

## Using ODS to Capture Statistics from PROC GLM

Frank Roediger, Meridian Software, Inc.

### ABSTRACT

Through SAS V6.12, there were limitations on what statistics were readily available from statistical PROCs. A rich set of statistics was available, just not readily. Programmers commonly worked around this limitation by using PROC PRINTTO to redirect a PROC's output from the Output Window (its default destination) to a file and then programming a DATA step to extract the desired results from that file output. This awkward work-around has been made obsolete by the capabilities of ODS.

### INTRODUCTION

The purpose of this poster is to demonstrate techniques that use ODS to extract values from statistical PROCs. Even though the focus is on PROC GLM and extracting the p-values and least squares means that it can provide, the same techniques can be used to extract results from other statistical PROCs as well. The purpose of this paper is *not* to demonstrate or make the case for the application of a particular statistical approach; that is beyond the scope of the paper (as well as beyond the scope of the author, who is not a statistician!). However, the techniques can help you set up an ODS statistics-extraction tool kit that will be useful in a variety of contexts.

### SETTING THE STAGE

Let's begin by describing a situation (depending upon your own work situation, you can decide whether it is fanciful or not). An analyst at Company X has the responsibility for providing a management report that shows how the sales force for each of Company X's three regions is doing. Upper management has implemented an incentive program and wants to know how well it is working. After a great many meetings with upper management, the programmer has developed a V6.12 SAS program that produces a report that upper management is satisfied with (see Exhibit 1).

The report shows basic descriptive statistics (extracted conventionally from PROC UNIVARIATE) for each region for each tracking interval, but it also includes p-values and least squares means (painstakingly extracted via PROC PRINTTO from PROC GLM). Upper management, now that it has the report that it wants, has decreed that it wants this output ongoing and exactly the same every tracking interval. However, the analyst has just received a memo from the IT department that before the next report is due, SAS V6.12 will no longer be available on Company X's computers and that SAS V8 will be the new company standard.

Because the analyst knows that, historically, new releases of SAS have been backwardly compatible, he figures that he'll be able to continue running the program under the new V8 release. However, because he has free time in his schedule (remember: you can decide whether this scenario is fanciful or not), he decides to test the program ahead of time under V8, just to make sure. To his satisfaction, it runs without any warnings or errors; to his horror, all of the p-values and least squares means from PROC GLM are now missing.

After some digging, he discovers the problem: PROC GLM output is formatted differently in V8 than it is under V6.12 (see Exhibits 2A & 2B). Here, he is faced with two alternatives: 1) change the INPUT statements in his DATA step so that they extract results from their new V8-specific location; or 2) bite the bullet and take the opportunity to convert his program to V8-specific techniques that involve ODS. Wisely, he chooses the latter.

### METHODOLOGY

#### 1. IDENTIFY VALUES IN STANDARD OUTPUT

Actually, our hypothetical analyst has already begun the conversion process: the first step is to identify from a standard PROC listing the values that are needed (see highlighted values on Exhibit 2C). His challenge now is to determine where ODS has stored these values.

#### 2. IDENTIFY CANDIDATE ODS OUTPUTS

ODS stores the values that appear in the standard PROC output listing as fields in SAS data sets. Every PROC has its own set of ODS-related data sets (the ones for PROC GLM are listed in Exhibit 3). However, ODS does not create all of these data sets every time the PROC is run: depending upon the options that have been used, ODS creates just the subset of these data sets that is needed to represent each of the values from the standard PROC listing as a variable in a data set.

SAS has provided a tool to help focus the search: the ODS trace statement. Inserting this statement before the block of code that contains the PROC GLM step indicates to SAS that you want to see a list of all the outputs that ODS will need to create to store the values generated by the PROC GLM step. (The ODS trace statement works like an OPTIONS statement; the statement before the PROC GLM block turns the feature on, the one after the block turns the feature off.) ODS displays this list in the Log Window (see Exhibit 4).

One (or more) of the entries in this list refers to the outputs that will contain the results we are seeking. Each entry in the list has four identifiers: Name, Label, Template, and Path. The two identifiers we are particularly interested in are "Name," which is the SAS data set reference, and "Label," which describes the type of information that is stored in that SAS data set (e.g., the "Diff Name is the reference for the output whose contents have the "Difference Matrix" Label).

It is important to note that, at this point, we have only gathered information that we will use later on to create SAS data sets – we have not yet created any of the data sets themselves. In fact, we are not even sure we know which of these outputs contains the values that we want. All we do know is that those values will be in one (or more) of the outputs listed in the Log by the TRACE statement.

