

Paper 198-25

Publishing Statistical Analyses on the Web Using Version 7 of the SAS® System

David J. Meade, Advanced Micro Devices, Austin, TX
Scott Lacey, Advanced Micro Devices, Austin, TX

ABSTRACT

One of the most difficult parts of data analysis can be communicating the results to other interested parties. Today, one of the most popular and effective method of communication is the World Wide Web. Recent advances in Web based programming in SAS® version 7, have made the process of publishing information on the Web a relatively simple process. In this paper we demonstrate how output from standard analysis procedures such as PROC ANOVA, PROC GLM, and PROC FREQ can be published on the Web using the tools available in SAS version 7.

INTRODUCTION

When analyzing data, researchers are always faced with the decision as to how the information should be reported. In the past we have relied on vehicles such as printed output. In this case stacks of paper end up being circulated from desk to desk. If errors are discovered after the report has been printed, then even more stacks of paper are run through the copier and circulated. Graphics included in these reports are normally black and white, as color copies are very expensive.

Another method of reporting information is via e-mail. This method is much more efficient than the paper copy method. Full color, paperless reports can be sent with minimal effort. However, you still face problems when errors are discovered or changes need to be made to the report itself.

In recent years, the World Wide Web (WWW) has become a very popular method for communicating and reporting information. The WWW has many advantages. Stacks of paper reports are eliminated. Tables and graphics are presented in full color. Changes can be made to

the report at any time. The list goes on and on. With recent advances in SAS version 7, publishing information on the WWW has never been easier. The new SAS Output Delivery System (ODS), makes it possible to convert SAS reports into the HTML format, which can then be directly displayed on the internet. In this paper we will introduce the new ODS and provide clear examples which demonstrate how it can be used to publish information on the WWW.

THE OUTPUT DELIVERY SYSTEM

The following excerpt from the SAS Version 7 online help provides a good introduction to the Output Delivery System. "Prior to Version 7, SAS procedures that produced printed output (that is, output destined for the procedure output file) generated output that was designed for a traditional line-printer. In Version 7, procedure output is much more flexible. "

"The Output Delivery System (ODS) enhances your ability to manage procedure output. Procedures that fully support ODS "

- "combine the raw data that they produce with one or more templates to produce one or more output objects that contain the formatted results"
- "store a link to each output object in the Results folder in the Results window"
- "can generate HTML files that contain the formatted results and links (such as a table of contents) to the results"
- "can generate output data sets from procedure output"
- "provide a way for you to customize the procedure output by creating templates that you can use whenever you run the procedure."

These new features combined with the analytical tools available in SAS, is a powerful combination. Throughout the remainder of this paper we will provide examples of how the ODS can be used to convert output from three SAS Procedures into HTML. The procedures we will illustrate are PROC ANOVA, PROC GLM, and PROC FREQ.

USING THE OUTPUT DELIVERY SYSTEM IN CONJUNCTION WITH PROC ANOVA

PROC ANOVA is frequently used to identify differences between averages of three or more treatment groups. Consider the follow SAS Version 7 dataset (fitness.sas7bdat), which can be automatically generated by SAS. This dataset contains information regarding the fitness of three groups of people, see Figure 1. The treatment groups are labeled in the “group” column as 0, 1, and 2. The variable labeled “runtime” is the number of minutes it took each person to run 1.5 miles. PROC ANOVA can be used to determine whether a significant difference exists for the “runtime” variable between treatment groups.

