Guido's Guide to PROC UNIVARIATE: A Tutorial for SAS® Users

Joseph J. Guido, University of Rochester Medical Center, Rochester, NY

ABSTRACT

PROC UNIVARIATE is a procedure within BASE SAS® used primarily for examining the distribution of data, including an assessment of normality and discovery of outliers. PROC UNIVARIATE goes beyond what PROC MEANS does and is useful in conducting some basic statistical analyses and includes high resolution graphical features. Output Delivery System (ODS) interface for this procedure will also be examined and demonstrated. This tutorial will cover both basic and intermediate uses of PROC UNIVARIATE including some helpful tips to expand the use of numeric type data and give a framework to build upon and extend knowledge of the SAS System.

INTRODUCTION

The Version 9.2 SAS® Procedure Manual states, "The UNIVARIATE procedure provides a variety of descriptive measures, high-resolution graphical displays, and statistical methods, which you can use to summarize, visualize, analyze, and model the statistical distributions of numeric variables. These tools are appropriate for a broad range of tasks and applications". Compare this with what the same manual states about PROC MEANS, "The MEANS procedure provides data summarization tools to computer descriptive statistics across all observations and within groups of observations. For example, PROC MEANS calculates descriptive statistics based on moments, estimates quantiles, which includes the median, calculates confidence limits for the mean, identifies extreme values and performs a t-test".

The following statements are used in PROC UNIVARIATE according to the SAS® 9.2 Procedure Manual:

PROC UNIVARIATE < options > **;**
CLASS variable(s) < / KEYLEVEL= value >**;**
VAR variable(s) **;**
BY variable(s) **;**
HISTOGRAM < variables > < / options > **;**
FREQ variable **;**
ID variable(s) **;**
INSET keyword-list < / options > **;**
OUTPUT < **OUT=**SAS-data-set > . . . < percentile-options >**;**
PROBPLOT < variable(s) > < / options > **;**
QQPLOT < variable(s) > < / options > **;**
WEIGHT variable **;**
RUN;

I have underlined the 5 statements in PROC UNIVARIATE which I will be discussing in this paper. The PROC UNIVARIATE statement is the only required statement for the UNIVARIATE procedure. It is only necessary to specify the following statements:

PROC UNIVARIATE;
RUN;

PROC UNIVARIATE will produce for each numeric variable in the last created dataset:

(1) Moments - N, Mean, Standard Deviation, Skewness, Uncorrected Sum of Squares, Coefficient of Variation, Sum of Weights, Sum of Observations, Variance, Kurtosis, Corrected Sum of Squares and Standard Error of the Mean

(2) Basic Statistical Measures - Mean, Median, Mode, Standard Deviation, Variance, Range and Interquartile Range

(3) Tests for Location Mu0=0 - Student's t, Sign and Signed Rand tests with their associated p-values

(4) Quantiles - 0%(Min), 1%, 5%, 10%, 25%(Q1), 50%(Median), 75%(Q3), 90%, 95%, 99% and 100%(Max)

(5) Extreme Observations - the five lowest and the five highest values

Some common options used in the PROC UNIVARIATE statement are DATA=, NORMAL, FREQ, PLOT. These options are explained as follows:

a) The **DATA=** option specifies which SAS Dataset to use in the PROC
b) The **NORMAL** option indicates a request for several tests of normality of variable(s)
c) The **FREQ** option produces a frequency table of the variable(s)
d) The **PLOT** option produces stem-and-leaf, box and qq plots of the variable(s)

DISCUSSION

I will use a SAS dataset called "choljg.sas7bdat" (a dataset that I created). In this fictitious dataset there are 100 patients. The variables are IDNUM (Identification Number), SEX (Sex of Patient), SITE (Clinic Site), AGE (Age of Patient), CHOL1 (Baseline Cholesterol mg/dl), CHOL2 (Follow-up Cholesterol mg/dl) and CHOLDIFF (Difference of CHOL1 – CHOL2). Let's begin with an analysis of all the numeric analytic variables in Mylib.choljg

Example 1

```
/*This is Example 1 which requests an analysis of all numeric variables in
Mylib.choljg*/
OPTIONS NODATE NONUMBER;
LIBNAME Mylib 'C:\Users\GVSUG\Desktop\Joseph Guido';
TITLE "Guido's Guide to PROC Univariate, NESUG 2009, Burlington, VT";
PROC UNIVARIATE DATA=Mylib.Choljg;
TITLE2 "Analysis of All Numeric Variables in Mylib.choljg";
   VAR Age Chol1 Chol2 Choldiff;
RUN;
```

Guido's Guide to PROC Univariate, NESUG 2009, Burlington, VT
Analysis of All Numeric Variables in Mylib.choljg

The UNIVARIATE Procedure
Variable: AGE (Age of Patient)

(1) Moments

N	100	Sum Weights	100
Mean	47.3	Sum Observations	4730
Std Deviation	18.6939735	Variance	349.464646
Skewness	0.21235988	Kurtosis	-1.0590988
Uncorrected SS	258326	Corrected SS	34597
Coeff Variation	39.5221428	Std Error Mean	1.86939735

(2) Basic Statistical Measures

Location		Variability	
Mean	47.30000	Std Deviation	18.69397
Median	45.00000	Variance	349.46465
Mode	33.00000	Range	68.00000
		Interquartile Range	31.50000

NOTE: The mode displayed is the smallest of 2 modes with a count of 6.

