

Paper 036-30
Reporting Statistical Results with PROC REPORT®
Nestor Rohowsky, IDCS Inc., La Grange, IL

Introduction

PROC REPORT® is a very flexible and user-friendly tool which can be applied to a wide range of data displays. In the biotech and pharmaceutical industries, PROC REPORT has been used for several years to generate production-quality data listings. This paper presents another use for this handy procedure, namely as a method for producing statistical summary tables, including p-values where appropriate. Examples will be given showing how easily a base table can be expanded or contracted, as the situation dictates.

Why PROC REPORT?

The traditional method used to produce customized statistical tables is to output the results using the DATA _NULL_ step. This allows for the maximum amount of control of placement of the various data items, including the results of performed statistical analyses, *i.e.* p-values. Titles, footnotes and horizontal and vertical lines to divide blocks of output, similar in concept to stock PROC TABULATE® output, can also be placed where needed.

Despite the obvious benefits of programming tables this way, there are drawbacks. First, one has to mock-up the appearance of the table and determine the location of each entry. Then one has to write program code so that the starting column numbers, lengths of entries, any special formatting for each entry, and column labels are determined. Then, if the output for any subset of the whole carries beyond the end of the physical page code has to be written such that important header information is carried over as well. Finally, there is the problem of modifying existing code in the case where the table grows or shrinks in the number of columns. An example is code written for an original clinical study where results for three treatments are presented being modified for a follow-up study where results for two (or four) treatments are to be displayed.

This can be done, but it requires an expenditure of time to mock-up the new table and modify existing code to account for the different number of columns in the new table. In companies with dedicated programmers who can write the code while the statistician simultaneously performs analyses and outputs p-values, this may not

be an issue. But in companies with no dedicated programmers, where one person acts as statistician and programmer, this may become a time and cost consideration. In this situation, using the built-in capabilities of PROC REPORT can help reduce the amount of extra time and associated cost spent programming, and at the same time generate quality data displays that closely resemble those produced using traditional methods.

Tables of counts and percents

In clinical trials work, summary displays of discrete variables are commonplace, both with and without p-values. Examples include distributions of patients by sex, race, shifts in efficacy or safety variables compared to baseline values (improved, no difference, worsened), or counts of events which happened during the study, such as incidence of treatment emergent adverse events. All of these tables are usually summarized using counts and percents. The last example is more difficult in that adverse events are summarized by overall presence of any event, by body system and individual adverse events within body system. When p-values are called for, they typically appear for all three levels of description.

To create a table of adverse events, it is necessary to first generate overall counts of patients, patients with at least one adverse event during treatment, patients with at least one adverse event in each body system, patients by adverse events within each body system. This can be done either using a series of data steps or storing the results of a series of procedure output such as from PROC FREQ®. Then the appropriate percents are calculated using the intermediate counts. Finally, desired p-values can be output for each appropriate level of description.

In order to get the most benefit from the proposed method, the programmer should name all the classes of variables the same, i.e. counts and percents involving body systems should all be called the same (e.g. syscnt, syspct), counts and percents of individual adverse events could be called advcnt and advpct, etc. If counts and percents are generated for several treatments, then the counts and percents could have a treatment prefix or suffix added.

Sample discrete variable output

INVESTIGATOR	INITIAL TREATMENT	WITH CS IMPROVEMENT		WITHOUT CS IMPROVEMENT		P-VALUE
		N	%	N	%	
SITE 1	2.5 MG	3	(21.4%)	6	(42.9%)	
	5.0 MG	3	(21.4%)	2	(14.3%)	
SITE 2	2.5 MG	7	(46.7%)	2	(13.3%)	
	5.0 MG	2	(13.3%)	4	(26.7%)	
COMBINED	2.5 MG	10	(34.5%)	8	(27.6%)	0.5769
	5.0 MG	5	(17.2%)	6	(20.7%)	

Cochran-Mantel-Haenszel test of differences between row means, stratifying by investigator.

Tables of continuous summary statistics

Other tables appearing in clinical reports describe findings of continuous variables. Examples include patient ages, vital signs, laboratory data and many efficacy variables. The displays typically show n, mean, std, min, median, and max for each treatment, possibly over multiple time points. Often there is at least an overall p-value comparing treatments as well as within-treatment p-values.

PROC REPORT may not seem like the best alternative to traditional table generation or even use of other procedures such as TABULATE. However, there are advantages to using PROC REPORT as the table generator, notably when data are presented in several blocks per column (e.g. baseline, visit and change from baseline).

Descriptive statistics can be generated using data steps or as output from other procedures (e.g. MEANS, UNIVARIATE, SUMMARY). P-values for the desired analyses are output from other procedures such as GLM, ANOVA, MIXED, and are renamed to correspond with the MEANS of the data. The combined output is then printed out in neat, ordered columns using PROC REPORT.

