Obtaining Summary Information from a Data Set

Two procedures that can generate summary information for n specified variables are the UNIVARIATE procedure and the MEANS procedure. The syntaxes for the two procedures are quite similar.

The UNIVARIATE Procedure

PROC UNIVARIATE DATA=*tablename* <options>*;*
BY *variable list*; *(optional statement)*
WHERE *list condition* ; *(optional statement, see DATA Step Information 2)*
CLASS *variable list;* *(optional statement)*
VAR *variable1 variable2 ... variablen;*
HISTOGRAM *variablelist* </options> ;
OUTPUT OUT= *newtablename* <list options and new variable names> ;
RUN;

PROC UNIVARIATE statement

Some of the possible options for the UNIVARIATE procedure are:

ALPHA=n	specifies the default Type I error probability for the confidence intervals. Select n such that $0 < n < 0.5$. The default setting is 0.05.
CIBASIC	requests confidence intervals for the mean, variance, and standard deviation. Confidence level is set by the ALPHA option.
FREQ	requests a frequency table consisting of variable value, frequencies, percentages, and cumulative percentages
MU0=	specify the null hypothesis value for the test of the location parameter; that is H_0: $\mu = \mu_0$, MU0= value specifies μ_0
NORMAL	computes four test statistics and their significance levels for the test of the null hypothesis that the input data come from a normal distribution versus the alternative hypothesis that the distribution is not normal.
PLOT	produces a stem-and-leaf plot, a box plot, and a normal probability plot. If a BY statement is used, side-by-side box plots labeled Schematic Plots appear for groups defined by the BY variables.
PLOTSIZE=	specify the number of rows the plots should take. Sometimes, plots can appear "scrunched." The PLOTSIZE option can be used to address this.

This list of options is not complete. For a complete list of options consult the SAS Online Documentation. It is not necessary to specify any options in the UNIVARIATE procedure. Several summary statistics are computed when no options are specified.

BY statement

The UNIVARIATE procedure can operate on subgroups of the data set defined by the values of another variable. These subgroups are identified using the BY statement. This statement is optional. Data must first be sorted BY the variables (in the same order) as those appearing in a BY statement.

CLASS Statement

The CLASS statement is also an optional statement. The CLASS statement assigns a variable or variables to form subgroups. CLASS variables can be either numeric or character. The CLASS statement has an effect similar to a BY statement. The CLASS statement does NOT require the data to be sorted first.

VAR Statement

All of the numeric variables for which summary statistics are to be calculated are specified in the VAR statement. If you omit the VAR statement, the UNIVARIATE procedure will compute summary statistics for every numeric variable in your data table.

HISTOGRAM Statement

Specify the variables for which histograms are to be produced. These variables can be a smaller subset of those variables identified in the VAR statement. Histograms are available as ODS Graphics. You must have ODS Graphics enabled to view this image. HISTOGRAM statement options include:

NORMAL this option will overlay a normal curve on the histogram. There are some advanced options controlling colors and placement of the normal curve. Other distributions are also available. See SAS Help and Documentation for more information.

MIDPOINTS = numberlist SAS/UNIVARIATE will select the midpoints at which bars of the histogram are to be centered. List the numbers you wish for midpoints if you want to control the number and placement. Use no commas to separate the values of the midpoints. For example: MIDPOINTS = 1.2 2.4 3.6 4.8 6.0 ;

OUTPUT Statement

Some of the information produced by the UNIVARIATE procedure can be output to a new data set. This is another way to create a data table. The new data table can then be used in other SAS procedures. The syntax for the OUTPUT statement identifies a new data table name, and the statistics one wishes to recover for each variable in the VAR statement. Note the number and order of the variables in the VAR statement determine the number and order of variables to be newly created for each requested statistic. The syntax is

VAR *variable1 variable2 . . . variablen* ;
OUTPUT OUT=*newdatatablename* statistic1 = *stat1var1 stat1var2 . . stat1varn*
statistic2 = *stat2var1 stat2var2 . . . stat2varn*
:
statisticm = *statmvar1 statmvar2 . . . statmvarn* ;

where statistic# can be any of the following:

N	the number of nonmissing values
NMISS	the number of observations having missing values
NOBS	the number of observations
MIN	the minimum value
MAX	the maximum value
RANGE	the range
SUM	the sum
MEAN	the mean

VAR	the variance
STD	the standard deviation
STDMEAN	the standard error of the mean
CV	coefficient of variation
MEDIAN	the median
Q3	the upper quartile or 75th percentile
Q1	the lower quartile or 25th percentile
P1	the 1st percentile
P5	the 5th percentile
P10	the 10th percentile
P90	the 90th percentile
P95	the 95th percentile
P99	the 99th percentile
MODE	the mode
T	Student's t value for testing the hypothesis that the population mean is 0
PROBT	probability of a greater absolute value for the Student's t value (observed significance level of a two-tailed test)

The following data are annualized returns on two groups of stocks.