(3) Tests for Location: Mu0=0

Test	-Statistic-		-----p Value------	
Student's t	t	25.30227	Pr > |t|	<.0001
Sign	M	50	Pr >= |M|	<.0001
Signed Rank	S	2525	Pr >= |S|	<.0001

(4) Quantiles (Definition 5)

Quantile	Estimate
100% Max	86.0
99%	84.0
95%	77.5
90%	75.5
75% Q3	64.5
50% Median	45.0
25% Q1	33.0
10%	23.0
5%	19.5
1%	18.0
0% Min	18.0

```
Guido's Guide to PROC Univariate, NESUG 2009, Burlington, VT
     Analysis of All Numeric Variables in Mylib.choljg

                The UNIVARIATE Procedure
              Variable:  AGE  (Age of Patient)

                (5)   Extreme Observations

          ----Lowest----        ----Highest---

          Value      Obs        Value      Obs

             18       48           78       10
             18        1           80       42
             19       93           80       66
             19       83           82       50
             19       57           86       62
```

-- NOTE: Output for Chol1, Chol2 and Choldiff not shown to save space --

----------§§§§----------

If I want to look at the variable AGE grouped by the SEX variable, I can do this by using a CLASS statement. The dataset does not have to be sorted. The SAS syntax below for Example 2 will give us two reports. One will be all the moments, basic statistical measures, tests for location, quantiles and extreme observations for females. The other report will be the same five reports for the males.

Example 2

```
/*This is Example 2 which requests an analysis of Age grouped by Sex*/
LIBNAME Mylib 'C:\Users\GVSUG\Desktop\Joseph Guido';
TITLE "Guido's Guide to PROC Univariate, NESUG 2009, Burlington, VT";
PROC UNIVARIATE DATA=Mylib.Choljg;
TITLE2 "Analysis of Age grouped by Sex";
   CLASS Sex;
   VAR Age;
RUN;
```

Guido's Guide to PROC Univariate, NESUG 2009, Burlington, VT
Analysis of Age grouped by Sex

The UNIVARIATE Procedure
Variable: AGE (Age of Patient)
SEX = F

Moments

N	59	Sum Weights	59
Mean	50.6101695	Sum Observations	2986
Std Deviation	18.9017578	Variance	357.276447
Skewness	-0.0281795	Kurtosis	-1.0058182
Uncorrected SS	171844	Corrected SS	20722.0339
Coeff Variation	37.3477464	Std Error Mean	2.46079926

-- NOTE: Only Moments shown for SEX = F to save space

Guido's Guide to PROC Univariate, NESUG 2009, Burlington, VT
Analysis of Age grouped by Sex

The UNIVARIATE Procedure
Variable: AGE (Age of Patient)
SEX = M

Moments

N	41	Sum Weights	41
Mean	42.5365854	Sum Observations	1744
Std Deviation	17.5343913	Variance	307.454878
Skewness	0.57290985	Kurtosis	-0.7591314
Uncorrected SS	86482	Corrected SS	12298.1951
Coeff Variation	41.2219061	Std Error Mean	2.73841185

-- NOTE: Only Moments shown for SEX = M to save space

----------§§§§----------

Example 3

```
/*This is Example 3 which requests an analysis of Age sorted by Sex*/
LIBNAME Mylib 'C:\Users\GVSUG\Desktop\Joseph Guido';
TITLE "Guido's Guide to PROC Univariate, NESUG 2009, Burlington, VT";
/*First we must sort the dataset Mylib.choljg by Sex*/
PROC SORT DATA=Mylib.Choljg OUT=Choljg;
   BY Sex;
PROC UNIVARIATE DATA=Choljg PLOT;
TITLE2 "Analysis of Age by Sex";
TITLE3 "Plot Statement added for side by side Box Plot";
   BY Sex;
   VAR Age;
RUN;
```

Guido's Guide to PROC Univariate, NESUG 2009, Burlington, VT
Analysis of Age by Sex
Plot Statement added for side by side Box Plot

---------- Sex of Patient=F ----------

The UNIVARIATE Procedure
Variable: AGE (Age of Patient)

Moments

N	59	Sum Weights	59
Mean	50.6101695	Sum Observations	2986
Std Deviation	18.9017578	Variance	357.276447
Skewness	-0.0281795	Kurtosis	-1.0058182
Uncorrected SS	171844	Corrected SS	20722.0339
Coeff Variation	37.3477464	Std Error Mean	2.46079926