31	8	Int	Int	Int	Int	Int	Int	Int	Int
		age	weight	runtime	rstpulse	runpulse	maxpulse	oxygen	group
1	57	73.37	12.63	58	174	176	39,407	2	
2	54	79.36	11.17	62	156	165	46,090	2	
3	52	76.32	9.63	48	164	166	45,441	2	
4	50	70.87	8.32	48	146	155	54,625	2	
5	51	67.25	11.08	48	172	172	45,118	2	
6	54	91.63	12.88	44	168	172	39,203	2	
7	51	73.71	10.47	59	186	188	45,790	2	
8	57	59.08	9.93	49	148	155	50,545	2	
9	49	76.32	9.40	56	186	188	48,673	2	
10	48	61.24	11.50	52	170	176	47,920	2	
11	52	82.78	10.50	53	170	172	47,467	2	
12	44	73.03	10.13	45	168	168	50,541	1	
13	45	87.66	14.03	56	186	192	37,388	1	
14	45	66.45	11.12	51	176	176	44,754	1	
15	47	79.15	10.60	47	162	164	47,273	1	
16	54	83.12	10.33	50	166	170	51,895	1	
17	49	81.42	8.95	44	180	185	49,156	1	
18	51	69.63	10.95	57	168	172	40,836	1	
19	51	77.91	10.00	48	162	168	46,672	1	
20	48	91.63	10.25	48	162	164	46,774	1	
21	49	73.37	10.08	76	168	168	50,388	1	
22	44	89.47	11.37	62	178	182	44,609	0	
23	40	75.07	10.07	62	185	185	45,313	0	
24	44	85.84	8.65	45	156	168	54,297	0	
25	42	68.15	8.17	40	166	172	59,571	0	
26	38	89.02	9.22	55	178	180	49,874	0	
27	47	77.45	11.63	58	176	176	44,811	0	
28	40	75.98	11.95	70	176	180	45,681	0	
29	43	81.19	10.85	64	162	170	49,091	0	
30	44	81.42	13.08	63	174	176	39,442	0	
31	38	81.87	8.63	48	170	186	60,055	0	

Figure 1: The SAS Version 7 “fitness.sas7bdat” dataset. Fitness data is shown for three treatment groups.

The ANOVA can be executed without the aid of the ODS by using the standard SAS code shown in Figure 2. The standard PROC ANOVA output for this example is shown in Figure 3. In this example, no significant difference exists between the treatment groups.

However, the analysis output takes on a much more polished appearance and has the advantage of being available via the WWW if PROC ANOVA is executed in conjunction with the ODS. Figure 4 illustrates how the original SAS code needs to be modified in order to implement the ODS. Figure 5 illustrates how the PROC ANOVA output appears after it has been converted to HTML and posted on the Web.

```
proc anova data=sasuser.fitness;
  class group;
  model runtime = group;
  title;
run;
```

Figure 2: Standard SAS code necessary to execute PROC ANOVA using the fitness dataset.

```

The GLM Procedure
Class Level Information
Class      Levels  Values
group      3      0 1 2

Number of observations   31

The GLM Procedure
Dependent Variable: runtime   Min. to run 1.5 miles

Source      DF      Squares    Mean Square    F Value    Pr > F
Model       3      2.38182273  0.79394091    0.39      0.7631
Error      27      55.36571275  2.05058195

Corrected Total      30      57.74753548

```

Figure 3: Standard PROC ANOVA output comparing the “runtime” variable across treatment groups. No significant difference appears to exist.

```

SAS Code for Standard ANOVA HTML

/* start redirecting output to html */
ods html body= 'plain_anova_body.html'
  contents='plain_anova_toc.html'
  page= 'plain_anova_page.html'
  frame= 'plain_anova_all.html';

proc anova data=sasuser.fitness;
  class group;
  model runtime = group;
  title;
run;

/* close the HTML file */
ods html close;
```

Figure 4: SAS code enabling PROC ANOVA to be used in conjunction with the Output Delivery System.

The ODS can also be used to include graphics in the analysis output. For the example at hand, we use PROC GPLOT to generate a set of comparative boxplots. These boxplots enable the reader to visualize the difference in “runtime” between the various treatment groups. The code shown in Figure 6, uses PROC ANOVA, PROC GPLOT, and the ODS to produce an online report containing text output as well as the

comparative boxplots discussed above. Figure 7 illustrates the final report.

USING THE OUTPUT DELIVERY SYSTEM IN CONJUNCTION WITH PROC GLM

PROC GLM is frequently used to fit a general linear model to a set of data. The ODS can, once again, be used to format the output in HTML. We will illustrate this process using the fitness data. For this example, we will fit a general linear model using the “runtime” variable as the response. The “age” and “group” variables will be used as the explanatory variables in the model. The standard SAS code necessary to execute PROC GLM is shown in Figure 8. The standard output from the analysis is shown in Figure 9. Once again, the output can be formatted in HTML and then displayed on the Web by incorporating the ODS. Figure 10 illustrates how the ODS can be integrated with PROC GLM. The resulting online output is shown in Figure 11. As shown previously, the ODS can be used in conjunction with plotting procedures to include graphics in the HTML output. We can modify the code shown in Figure 10, to include a call to PROC GPLOT. The resulting graphic illustrates “runtime” plotted by “group” across the various levels of “age”. Figure 12 illustrates how the SAS code needs to be modified in order to generate the graphic. Figure 13 illustrate the final online output.

