. The method proposed here offers an efficient alternative for both programmers and statisticians, and can save all team members precious time when trying to resolve discrepancies between the final output and comparative alternatives. Admittedly, this proposed method is just one perspective, and simplistically illustrated. There are many options that could be employed to enhance the review tool, depending on the needs of the reviewer.

ACKNOWLEDGEMENTS

I would like to thank my colleagues in the Biostatistics and Statistical Programming groups at inVentiv Health Clinical who provided comments, feedback, and real-world examples which provided inspiration for the paper topic itself. Their insights and guidance are appreciated.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Please contact the author at:

Bradford J. Danner
 inVentiv Health Clinical
 1787 Sentry Parkway West, Suite 300, Building 16
 Blue Bell, PA 19422 USA
 Work Phone: (615) 806-3017
brad.danner@inventivhealth.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.