

Paper 244-2009

Image is Almost Everything: Displaying Statistics Via ODS

R. Scott Leslie, MedImpact Healthcare Systems, Inc., San Diego, CA

ABSTRACT

After conducting statistical analyses you must find a way to report the results. A simple print of the traditional SAS[®] output can disinterest most readers, especially those not familiar with statistics. The Output Delivery System (ODS) gives you the ability to output results of statistical tests into data tables and graphs. With some manipulation and formatting these data tables can be sent to the possible ODS destinations to generate appealing results therefore avoiding manual data entry. This paper offers ways to make a nice presentation of numbers to assist an audience in reading and interpreting results. Examples include using the LOGISTIC and GLM procedures to better display odds ratios and adjusted least-squares means.

INTRODUCTION

Have you ever had the need to show results of statistical analyses to someone less familiar with statistics than yourself? After conducting your analyses you are left with a bunch of output to explain. Traditional SAS[®] output from procedures is familiar and easy to read to some but less easy and cumbersome for others. You are usually left with tedious cutting, pasting and formatting to get results of tests out to tables or text.

Using the ODS OUTPUT statement you can create dressed up tables that are easy to read and cut/paste to a document or spreadsheet for graphing. The purpose of this paper is to offer ways to transform traditional SAS[®] output to more reader-friendly content for presentation. This code can be a helpful in displaying statistics or any SAS[®] output.

USING THE OUTPUT DELIVERY SYSTEM (ODS) TO YOUR ADVANTAGE

Briefly, the ODS system allows you to generate, store, and format DATA step and SAS[®] procedure output. Instead of just using output from the listing, the ODS allows you to output data to multiple third party destinations that include HTML, RTF and PRINTER (PDF). For more detail on the ODS functions and capabilities, read the online documentation, SAS OnlineDoc[®] (link in reference section).

One of the great features of ODS is the ability to output the results of SAS[®] procedures to data sets. I find this extremely useful in displaying results of statistical tests. I first output the results of a procedure to a data set. Then I use PROC PRINT or PROC TABULATE to create tables for insertion into a document or spreadsheet for further discussion.

The ODS TRACE statement is a great starting point because it identifies parts of the output that procedures generate. Below is an example for a UNIVARIATE procedure. This statement will write all the objects produced by the PROC UNIVARIATE to the SAS[®] log.

```
ODS TRACE On;
PROC UNIVARIATE data=sample;TABLES age;RUN;
ODS TRACE Off;
```

The text box below contains some of the result of running the above code. The SAS[®] log lists the ODS objects for the PROC UNIVARIATE. Here are two of the five objects. You can match the name from the listing to the names of the objects in the log.

```

Output Added:
-----
Name:      Moments
Label:     Moments
Template:  base.univariate.Moments
Path:     Univariate.age.Moments
-----

Output Added:
-----
Name:      BasicMeasures
Label:     Basic Measures of Location and Variability
Template:  base.univariate.Measures
Path:     Univariate.age.BasicMeasures
-----

```

After identifying all the elements of the desired procedure via the ODS TRACE statement, you can select some or include all to be written to the many ODS destinations. One destination I use with frequency is the RTF destination because I can easily cut and paste into an existing MS WORD document that I use to explain results of many SAS® procedures. Wrapping the above code between the ODS RTF statements shown below will create a RTF file in the location specified.

```

ods rtf file='E:\project123\results_glm.rtf' style=journal startpage=no;
ods noptitle;
/*above univariate code or any SAS code*/
ods rtf close;

```

Steven LaLonde's paper from SAS Global Forum 2008 contains many helpful tricks to create simplified RTF files that help the cutting and pasting to a MS WORD document. Some of the options on the ODS RTF statement he uses include the StartPage=No option to suppress page breaks before each procedure and the ODS NoProctitle statement to prevent the writing of the procedure title.