-- NOTE: Only Moments shown for SEX = F to save space

Guido's Guide to PROC Univariate, NESUG 2009, Burlington, VT
Analysis of Age by Sex
Plot Statement added for side by side Box Plot

---------- Sex of Patient=M ----------

The UNIVARIATE Procedure
Variable: AGE (Age of Patient)

Moments

N	41	Sum Weights	41
Mean	42.5365854	Sum Observations	1744
Std Deviation	17.5343913	Variance	307.454878
Skewness	0.57290985	Kurtosis	-0.7591314
Uncorrected SS	86482	Corrected SS	12298.1951
Coeff Variation	41.2219061	Std Error Mean	2.73841185

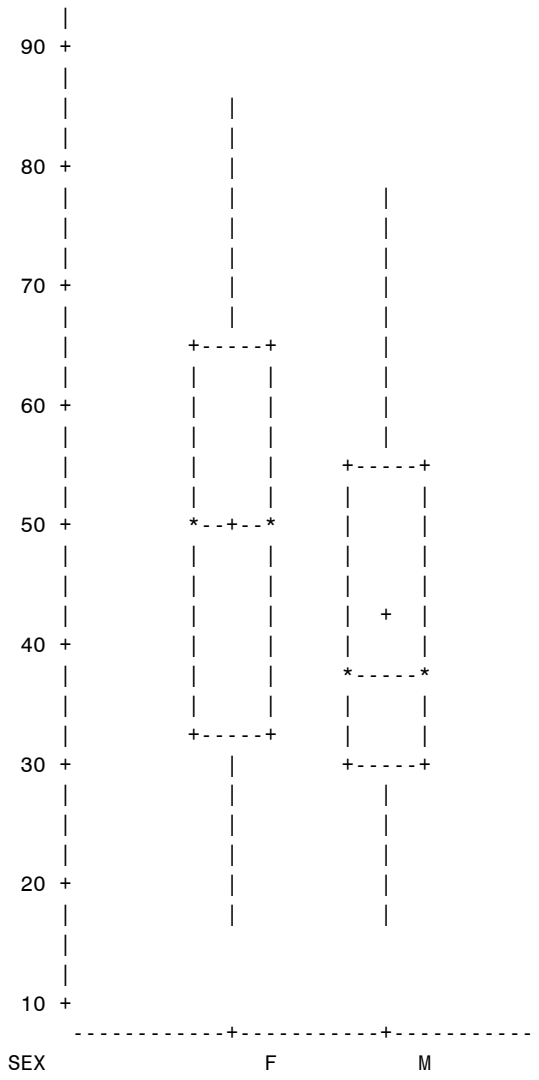
-- NOTE: Only Moments shown for SEX = M to save space

Here is the box plot of Age for the females and males. This was produced by using the keyword PLOT on the PROC UNIVARIATE Statement. The side-by-side plots are a result of the BY statement.

```
Guido's Guide to PROC Univariate, NESUG 2009, Burlington, VT
                    Analysis of Age by Sex
          Plot Statement added for side by side Box Plot

                       The UNIVARIATE Procedure
                    Variable:  AGE  (Age of Patient)

                            Schematic Plots
```



----------§§§§----------

Up until now, I have only been showing you the Moments section (first section) of PROC UNIVARIATE because I have wanted to save space. Now let's say that I wanted another section (of the five available) in PROC UNIVARIATE. How would I do this? Before I answer this question, let me just say something about output styles. I have been showing the regular SAS Listing Output up until now. If you wanted output such as HTML, RTF, PDF, etc you could use the Output Delivery System (ODS) in SAS. I refer you to an URL on the SAS website

http://support.sas.com/rnd/base/ods/index.html

This is a great resource for learning all about ODS up to and including SAS version 9.2
There are even quick tri-fold guides that you can print out and use as reference.

Let's suppose that I only want the Basic Statistical Measure for the variable CHOL1. It turns out that the Output Delivery System (ODS) in SAS can do this. In fact, not only does ODS allow for this in PROC UNIVARIATE, but it is virtually every PROC in SAS!

Well, since I don't know the name of the section of PROC UNIVARIATE that I want, I will use the ODS TRACE ON statement.