SUMMARY

Effectively communicating information is one of the most important parts of data analysis. The World Wide Web is rapidly becoming one of the most popular, if not the most popular, and efficient methods of communication available. The Output Deliver System (ODS) in SAS Version 7, is a remarkable advancement in SAS programming. Data analysts can now format the output of any procedure in HTML. Application of the ODS to specific situations are simply modifications or extensions of the examples in this paper. Appendix 1 illustrates how the ODS can be used in conjunction with PROC FREQ.

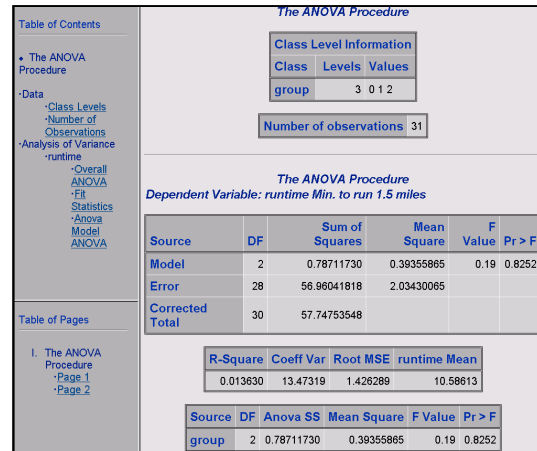


Figure 5: PROC ANOVA output formatted in HTML using the Output Delivery System. Please note that some rows of output are not visible in this figure.

```

SAS Code for Graph + Standard ANOVA HTML

/* start redirecting output to html */
ods html body= 'graph_anova_body.html'
  contents='graph_anova_toc.html'
  page= 'graph_anova_page.html'
  frame= 'graph_anova_all.html';

/* generate graphics */
goptions colors=(black) cback=white
device=imggif hsize=5 in vsize=2 in
  htext=4 pct gsfname=grafout
gsfmode=replace;
symbol1 v=plus i=boxt color=green;
symbol2 v=plus i=boxt color=blue;
symbol3 v=plus i=boxt color=red;
axis1 offset=(2,2);
filename grafout './anova.gif';
title h=5 pct "Runtime across Group";
proc gplot data=sasuser.fitness;
  plot runtime*group=group / haxis=axis1
;
run; quit;

proc anova data=sasuser.fitness;
  class group;
  model runtime = group;
  title;
run;

/* be sure to quit interactive procedures
while under ods directive */
quit;

/* close the HTML file */
ods html close;

```

Figure 6: SAS code combining PROC ANOVA, PROC GPLOT, and the Output Delivery System.

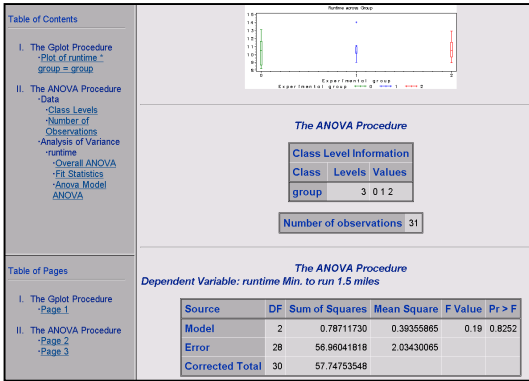


Figure 7: Online report generated from the code shown in above. PROC GPLOT, PROC ANOVA, and the Output Delivery System are used. Please note that some rows of the final output are not visible in the figure.

```

SAS Code for Standard GLM HTML

/* start redirecting output to html */
ods html body= 'plain_glm_body.html'
  contents='plain_glm_toc.html'
  page= 'plain_glm_page.html'
  frame= 'plain_glm_all.html';

proc glm data=sasuser.fitness;
  class group;
  model runtime = age group;
  title;
run; quit;

/* close the HTML file */
ods output close;
    
```

Figure 10: SAS Code incorporating PROC GLM with the Output Delivery System.

```

Proc glm data=sasuser.fitness;

      class group;
      model runtime= age group;
      title;

run;
quit;
    
```

Figure 8: Standard SAS code necessary to execute PROC GLM using the fitness data.