EXAMPLE 1: REPORTING LEAST-SQUARES MEANS USING ODS AND PROC GLM

Using all the steps explained above, this example shows how to easily export a comparison of means from a GLM procedure. The objective in this study was to compare compliance rates for 6 different treatments. Comparing unadjusted means is easily done with PROC MEANS or PROC TTEST. But to get adjusted means requires the GLM procedure to control for believed confounding factors. The LSMEANS statement in PROC GLM computes the least-squares means and performs multiple comparisons of means when specifying certain options.

As explained earlier, using the ODS TRACE option lists the objects generated from the GLM procedure. In this case I want to output parameter estimates, adjusted means using least-squares, multiple comparisons of pairwise differences, and the p-values to show if differences between means are statistically significant. So with the use of the ODS OUTPUT statement, I can specify which elements of the PROC GLM I want to output to data sets. The code below outputs multiple objects to multiple data sets using the ODS OUTPUT statement.

```

ods rtf file='E:\project123\final_results_GLM.rtf' style=journal;

ods output ModelANOVA=a ParameterEstimates=b diff=c LSMeanCL=d LSMeanDiffCL=e;
proc glm data=sample order=formatted;
class idx_drug agegroup gender region;
model mpr_idxdrug = treatment agegroup gender region/solution;
lsmeans treatment /OM ADJUST=TUKEY PDIFF CL;
format idx_drug $drug.;
quit;

```

The ODS OUTPUT statement creates data sets containing output of each object (e.g., the ANOVA goes to data set "a"). The least-squares means are computed for the treatment variable listed in the LSMEANS statement. The ADJUST statement requests a multiple comparison adjustment using the Tukey-Kramer method. Other options in the

LSMEANS statement call for the p-values and confidence limits for the multiple comparisons of pairwise differences of the least-squares means. The FORMAT statement is used to format the treatment variable.

After getting the data sets, you can print the adjusted means and the difference between the adjusted means using the following code. The FORMAT statement formats the treatment variable to preferred names and mean values to a percent. The LABEL statement and LABEL option in the PROC PRINT lets you describe variables to your audience (as opposed to showing the variable names).

```
proc print data=d noobs label;
var idx_drug lowercl lsmean uppercl;
format idx_drug $drug. lowercl uppercl lsmean percent8.1;
label idx_drug='Treatment' LSMean='Adjusted Mean';
run;

proc print data=e noobs label;
var i j lowercl difference uppercl;
format i j numcharad. lowercl uppercl difference percent8.1;
label lowercl='Lower CL'
      uppercl='Upper CL'
      i='Treatment i' j='Treatment j'
      difference='Difference Between LS Means= LSMean(i)-LSMean(j)';
run;

ods rtf close;
```

The final result of the above code is shown below. As you can see, using the ODS RTF FILE statement places all the output into a RTF document that is ready to for cutting and pasting into a working document that explains the tables (i.e., describing confidence limits and recognizing differences that are significantly different).

	Treatment	LowerCL	Adjusted Mean	UpperCL
	A	31.6%	33.5%	35.5%
	B	24.2%	24.6%	25.0%
	C	28.5%	29.1%	29.8%
	D	29.3%	29.7%	30.2%
	E	16.0%	18.7%	21.4%
	F	30.5%	32.3%	34.1%

Treatment i	Treatment j	Lower CL	Difference Between LS Means= LSMean(i)-LSMean(j)	Upper CL
A	B	6.0%		11.8%
A	C	1.4%		7.4%
A	D	0.9%		6.7%
A	E	10.0%		19.8%
A	F	(2.6%)		5.1%
B	C	(5.7%)	(4.5%)	(3.4%)
B	D	(6.1%)	(5.1%)	(4.2%)
B	E	1.9%		10.0%
B	F	(10.4%)	(7.7%)	(5.0%)
C	D	(1.8%)	(0.6%)	0.6%
C	E	6.4%		14.5%
C	F	(6.0%)	(3.2%)	(0.4%)
D	E	7.0%		15.1%
D	F	(5.3%)	(2.6%)	0.1%
E	F	(18.4%)	(13.6%)	(8.9%)