Example 4

SAS PROGRAM

```
LIBNAME Mylib 'C:\Users\GVSUG\Desktop\Joseph Guido';
ODS TRACE ON;
PROC UNIVARIATE Data=Mylib.Choljg;
   VAR Chol1;
RUN;
```

SAS LOG

```
1    LIBNAME Mylib 'C:\Users\GVSUG\Desktop\Joseph Guido';
NOTE: Libref MYLIB was successfully assigned as follows:
      Engine:        V9
      Physical Name: C:\Users\GVSUG\Desktop\Joseph Guido
2    ODS TRACE ON;
3    PROC UNIVARIATE Data=Mylib.Choljg;
4       VAR Chol1;
5    RUN;

NOTE: Writing HTML Body file: sashtml1.htm

Output Added:
-------------
Name:       Moments
Label:      Moments
Template:   base.univariate.Moments
Path:       Univariate.CHOL1.Moments
```

```
-------------
Output Added:
-------------
Name:       BasicMeasures
Label:      Basic Measures of Location and Variability
Template:   base.univariate.Measures
Path:       Univariate.CHOL1.BasicMeasures
-------------

Output Added:
-------------
Name:       TestsForLocation
Label:      Tests For Location
Template:   base.univariate.Location
Path:       Univariate.CHOL1.TestsForLocation
-------------

Output Added:
-------------
Name:       Quantiles
Label:      Quantiles
Template:   base.univariate.Quantiles
Path:       Univariate.CHOL1.Quantiles
-------------

Output Added:
-------------
Name:       ExtremeObs
Label:      Extreme Observations
Template:   base.univariate.ExtObs
Path:       Univariate.CHOL1.ExtremeObs
-------------
NOTE: PROCEDURE UNIVARIATE used (Total process time):
      real time           0.06 seconds
      cpu time            0.04 seconds
```

Since I want the Basic Measures, I will use the ODS SELECT statement in my next example to indicate how to accomplish this in SAS.

----------§§§§----------

Example 5

```
/*This is Example 5 which requests Basic Measures*/
LIBNAME Mylib 'C:\Users\GVSUG\Desktop\Joseph Guido';
ODS SELECT BASICMEASURES;
TITLE "Guido's Guide to PROC Univariate, NESUG 2009, Burlington, VT";
PROC UNIVARIATE Data=Mylib.Choljg;
TITLE2 "Basic Measures selected";
   VAR Chol1;
RUN;
/*Other Selections: Moments, TestsforLocation, Quantiles, ExtremeObs*/
```

```
                    Guido's Guide to PROC Univariate, NESUG 2009, Burlington, VT
                                    Basic Measures selected

                                    The UNIVARIATE Procedure
                       Variable:  CHOL1  (Baseline Cholesterol (mg/dl))

                                  Basic Statistical Measures

                           Location                      Variability

                    Mean     227.2700     Std Deviation           46.42510
                    Median   223.0000     Variance                    2155
                    Mode     189.0000     Range                  216.00000
                                          Interquartile Range     71.00000

                NOTE: The mode displayed is the smallest of 2 modes with a count of 4.
```

OK, now that I've shown you that you can select any of the five standard sections of PROC UNIVARIATE, what if you want more. For example, let's say that you want more than just the 5 extreme values which you get by default. In other words, PROC UNIVARIATE gives the 5 highest and 5 lowest values of a given variable. What if we want the 10 highest and 10 lowest values? SAS can do that!

----------§§§§----------

Example 6

```
/* Example 6 Select the 10 lowest and 10 highest values
    of Cholesterol Difference */
LIBNAME Mylib 'C:\Users\GVSUG\Desktop\Joseph Guido';
ODS SELECT EXTREMEVALUES;
TITLE "Guido's Guide to PROC Univariate, NESUG 2009, Burlington, VT";
PROC UNIVARIATE Data=Mylib.Choljg NEXTRVAL=10;
TITLE2 "10 lowest and 10 highest values of Cholesterol Difference";
VAR Choldiff;
RUN;
```

Guido's Guide to PROC Univariate, NESUG 2009, Burlington, VT
10 lowest and 10 highest values of Cholesterol Difference

The UNIVARIATE Procedure
Variable: CHOLDIFF (Difference of CHOL1 - CHOL2)

Extreme Values

---------Lowest--------			--------Highest--------		
Order	Value	Freq	Order	Value	Freq
1	-42	1	48	45	1
2	-40	1	49	50	1
3	-33	2	50	56	1
4	-31	1	51	57	1
5	-24	2	52	60	3
6	-21	1	53	64	1
7	-18	1	54	73	1
8	-13	1	55	78	1
9	-11	1	56	87	1
10	-10	1	57	98	1

I mentioned the 5 basic ODS table name (Moments, BasicMeasures, TestsforLocations, Quantiles and ExtremeObs) and one optional one (ExtremeValues) with the Nextrvals option. The table below contains those plus all table names produced by the PROC UNIVARIATE statement. These can be found at the following URL:
http://support.sas.com/documentation/cdl/en/procstat/59629/HTML/default/procstat_univariate_sect051.htm

(table continues on next page)

Table 4.93 ODS Tables Produced with the PROC UNIVARIATE Statement

ODS Table Name	Description	Option
BasicIntervals	confidence intervals for mean, standard deviation, variance	CIBASIC
BasicMeasures	measures of location and variability	default
ExtremeObs	extreme observations	default
ExtremeValues	extreme values	NEXTRVAL=
Frequencies	frequencies	FREQ
LocationCounts	counts used for sign test and signed rank test	LOCCOUNT
MissingValues	missing values	default, if missing values exist