To narrow down our search, we can use the headings in the standard PROC listing (see Exhibit 2C). By studying that listing, we can observe that the values we seek are grouped into three clusters; therefore, we can reasonably guess that the values will be in three different outputs.

An obvious guess for the source of the least squares means is the output labeled "LSMeans." The overall p-value appears by itself in a section of the standard listing that is titled, "Type III SS," so a reasonable candidate for its output source would be the one labeled, "Type III Model ANOVA." The inter-group differences p-values do not have a similarly apparent source, but we can make a guess that their output source might be the one labeled, "Difference Matrix."

#### 3. CREATE DATA SETS FROM ODS OUTPUTS

Now that we have candidate outputs for the values we are seeking, we need to tell SAS that we would like those outputs created as SAS data sets. To do that, we use the ODS OUTPUT statement. Because, at this point, we no longer need the ODS trace statement, we can replace it with a set of ODS output

statements. One OUTPUT statement is needed for each data set that we want SAS to create – three outputs, three output statements (see Exhibit 5). Following the PROC GLM block, we insert another ODS statement to close the three data sets. Now, we have created the data sets, but, before we use them, we first need to make sure that they contain the values we are seeking.

#### 4. IDENTIFY VALUES' FIELDS IN DATA SETS

We need to view a raw dump of the three data sets. Exhibit 6A is a PROC CONTENTS listing of the LSMEANS data set. Note that the FORMAT attribute for the PROB\_T field is PVALUE6.4, a w.d-type format that SAS provides specifically for the presentation of p-values. Exhibit 6B is a PROC PRINT of the LSMEANS data set with the target values highlighted; this confirms that the LSMEANS data set does contain the values that we seek. Similar PROC CONTENTS and PROC PRINT listings of the other two data sets that we created will similarly confirm that they contain the other values that we seek.

The only remaining task is to replace the old DATA step, which under V6.12 had used INFILE/INPUT statements to extract the values from the file that contained the re-directed PROC GLM output. Its replacement will be a series of DATA steps that process the values from the three data sets we have identified. Admittedly, under V6.12, one DATA step did the work that under V8 now requires three DATA steps, but each of these three DATA

steps is a simple extraction process. Also, the values that we will now have are direct numeric values, not the indirect ones that we had to translate from the character strings in the re-directed PROC GLM output under V6.12.

#### 5. HOUSEKEEPING

One additional – and optional, though highly recommended – task remains: housekeeping. Under V8, ODS creates an entry in the Results window for each execution of the PROC GLM step. Our particular report has three executions of that PROC GLM step, so there will be three new entries in the Results window after the job has run. If we are submitting only this one job, then these Results window entries are not an issue. However, if we are submitting a whole string of jobs within the same SAS session to create a series of reports, it is possible for these Results window entries to overwhelm a PC's resources and cause an abnormal termination of the SAS session. Therefore, it is a good practice to programmatically clear out the Results window as the first step in the job (if you clear it out as the last step, the entries will not be available in case you need them for debugging). The following display manager command can achieve this for you:

```
dm odsresults 'clear';
```

### EXHIBIT 1 - SAMPLE REPORT

Table xx.y  
Summary of Sales Results  
Sales Group

	1	2	3	Sales Group Comparisons			
				1 vs. 2	1 vs. 3	2 vs. 3	Overall
	Eastern Region (N=285)	Central Region (N=313)	Western Region (N=305)				
Baseline							
n	285	313	305				
Mean	29,917	31,996	31,732				
SD	15,771	15,243	15,833				
Median	30,995	32,690	32,258				
Min.	3,341	3,264	3,570				
Max.	57,837	58,175	58,135				
LS Mean	30,541	32,574	32,339	0.112	0.162	0.851	0.226
Week 24							
n	285	313	305				
Mean	31,946	30,693	31,267				
SD	15,053	15,807	15,633				
Median	33,376	31,398	31,274				
Min.	3,592	3,285	3,352				
Max.	58,099	58,217	58,168				
LS Mean	31,962	30,733	31,304	0.335	0.607	0.648	0.628
Week 48							
n	285	313	305				
Mean	29,193	31,110	30,524				
SD	15,924	15,900	15,719				
Median	28,001	31,530	31,278				
Min.	3,438	3,311	3,716				
Max.	57,893	58,128	58,130				
LS Mean	29,325	31,071	30,535	0.179	0.354	0.675	0.392