USING ODS GRAPHICS IN 9.2

The ODS GRAPHICS release in SAS[®] 9.2 gives users the opportunity to create statistical graphics for over 25 procedures in SAS/STAT[®]. In this PROC GLM example, specifying the LSMEANS statement with the PDIF option produces a plot for the type of LS-means comparison. Using the PDIF=ANOM option produces an "analysis of means" plot, comparing each LS-mean to the average LS-mean. The resulting plot is shown below the code.

```
ods graphics on;

ods rtf file='C:\wuss2008_scott\glm_model_9.2.rtf' style=analysis;
proc glm data=id.sample order=formatted;
class treatment agegroup gender region;
model mpr_idxdrug = treatment agegroup gender region/solution;
lsmeans treatment / PDIF=ANOM CL;
format treatment $drug.;
quit;
ods rtf close;

ods graphics off;
```

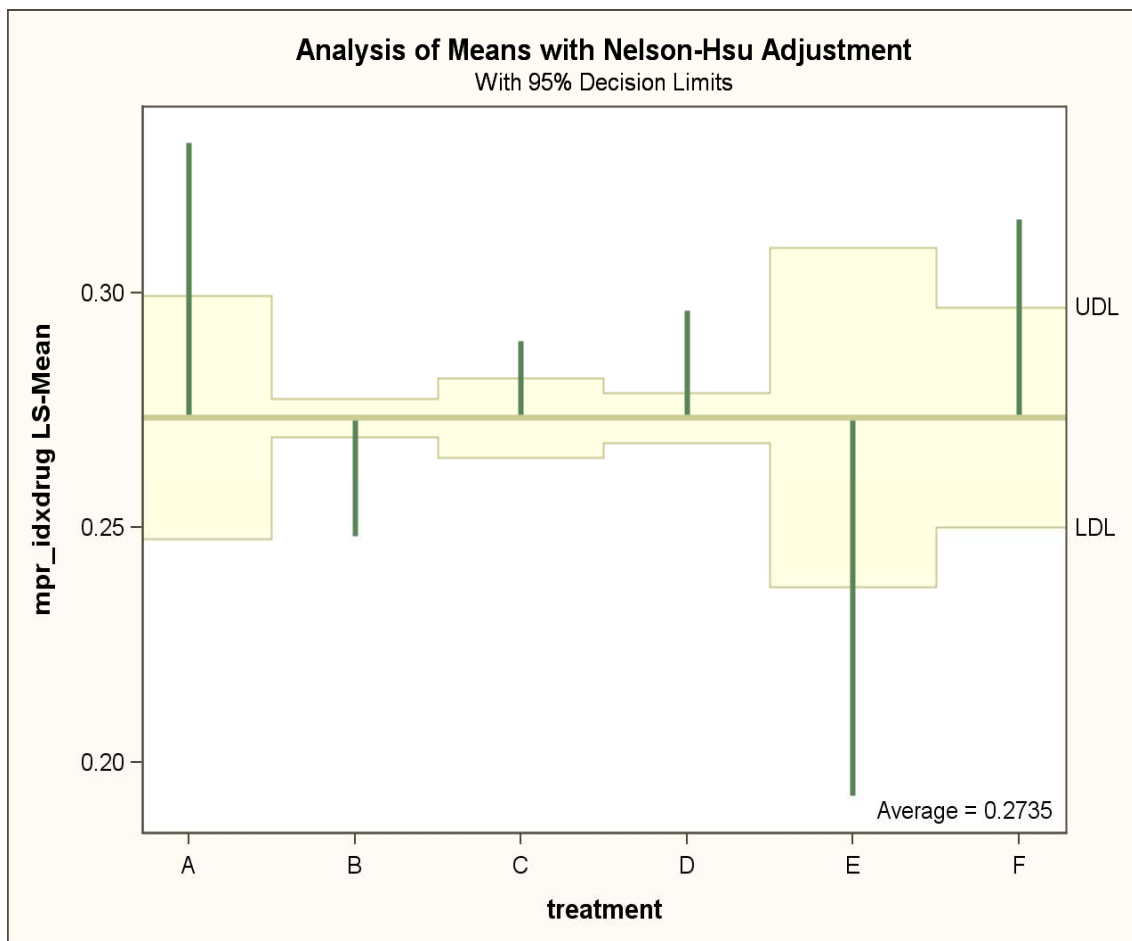


Figure 1. Plot of Analysis of Means with Nelson Hsu Adjustment

EXAMPLE 2: REPORTING ODDS RATIOS USING ODS AND PROC LOGISTIC

Odds ratios are a product of logistic regression that can be difficult to present and explain. Similar to the example above, you can use ODS to trace the objects of PROC LOGISTIC and rerun the procedure to output objects to data sets for easy formatting and printing. One of the desired objects of the LOGISTIC procedure is the odds ratios. In this example the objective was to compare probabilities of switching treatments among subjects.

In the code below, the ODS OUTPUT statement outputs the parameter estimates to a data set named "paramest" and the odds ratios to a data set named "or". The odds ratios are computed for the treatment variable listed in the MODEL statement. The EVENT option in the MODEL statement requests to model the probability of a switch in therapy. The REF option in the CLASS statement sets the reference category for each of