Sample continuous variable output

STUDY DAY	INVESTIGATOR	INITIAL DOSE GROUP				P-VALUES		
		2.5 MG bid	WITHIN-DOSE P-VALUE	5.0 MG bid	WITHIN-DOSE P-VALUE	COVARIATE	INVEST.	TREATMENT
35	SITE 1	N	13.0		7.0			
		MEAN	-4.8		-13.7			
		STD	18.2		22.4			
		MIN	-39.0		-38.0			
		MEDIAN	0.0		-16.0			
		MAX	18.0		13.0			
	SITE 2	N	13.0		11.0			
		MEAN	-23.2		-23.4			
		STD	21.7		25.1			
		MIN	-54.0		-69.0			
		MEDIAN	-19.0		-23.0			
		MAX	25.0		17.0			
	COMBINED SITES	N	26.0		18.0			
		MEAN	-14.0 (0.0030 **)		-19.6 (0.0028 **)	0.5172	0.0508	0.4230
		STD	21.8		23.9			
		MIN	-54.0		-69.0			
		MEDIAN	-14.0		-19.5			
		MAX	25.0		17.0			

Analysis of Covariance model with baseline scores as the covariate.
*, **, ***: Statistically significant with p<=0.05, 0.01, 0.001, respectively.

Sample code

Following is SAS code for generating the output shown above. It is a fairly complex table displaying descriptive statistics by treatment group and investigator and p-values for a variety of desired tests. This is intended to give an example of how to generate such a table using PROC REPORT, and is not to be construed as the best, or only, way to produce such a display.

***** Desired descriptive statistics are generated using other SAS procedures, such as MEANS, SUMMARY, TABULATE, or UNIVARIATE and output.

Statistical analyses are run using an appropriate procedure, e.g. GLM, and desired effect IDs and p-values are output.

The output datasets are combined and sorted and PROC REPORT code follows. *****

```
proc transpose data=contin out=table; * PROC TRANSPOSE required because of
                                         how summary data were obtained. This
                                         may not always be a necessary step;
by day invno;
  id trt;
var CHNG_n CHNG_mean CHNG_std CHNG_min CHNG_median CHNG_max;
run;
```

```

data table;
  set table;
    by day invno;
    if _NAME_='CHNG_N' then _NAME_='N';
    else if _NAME_='CHNG_Mean' then _NAME_='MEAN';
    else if _NAME_='CHNG_Std' then _NAME_='STD';
    else if _NAME_='CHNG_Min' then _NAME_='MIN';
    else if _NAME_='CHNG_Median' then _NAME_='MEDIAN';
    else if _NAME_='CHNG_Max' then _NAME_='MAX';
  run;

  options pageno=1;

  proc report data=table nowd headline split='*' spacing=4 missing;
    columns day invno _NAME_ ('_INITIAL DOSE GROUP_' _2D5 _2D6 _5 _5D1)
           ('_P-VALUES_' _55 _66 _77 _88);
    define day / order f=2. width=5 'STUDY*DAY' center;
    define invno / order f=inv. width=16 'INVESTIGATOR' right;
    define _NAME_ / order order=data f=$8. width=8 ' ' right;
    define _2D5 / display f=6.1 width=10 '2.5 MG*bid' center;
    define _2D6 / display f=with_p. /*f=within.*/ width=12 'WITHIN-DOSE*P-VALUE'
    left spacing=1;
    define _5 / display f=6.1 width=10 '5.0 MG*bid' center;
    define _5D1 / display f=with_p. /*f=within.*/ width=12 'WITHIN-DOSE*P-VALUE'
    left spacing=1;
    define _55 / display f=p_val. width=10 'COVARIATE' center;
    define _66 / display f=p_val. width=10 'INVEST.' center spacing=2;
    define _77 / display f=p_val. width=10 'TREATMENT' center spacing=2;
    define _88 / display f=p_val. width=10 'INV x TRT' center spacing=2;

    break after day / page;
    break after invno / skip;

*** add titles and footnotes as desired ***;

run;

```

References

Haworth LE, PROC TABULATE by Example, 1999 SAS Institute Inc., Cary NC

SAS® Guide to the REPORT Procedure, Release 6.11, 1995 SAS Institute Inc, Cary NC

SAS® Technical Report P-229, SAS/STAT Software: Changes and Enhancements, Release 6.07, 1992 SAS Institute Inc, Cary NC

Contact information: Nestor Rohowsky
IDCS Inc.
5718 Laurel Ave.
La Grange, IL 60525
Phone: (708) 784-0373
Fax: (708) 784-1619
e-mail: mrdata.idcsinc@earthlink.net

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.