Modes	modes	MODES
Moments	sample moments	default
Plots	line printer plots	PLOTS
Quantiles	quantiles	default
RobustScale	robust measures of scale	ROBUSTSCALE
SSPlots	line printer side-by-side box plots	PLOTS (with BY statement)
TestsForLocation	tests for location	default
TestsForNormality	tests for normality	NORMALTEST
TrimmedMeans	trimmed means	TRIMMED=
WinsorizedMeans	Winsorized means	WINSORIZED=

You may have noticed that from the above table we have covered the default options as well as the nextrval= and plots. Using these as examples, the user now has a guide to explore other ODS Table Names as needed.

I will demonstrate one more option as I have found in teaching my applied statistics course using SAS has been of great help to students. The question concerns whether the data are normally distributed or not.

----------§§§§----------

Example 7

```
/* Example 7 -  Are my analysis variables normally distributed? */
LIBNAME Mylib 'C:\Users\GVSUG\Desktop\Joseph Guido';
ODS SELECT TESTSFORNORMALITY;
TITLE "Guido's Guide to PROC Univariate, NESUG 2009, Burlington, VT";
PROC UNIVARIATE Data=Mylib.Choljg NORMALTEST;
TITLE2 "Are my analysis variables normally distributed?";
VAR Age Chol1 Chol2 Choldiff;
RUN;
```

There are several ways to determine this from the output found on the next page. To test for normality you must add the keyword **normal (or normaltest)** to the PROC UNIVARIATE statement. PROC UNIVARIATE displays two statistics in the Tests for Normality section. These are W and Pr<W. The values of W it will show range from 0.0 <= value <=1.00 . Values closer to 1 indicate a higher degree of normality. SAS uses the *Shapiro-Wilk* test when the sample size is below 2000 and the *Kolmogorov-Smirnov* test for sample size above 2000. Additionally, SAS uses the *Cramer-von Mises* and *Anderson-Darling* tests. You may want to consult a statistical text book such as Testing for Normality by Henry C. Thode, Jr. for more detail.

We can see from the output on the next page that all of our analysis variables are not normally distributed with Choldiff being the closest to normal (W = .973 and Pr<W = 0.04) of all four variables.

Guido's Guide to PROC Univariate, NESUG 2009, Burlington, VT
Are my analysis variables normally distributed?
The UNIVARIATE Procedure
Variable: AGE (Age of Patient)

Tests for Normality

Test		--Statistic---		-----p Value------
Shapiro-Wilk	W	0.955237	Pr < W	0.0019
Kolmogorov-Smirnov	D	0.0916	Pr > D	0.0384
Cramer-von Mises	W-Sq	0.168549	Pr > W-Sq	0.0143
Anderson-Darling	A-Sq	1.187565	Pr > A-Sq	<0.0050

Guido's Guide to PROC Univariate, NESUG 2009, Burlington, VT
Are my analysis variables normally distributed?
The UNIVARIATE Procedure
Variable: CHOL1 (Baseline Cholesterol (mg/dl))

Tests for Normality

Test		--Statistic---		-----p Value------
Shapiro-Wilk	W	0.963579	Pr < W	0.0073
Kolmogorov-Smirnov	D	0.094255	Pr > D	0.0272
Cramer-von Mises	W-Sq	0.183726	Pr > W-Sq	0.0086
Anderson-Darling	A-Sq	1.119607	Pr > A-Sq	0.0062

Guido's Guide to PROC Univariate, NESUG 2009, Burlington, VT
Are my analysis variables normally distributed?
The UNIVARIATE Procedure
Variable: CHOL2 (Follow-up Cholesterol (mg/dl))

Tests for Normality

Test		--Statistic---		-----p Value------
Shapiro-Wilk	W	0.964597	Pr < W	0.0087
Kolmogorov-Smirnov	D	0.07366	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.132373	Pr > W-Sq	0.0420
Anderson-Darling	A-Sq	0.890838	Pr > A-Sq	0.0227

Guido's Guide to PROC Univariate, NESUG 2009, Burlington, VT
Are my analysis variables normally distributed?
The UNIVARIATE Procedure
Variable: CHOLDIFF (Difference of CHOL1 - CHOL2)

Tests for Normality

Test		--Statistic---		-----p Value------
Shapiro-Wilk	W	0.973285	Pr < W	0.0396
Kolmogorov-Smirnov	D	0.078563	Pr > D	0.1312
Cramer-von Mises	W-Sq	0.180617	Pr > W-Sq	0.0092
Anderson-Darling	A-Sq	1.05766	Pr > A-Sq	0.0088

Before we finish with our publication quality or high-resolution graphics, let's take a quick look at whether or not there was a statistically significant difference (lowering) or cholesterol from baseline to follow-up in the subjects in either of our two study sites.