the specified variables (e.g., compare the probability of switching for males vs. females). Other options in the MODEL statement call for the generalized R^2 for the fitted model and 95% Wald confidence limits for the odds ratios. The FORMAT statement is used to format the treatment variable.

```
ods rtf file='E:\project123\logistic_model.rtf' style=journal;

ods Trace On;

ods output ParameterEstimates=paramest OddsRatios=or CLOddsWald=clodds;
proc logistic data= sample;
class treatment(ref='A') gender(ref='F') region(ref='Central')
    agegroup(ref='Age Group 18-44') / param=ref ;
model swt_otherlao (event='1')= treatment agegroup gender region
    /rsquare clodds=wald;
format treatment $drug.;
run;

ods Trace Off;
```

After creating the data sets, you can print the parameter estimates and odds ratios using the following code.

```
proc print data=paramest noobs label;
label variable='Class' classval0='Value';run;
proc print data=or noobs label;run;

ods rtf close;
```

The final result of the above code are parameter estimates and odds ratios tables shown below.

Class	Value	DF	Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square
Intercept		1	-0.8828	0.0746	140.2071	<.0001
treatment	B	1	-0.9400	0.0729	166.0554	<.0001
treatment	C	1	-0.6490	0.0767	71.5625	<.0001
treatment	D	1	-0.4247	0.0729	33.9182	<.0001
treatment	E	1	0.4260	0.1154	13.6251	0.0002
treatment	F	1	-0.1181	0.0959	1.5148	0.2184
agegroup	Age Group 45-54	1	0.0132	0.0331	0.1593	0.6898
agegroup	Age Group 56-64	1	-0.1035	0.0359	8.3297	0.0039
agegroup	Age Group 65+	1	-0.0829	0.0328	6.3957	0.0114
gender	M	1	-0.0598	0.0243	6.0396	0.0140
region	East	1	-0.1076	0.0282	14.6027	0.0001
region	Mountain	1	-0.2088	0.0570	13.3956	0.0003
region	Pacific	1	-0.1161	0.0368	9.9564	0.0016
region	Southeast	1	-0.1161	0.0669	3.0171	0.0824