**EXHIBIT 2A. SAMPLE SAS V6.12 OUTPUT**

General Linear Models Procedure

Source	DF	Type III SS	Mean Square	F Value	Pr > F
REGION	2	224560940.42748600	112280470.21374300	0.47	0.6283
BASRLT	1	26788745.36482850	26788745.36482850	0.11	0.7391
STRATA	1	1475228.00555101	1475228.00555101	0.01	0.9377
SEX	1	844696.35443265	844696.35443265	0.00	0.9528

**EXHIBIT 2B. SAMPLE SAS V8.02 OUTPUT**

The GLM Procedure

Source	DF	Type III SS	Mean Square	F Value	Pr > F
REGION	2	224560940.4	112280470.2	0.47	0.6283
basrlt	1	26788745.4	26788745.4	0.11	0.7391
strata	1	1475228.0	1475228.0	0.01	0.9377
SEX	1	844696.4	844696.4	0.00	0.9528

**EXHIBIT 2C. SAS V8.02 OUTPUT**

The GLM Procedure

Source	DF	Type III SS	Mean Square	F Value	Pr > F
REGION	2	224560940.4	112280470.2	0.47	<u>0.6283</u>
basrlt	1	26788745.4	26788745.4	0.11	0.7391
strata	1	1475228.0	1475228.0	0.01	0.9377
SEX	1	844696.4	844696.4	0.00	0.9528

Least Squares Means

REGION	TOTSALES LSMEAN	Standard Error	Pr >  t	LSMEAN Number
1	<u>31962.1545</u>	991.5644	<.0001	1
2	<u>30732.6810</u>	943.3254	<.0001	2
3	<u>31303.5004</u>	959.6158	<.0001	3

Least Squares Means for effect REGION  
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: TOTSALES

i/j	1	2	3
1		0.3351	0.6074
2	<u>0.3351</u>		0.6482
3	<u>0.6074</u>	<u>0.6482</u>	

**EXHIBIT 3. PROC GLM'S POTENTIAL ODS DATA SETS\***

ODS Table Name	Description	Statement / Option
Aliasing	Type 1,2,3,4 aliasing structure	MODEL / (E1 E2 E3 or E4) and ALIASING
AltErrContrasts	ANOVA table for contrasts with alternative error	CONTRAST / E=
AltErrTests	ANOVA table for tests with alternative error	TEST / E=
Diff	<b>PDiff matrix of Least-Squares Means</b>	<b>LSMEANS / PDIFF</b>
LSMeans	<b>Least-Squares means</b>	<b>LSMEANS statement</b>
Means	Group means	MEANS statement
ModelANOVA	<b>ANOVA FOR MODEL TERMS</b>	<b>default</b>
SimDetails	Details of difference quantile simulation	LSMEANS / ADJUST=SIMULATE(REPORT)
SimResults	Evaluation of difference quantile simulation	LSMEANS / ADJUST=SIMULATE(REPORT)
SlicedANOVA	Sliced effect ANOVA table	LSMEANS / SLICE
Sphericity	Sphericity tests	REPEATED / PRINTE
Tests	Summary ANOVA for specified MANOVA H= effects	MANOVA / H= SUMMARY
Tolerances	<b>X'X</b> Tolerances	MODEL / TOLERANCE
Welch	Welch's ANOVA	MEANS / WELCH
XpX	<b>X'X</b> matrix	MODEL / XPX

\*A portion of the table from SAS OnlineDOC: SAS/Stat→User's Guide→GLM Procedure→Details→ ODS Table Names  
The table has been truncated to save space.

**EXHIBIT 4. SAS LOG FOR PROC GLM WITH ODS TRACE ON/OFF\***

```

461 ods trace on;
462 proc glm data=analysis.SALES (where=(week eq &week));
463   class REGION strata sex;
464   model TOTSALES=REGION basrlt strata sex;
465   lsmeans REGION / stderr pdiff;
466   means REGION;
467 quit;

```

Output Added:

-----

```

Name:      ClassLevels
Label:     Class Levels
Template:  STAT.GLM.ClassLevels
Path:     GLM.Data.ClassLevels
-----

```

Output Added:

-----

```

Name:      ModelANOVA
Label:     Type III Model ANOVA
Template:  stat.GLM.Tests
Path:     GLM.ANOVA.totfat.ModelANOVA
-----

```



## EXHIBIT 4. (CONT'D)

Output Added:

-----

Name: LSMeans  
Label: LSMeans  
Template: stat.GLM.LSMeans  
Path: GLM.LSMEANS.REGION.TOTSALES.LSMeans

←

Output Added:

-----

Name: Diff  
Label: Difference Matrix  
Template: stat.GLM.PDiff  
Path: GLM.LSMEANS.REGION.TOTSALES.Diff

←

NOTE: PROCEDURE GLM used:

real time 0.27 seconds  
cpu time 0.25 seconds

468 ods trace off;

\*The actual log has been truncated to save space.