----------§§§§----------

Example 8

```
/*Example 8-Is there a statistically significant difference in Cholesterol
between Site 1 and Site 2?*/
LIBNAME Mylib 'C:\Users\GVSUG\Desktop\Joseph Guido';
ODS SELECT TESTSFORLOCATION;
TITLE "Guido's Guide to PROC Univariate, NESUG 2009, Burlington, VT";
PROC UNIVARIATE Data=Mylib.Choljg;
TITLE2 "Is there a statistically significant difference in Cholesterol";
TITLE3 "between Site 1 and Site 2?";
VAR Choldiff;
RUN;
```

Guido's Guide to PROC Univariate, NESUG 2009, Burlington, VT
Is there a statistically significant difference in Cholesterol
among Site 1 and Site 2?

The UNIVARIATE Procedure
Variable: CHOLDIFF (Difference of CHOL1 - CHOL2)
SITE = 1

Tests for Location: MuO=0

Test	-Statistic-		-----p Value------	
Student's t	t	7.239629	Pr > |t|	<.0001
Sign	M	19	Pr >= |M|	<.0001
Signed Rank	S	535	Pr >= |S|	<.0001

Guido's Guide to PROC Univariate, NESUG 2009, Burlington, VT
Is there a statistically significant difference in Cholesterol
among Site 1 and Site 2?

The UNIVARIATE Procedure
Variable: CHOLDIFF (Difference of CHOL1 - CHOL2)
SITE = 2

Tests for Location: MuO=0

Test	-Statistic-		-----p Value------	
Student's t	t	5.204728	Pr > |t|	<.0001
Sign	M	17	Pr >= |M|	<.0001
Signed Rank	S	401.5	Pr >= |S|	<.0001

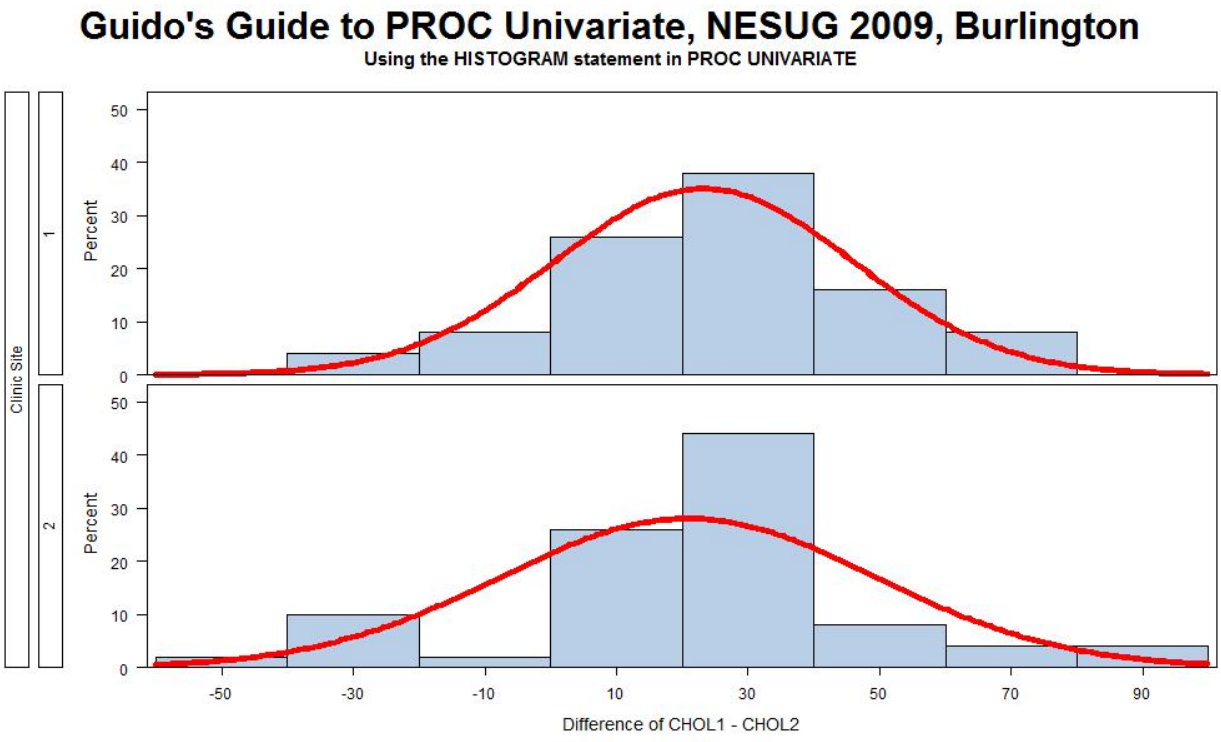
From the Student's t statistic we can see that we had a significant lowering in the subjects from Site 1 and the subjects from Site 2 and so our intervention was successful.