Effect	Odds Ratio Estimate	Lower 95% Confidence Limit for Odds Ratio	Upper 95% Confidence Limit for Odds Ratio
treatment B vs A	0.391	0.339	0.451
treatment C vs A	0.523	0.450	0.607
treatment D vs A	0.654	0.567	0.754
treatment E vs A	1.531	1.221	1.920
treatment F vs A	0.889	0.736	1.072
agegroup Age Group 45-54 vs Age Group 18-44	1.013	0.950	1.081
agegroup Age Group 56-64 vs Age Group 18-44	0.902	0.841	0.967
agegroup Age Group 65+ vs Age Group 18-44	0.920	0.863	0.982
gender M vs F	0.942	0.898	0.988
region East vs Central	0.898	0.850	0.949
region Mountain vs Central	0.812	0.726	0.908
region Pacific vs Central	0.890	0.828	0.957
region Southeast vs Central	0.890	0.781	1.015

USING ODS GRAPHICS IN 9.2

The LOGISTIC procedure has many graphs available in ODS GRAPHICS. The odds ratio plot is fantastic for turning a table of results into an easy to interpret diagram. The reader can instantly compare covariates and identify those point estimates that are statistically significant. The code below shows how adding the plots option to the LOGISTIC statement creates an odds ratio plot (Figure 2).

```
ods graphics on;

proc logistic data= id.sample plots(only)=(oddsratio effect);
class treatment (ref='A') gender (ref='F') region (ref='Central')
    agegroup (ref='Age Group 18-44') / param=ref;
model swt_otherlao (event='1')= treatment agegroup gender region /rsquare
    clodds = wald lackfit;
format treatment $drug.;
run;

ods graphics off;
```