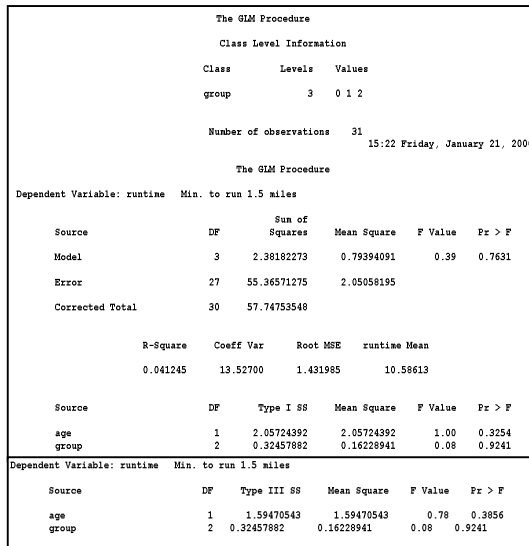


Figure 9: Standard SAS output for PROC GLM and the fitness data.

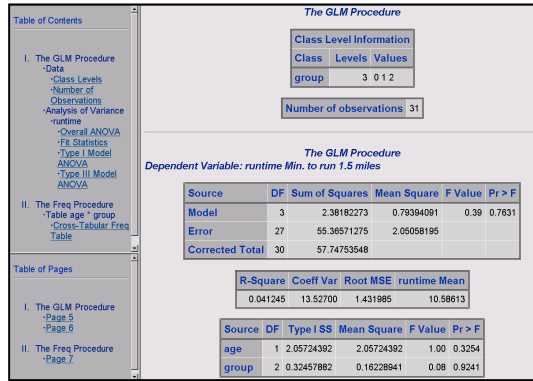


Figure 11: Online output generated using the Output Delivery System code shown above. Please note that some lines of output are not visible in this figure.

SAS Code for Graph + Standard GLM HTML

```

/*Note that the ODS does not handle the optional output datasets
generated by the output statement of stat procedures like glm. It
only handles the printed output of these procedures.*/
/* cut down to necessary columns - will use
append later */
data tmp; set sasuser.fitness(keep=age group
runtime); run;

/* get predicted values from model */
proc glm data=tmp;
  class group;
  model runtime = age group;
  output out=outstat
  predicted=runtime_pred;
run; quit;

/* assign predicted values to a new group
for graph */
data outstat(rename=(runtime_pred=runtime));
  set outstat(drop=runtime);
  group=4;
run;
/* append to original data */
proc append data=outstat base=tmp; run;

/* start redirecting output to HTML */
ods html body= 'graph_glm_body.html'
  contents='graph_glm_toc.html'
  page= 'graph_glm_page.html'
  frame= 'graph_glm_all.html';

/* generate graphics */
goptions colors=(black) cback=white
device=imggif hsize=5 in vsize=2 in
  htext=4 pct gsfname=grafout
gsfmode=replace;
symbol1 v=circle i=r1 line=2
color=green;
symbol2 v=square i=r1 line=3 color=blue;
symbol3 v=triangle i=r1 line=4 color=red;
symbol4 v=point i=join line=1
color=black;
legend1 label=('Group') value=(tick=4
'Pred');

/* be sure predicted values will line up
properly */
proc sort data=tmp; by age; run;

filename grafout './glm.gif';
title h=5 pct "Runtime vs Age across Group";
proc gplot data=tmp;
  plot runtime*age=group / legend=legend1;
run; quit;

/* generate output for HTML */
proc glm data=sasuser.fitness;
  class group;
  model runtime = age group;
  title;
run; quit;

/* close the HTML file */
ods output close;

```

Figure 12: SAS code incorporating PROC GLM, PROC GPLOT, and the Output Delivery System.

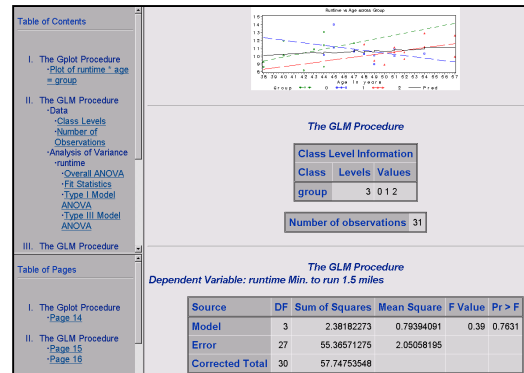


Figure 13: Final online output resulting from the SAS code shown in Figure 12.