For our last examples we will now consider producing some high resolution or publication quality graphics using the HISTOGRAM statement and ODS.

----------§§§§----------

Example 9a

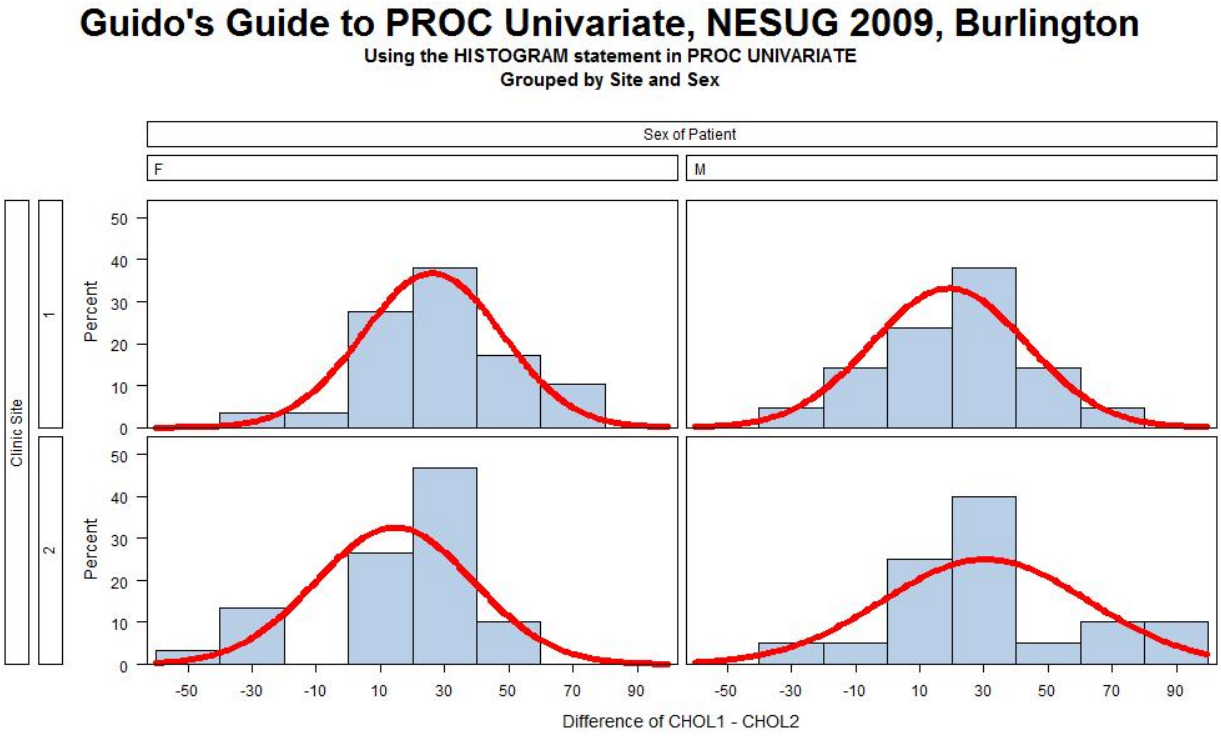
```
/*Example 9a - Using the HISTOGRAM statement in PROC UNIVARIATE */
LIBNAME Mylib 'C:\Users\GVSUG\Desktop\Joseph Guido';
TITLE "Guido's Guide to PROC Univariate, NESUG 2009, Burlington";
PROC UNIVARIATE Data=Mylib.Choljg NOPRINT;
CLASS Site;
TITLE2 "Using the HISTOGRAM statement in PROC UNIVARIATE";
VAR Choldiff;
HISTOGRAM / NORMAL (COLOR=RED W=5) NROWS=2;
RUN;
```