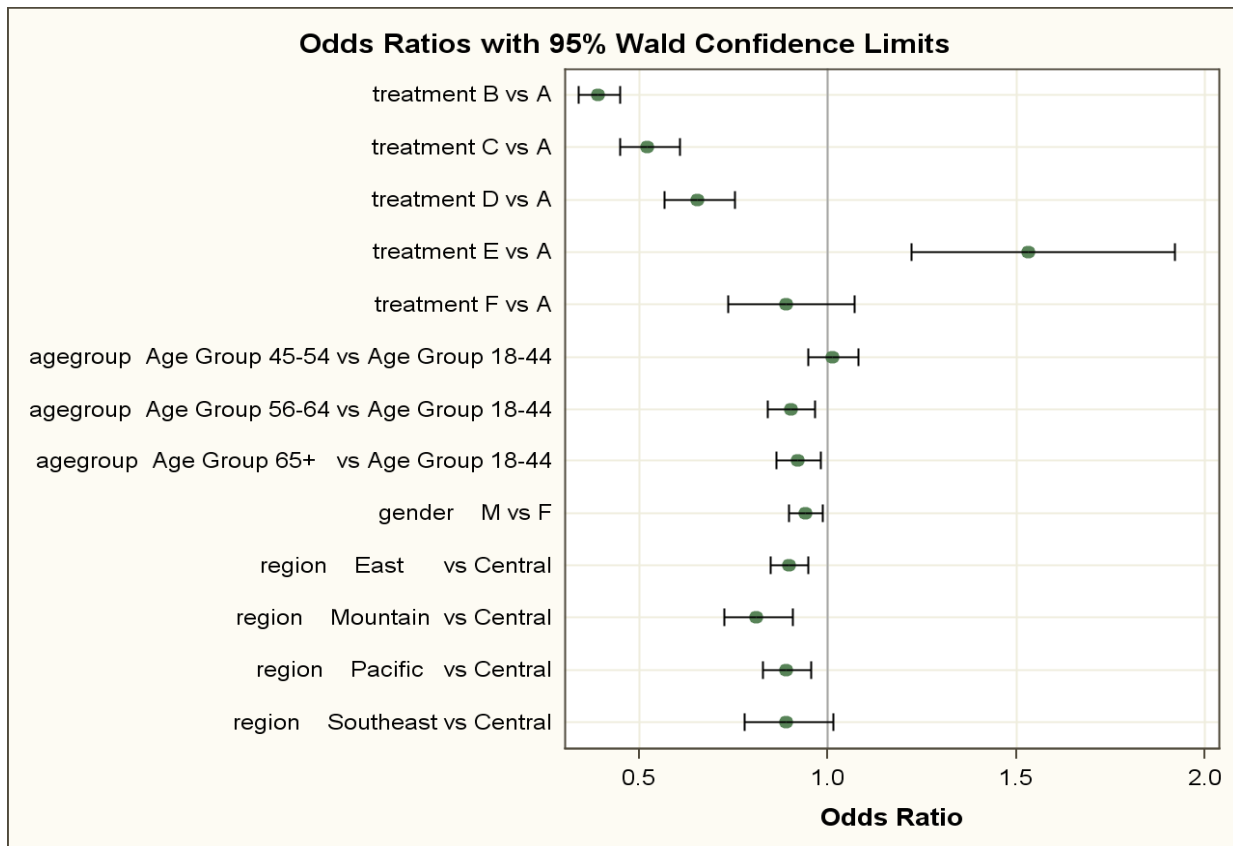


Figure 2. Odds Ratio Plot

CONCLUSION

Reporting results of statistical analyses can be equally difficult as conducting the analyses. Especially when explaining statistics to those not so familiar with statistics. Image can be almost everything. A nice visual presentation can help tremendously when explaining the results to your audience. ODS, ODS GRAPHICS, and the OUTPUT statement allows users to report statistics in a nice, presentable approach without spending too much time on manual formatting.

REFERENCES

- Cheng, Wei. 2006. "ODS Statistical Graphics 101", *Proceedings of the 2008 Western Users of SAS Software Conference*, Universal City, CA.
- LaLonde, Steven. 2008. "ODS for the Professional Statistician", *Proceedings of the 2008 SAS Global Forum*, San Antonio, TX. Paper 377-2008.
- Rodriguez, Robert N. 2004. "An Introduction to ODS for Statistical Graphics in SAS[®] 9.1", *Proceedings of the Twenty-Ninth Annual SAS[®] Users Group International Conference*, SAS Institute, Cary, NC, Paper 204-29.
- Rodriguez, Robert N., Balan, T. E. 2006. "Creating Statistical Graphics in SAS[®] 9.2: What Every Statistical User Should Know", *Proceedings of the Thirty-first Annual SAS Users Group International Conference*, Cary, NC: SAS Institute Inc.
- SAS Institute Inc. 2008. "Output Delivery System: User's Guide". *SAS OnlineDoc[®] 9.2*. Cary, NC: SAS Institute Inc. <http://support.sas.com/documentation/cdl/en/odsug/59523/HTML/default/odsugwhatsnew902.htm> (4/15/2008).

SAS Institute Inc. 2008. "SAS/STAT: User's Guide". *SAS OnlineDoc*[®] 9.2. Cary, NC: SAS Institute Inc. http://support.sas.com/documentation/cdl/en/statug/59654/HTML/default/chap0_toc.htm (4/15/2008).

SAS Institute Inc. 2008. "User's Guide: Statistical Graphics Using ODS". *SAS OnlineDoc*[®] 9.2. Cary, NC: SAS Institute Inc. <http://support.sas.com/documentation/cdl/en/statugstatgraph/61769/PDF/default/statugstatgraph.pdf> (4/15/2008).

Smith, Kevin D. 2007. "The Output Delivery System (ODS) from Scratch", *Proceedings of the 2007 SAS Global Forum*, Orlando, FL. Paper 219-2007.

RECOMMENDED READING

Rodriguez, Robert N., Balan, T. E. 2006. "Creating Statistical Graphics in SAS[®] 9.2: What Every Statistical User Should Know", *Proceedings of the Thirty-first Annual SAS Users Group International Conference*, Cary, NC: SAS Institute Inc.

CONTACT INFORMATION

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

Name: R. Scott Leslie
Enterprise: MedImpact Healthcare Systems, Inc.
Address: 10680 Treena Street
City, State ZIP: San Diego, CA 92131
Work Phone: 858-790-6685
Fax: 858-689-1799
E-mail: scott.leslie@medimpact.com
Web: www.medimpact.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.