APPENDIX 1

USING THE OUTPUT DELIVERY SYSTEM IN CONJUNCTION WITH PROC FREQ

Objective: Use PROC FREQ to determine age distribution across treatment groups. Format the output in HTML using the Output Delivery System. Figures 14, 15, and 16 show the SAS code and resulting online output.

SAS Code for Graph + Standard Freq HTML

```

/*A nice feature of the ODS is that output can be redirected into
a dataset. In order to do this one needs to know the name of the
output objects. The ods trace option puts a list of the output
objects to the log when the objects are created.*/

/* use trace output to identify output
objects generated */
ods trace output;

proc freq data=sasuser.fitness;
  table age*group;
  title;
run;

ods trace off;

/* this shows up in the log:

Output Added:
-----
Name:          CrossTabFreqs
Label:         Cross-Tabular Freq Table
Data Name:
Path:
Freq.age_by_group.CrossTabFreqs
-----
*/

```

Figure 14: SAS code incorporating PROC FREQ, PROC GPLOT, and the Output Delivery System.

```

/* get analysis results into a dataset */
ods output crosstabfreqs=stats;

proc freq data=sasuser.fitness;
  table age*group;
  title;
run;

ods output close;

/* start redirecting output to HTML - just
one table, so only generate one page */
ods html body='graph_freq.html';

/* generate graphics */
goptions colors=(black) cback=white
device=imggif hsize=5 in vsize=2 in
  htext=4 pct gsfname=grafout
gsfmode=replace;
symbol1 v='1' h=4 pct;
symbol2 v='2' h=4 pct;
symbol3 v='3' h=4 pct;
symbol4 v='4' h=4 pct;
symbol5 v='5' h=4 pct;
symbol6 v='6' h=4 pct;
symbol7 v='7' h=4 pct;
symbol8 v='8' h=4 pct;
symbol9 v='9' h=4 pct;
symbol10 v=plus h=2 pct;
axis1 offset=(2,2);

/* set symbol column - character sorting */
data stats; set stats(where=(type_='11'));
  length symbol $1; if frequency>0;
  /* : is after 9 in ascii */
  if frequency>9 then symbol=':.';
  else symbol=put(frequency,1.);
run;

filename grafout './freq.gif';
title h=5 pct "Freq of Age across Group";
proc gplot data=stats;
  plot group*age=symbol / vaxis=axis1
  haxis=axis1;
run; quit;

/* generate output for HTML */
proc freq data=sasuser.fitness;
  table age*group;
  title;
run;

/* close the HTML file */
ods html close;

```

Figure 15: Second half of the code necessary to incorporate PROC FREQ, PROC Gplot, and the Output Delivery System.

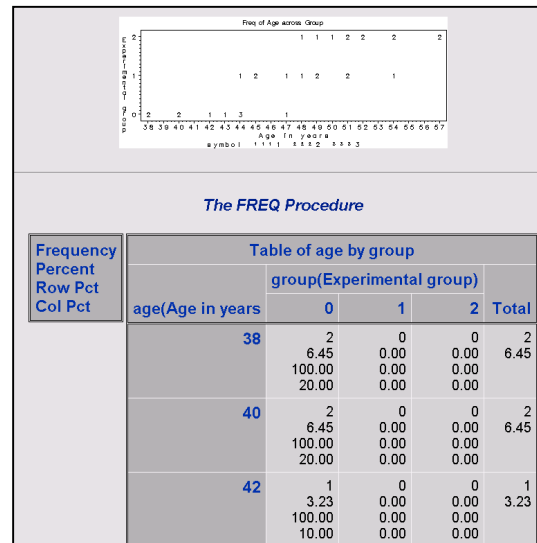


Figure 16: Online output resulting from the code shown above.

ACKNOWLEDGMENTS

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

CONTACT INFORMATION

David J. Meade
 5204 E. Ben White Blvd.
 Mail Stop 613
 Austin, TX 78741
 USA
 Email: david.meade@amd.com

Scott Lacey
 5204 E. Ben White Blvd.
 Mail Stop 601
 Austin, TX 78741
 USA
 Email: scott.lacey@amd.com