## EXHIBIT 5. SAS LOG FOR PROC GLM WITH ODS OUTPUT STATEMENTS \*

```
473 ods output ModelANOVA=modela;  
474 ods output Diff=diff;  
475 ods output LSMeans=LSMeans;  
  
476 proc glm data=analysis.SALES (where=(week eq &week));  
477 class REGION strata sex;  
478 model TOTSALES=REGION basrlt strata sex;  
479 lsmeans REGION / stderr pdiff;  
480 means REGION;  
481 quit;
```

NOTE: PROCEDURE GLM used:

real time 0.54 seconds  
cpu time 0.41 seconds

482 ods output close;

\*The actual log has been truncated to save space.

## EXHIBIT 6A. CONTENTS FOR PROC GLM'S LSMEANS OUTPUT DATA SET

The CONTENTS Procedure

Data Set Name:	WORK.LSMEANS	Observations:	3
Member Type:	DATA	Variables:	7
Engine:	V8	Indexes:	0
Created:	12:10 Wednesday, June 12, 2002	Observation Length:	56
Last Modified:	12:10 Wednesday, June 12, 2002	Deleted Observations:	0
Protection:		Compressed:	NO
Data Set Type:		Sorted:	NO
Label:	LSMeans		

## EXHIBIT 6A. (CONT'D)

-----Engine/Host Dependent Information-----

```
Data Set Page Size:      8192
Number of Data Set Pages:  1
First Data Page:         1
Max Obs per Page:       145
Obs in First Data Page:   3
Number of Data Set Repairs: 0
File Name:               g:\temp\SAS Temporary Files\_TD347\lsmeans.sas7bdat
Release Created:         8.0202M0
Host Created:            WIN_NT
```

-----Alphabetic List of Variables and Attributes-----

#	Variable	Type	Len	Pos	Format	Label
2	Dependent	Char	8	41		
1	Effect	Char	9	32		
4	LSMean	Num	8	0	12.4	TOTSALES LSMEAN
7	LSMeanNumber	Num	8	24	BEST8.	LSMEAN Number
6	Probt	Num	8	16	<u>PVALUE6.4</u>	Pr >  t
3	REGION	Char	1	49		
5	StdErr	Num	8	8	12.4	Standard Error

## EXHIBIT 6B. DUMP OF PROC GLM'S LSMEANS OUTPUT DATA SET

Obs	Effect	Dependent	REGION	LSMean	StdErr	Probt	LSMean Number
1	REGION	TOTSALES	1	<u>31962.1545</u>	991.5644	<.0001	1
2	REGION	TOTSALES	2	<u>30732.6810</u>	943.3254	<.0001	2
3	REGION	TOTSALES	3	<u>31303.5004</u>	959.6158	<.0001	3

### CONCLUSION

ODS has simplified the process of getting convenient access to statistical results. Following a few simple steps is all that is needed to get statistics that, with earlier versions of SAS, were available only from Output listings.

### REFERENCES

SAS Institute, Inc. (1999), *SAS Online Doc, Version 8, First Edition*, Cary, NC: SAS Institute, Inc.

### ACKNOWLEDGMENTS

Eric Brinsfield and Caroline Bahler (Meridian Software, Inc.) for helpful suggestions and recommendations.  
Cathy Brinsfield (Meridian Software, Inc.) for editorial suggestions and for help with putting this paper in its final form.

### TRADEMARKS

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. Meridian Software, Inc.® is a registered trademark of Meridian Software, Inc. Other brand and product names are registered trademarks or trademarks of their respective companies.

### CONTACT INFORMATION

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

Frank Roediger  
Meridian Software, Inc.  
12204 Old Creedmoor Rd.  
Raleigh, NC 27613-7213  
Work Phone: 919.786-0889  
Fax: 919.518-1170  
Email: merflr@meridiansoftware.com