----------§§§§----------

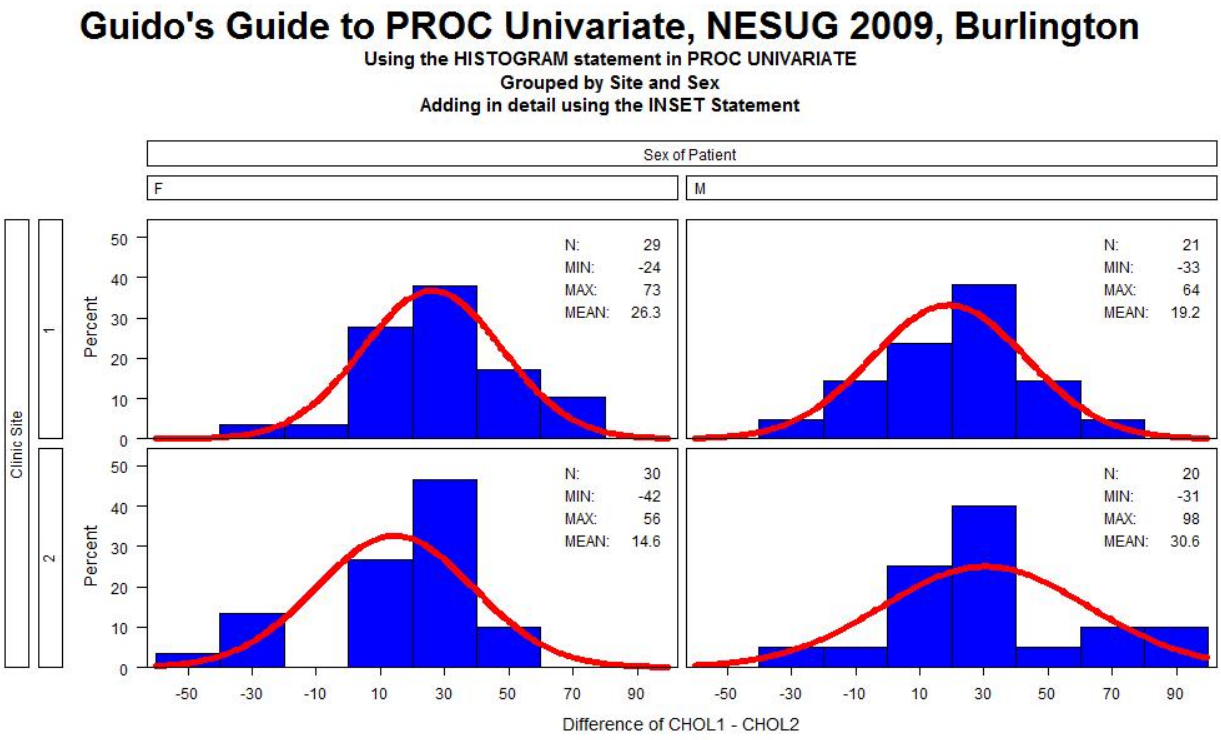
Example 9b

```
/*Example 9b - Using the HISTOGRAM statement in PROC UNIVARIATE */
LIBNAME Mylib 'C:\Users\GVSUG\Desktop\Joseph Guido';
TITLE "Guido's Guide to PROC Univariate, NESUG 2009, Burlington";
PROC UNIVARIATE Data=Mylib.Choljg NOPRINT;
CLASS Site Sex;
TITLE2 "Using the HISTOGRAM statement in PROC UNIVARIATE";
TITLE3 "Grouped by Site and Sex";
VAR Choldiff;
HISTOGRAM / NORMAL (COLOR=RED W=5) NROWS=2 NCOLS=2;
RUN;
```



----------§§§§----------

```
/*Example 9c - Using the HISTOGRAM statement in PROC UNIVARIATE */
LIBNAME Mylib 'C:\Users\GVSUG\Desktop\Joseph Guido';
TITLE "Guido's Guide to PROC Univariate, NESUG 2009, Burlington";
PROC UNIVARIATE Data=Mylib.Choljg NOPRINT;
CLASS Site Sex;
TITLE2 "Using the HISTOGRAM statement in PROC UNIVARIATE";
TITLE3 "Grouped by Site and Sex";
TITLE4 "Adding in detail using the INSET Statement";
VAR Choldiff;
HISTOGRAM / NORMAL (COLOR=RED W=5) NROWS=2 NCOLS=2 CFILL=BLUE;
INSET N='N:' (4.0) MIN='MIN:' (4.0) MAX='MAX:' (4.0) MEAN='MEAN:' (4.1)
                / NOFRAME POSITION=NE HEIGHT=2;
RUN;
```



CONCLUSION

PROC UNIVARIATE is a BASE SAS procedure which goes beyond the functionality of PROC MEANS. This procedure is extremely useful for examination of distributions of analysis variables and the production of high resolution graphics. This tutorial has just scratched the surface of the power of PROC UNIVARIATE and the author's hope is that from these simple examples that the SAS user will use it as a guide to extend their knowledge of PROC UNIVARIATE and experiment with other uses for this very versatile procedure.

REFERENCES

SAS Institute Inc. (2009) *Base SAS® 9.2 Procedures Guide*, Cary, NC: SAS Institute, Inc.

Guido, Joseph J. (2008) "Guido's Guide to PROC MEANS – A Tutorial for Beginners Using the SAS® System", Proceedings of the 21st annual North East SAS Users Group Conference, Pittsburgh, PA, 2008, paper #FF06.

Thode, Jr., Henry C. (2002) *Testing for Normality*, New York, Marcel Dekker

ACKNOWLEDGEMENTS

CONTACT INFORMATION

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

Joseph J Guido, MS
University of Rochester Medical Center
Department of Community and Preventive Medicine
Division of Social and Behavioral Medicine
120 Corporate Woods, Suite 350
Rochester, New York 14623
Phone: (585) 758-7818
Fax: (585) 424-1469
Email: Joseph_Guido@urmc.rochester.edu