Group A: 12.5 13 14.8 11 16.7 9 8.3 -1.2
Group B: 3.9 15.5 16.2 18 11.6 10 9.5

Objective 1: Using the UNIVARIATE procedure test whether or not the combined data are normally distributed and compute summary statistics. Experiment with the PLOT and PLOTSIZE options on the PROC UNIVARIATE statement also.

```
DM 'LOG; CLEAR; ODSRESULTS; CLEAR; ';

DATA one;
INPUT  group $ returns  @@;
DATALINES;
A 12.5 A 13 A 14.8 A 11 A 16.7 A 9 A 8.3 A -1.2
B 3.9 B 15.5 B 16.2 B 18 B 11.6 B 10 B 9.5
;
PROC UNIVARIATE DATA=one  NORMAL;
VAR returns;
TITLE 'Objective 1';
RUN;

QUIT;
```

You may also wish to examine the results when you include ODS LISTING; at the top of the program. This will enable the Listing Output (in the Output window). ODS LISTING CLOSE ; disables the Listing Output.

Note these are global statements. Once you close the Listing Output and/or the HTML Output they remain closed during the SAS session until you enable them, and vice versa. When enabling ODS Graphics, it is recommended to close the Listing Output at the top of the program. That is, the top lines would include

```
ODS HTML ;
ODS GRAPHICS ;
ODS LISTING CLOSE ;
```

Objective 2: Using the CLASS statement, compute the summary statistics, 99% confidence intervals, and tests for normality for each group and produce a HISTOGRAM with a normal curve for each group. Make certain that ODS Graphics are enabled to view the histogram.

```
PROC UNIVARIATE DATA=one NORMAL CIBASIC ALPHA=0.01;
CLASS group ;
VAR returns;
HISTOGRAM returns / NORMAL ;
TITLE 'Objective 2';
RUN;
```

Note: If you position the Objective 2 code after ODS GRAPHICS OFF; you will not obtain any HTML Output or graphics as you have closed those "communications" prior to running this UNIVARIATE code.

Objective 3: Modify the program in Objective 2. From the UNIVARIATE procedure recover the means, sample sizes and standard errors of the means in a new data table. Use the PRINT procedure to view the contents of the new data table.

```
PROC UNIVARIATE DATA=one ;
CLASS group ;
VAR returns;
HISTOGRAM returns / NORMAL ;
OUTPUT OUT=two MEAN=return_mean N=return_size STDMEAN=return_se ;
TITLE 'Objective 3';
PROC PRINT DATA=two ;
RUN;
```

Objective 2 and 3

The UNIVARIATE Procedure
Variable: returns
group = A

Moments			
N	8	**Sum Weights**	8
Mean	10.5125	**Sum Observations**	84.1
Std Deviation	5.49556639	**Variance**	30.20125
Skewness	-1.4712422	**Kurtosis**	2.88318709
Uncorrected SS	1095.51	**Corrected SS**	211.40875
Coeff Variation	52.2764936	**Std Error Mean**	1.94297613

Above and below are default output tables.

Basic Statistical Measures			
Location		**Variability**	
Mean	10.51250	**Std Deviation**	5.49557
Median	11.75000	**Variance**	30.20125
Mode	.	**Range**	17.90000
		Interquartile Range	5.25000

Produced by the CIBASIC ALPHA=0.01 options.

Basic Confidence Limits Assuming Normality			
Parameter	**Estimate**	**99% Confidence Limits**	
Mean	10.51250	3.71309	17.31191
Std Deviation	5.49557	3.22888	14.61865
Variance	30.20125	10.42566	213.70486

Default table

Tests for Location: Mu0=0				
Test		**Statistic**	**p Value**	
Student's t	**t**	5.410514	**Pr > |t|**	0.0010
Sign	**M**	3	**Pr >= |M|**	0.0703
Signed Rank	**S**	17	**Pr >= |S|**	0.0156

Produced by the NORMAL option on the PROC UNIVARIATE statement.

Tests for Normality				
Test		**Statistic**	**p Value**	
Shapiro-Wilk	**W**	0.884145	**Pr < W**	0.2062
Kolmogorov-Smirnov	**D**	0.218622	**Pr > D**	>0.1500
Cramer-von Mises	**W-Sq**	0.064568	**Pr > W-Sq**	>0.2500
Anderson-Darling	**A-Sq**	0.433707	**Pr > A-Sq**	0.2284

Default table

Quantiles (Definition 5)	
Quantile	**Estimate**
100% Max	16.70
99%	16.70
95%	16.70
90%	16.70
75% Q3	13.90
50% Median	11.75
25% Q1	8.65
10%	-1.20
5%	-1.20
1%	-1.20
0% Min	-1.20

Default table

Extreme Observations			
Lowest		**Highest**	
Value	**Obs**	**Value**	**Obs**
-1.2	8	11.0	4
8.3	7	12.5	1
9.0	6	13.0	2
11.0	4	14.8	3
12.5	1	16.7	5

Objective 2 and 3

The UNIVARIATE Procedure
Variable: returns
group = B

Moments			
N	7	**Sum Weights**	7
Mean	12.1	**Sum Observations**	84.7
Std Deviation	4.86209831	**Variance**	23.64
Skewness	-0.5374056	**Kurtosis**	-0.2106967
Uncorrected SS	1166.71	**Corrected SS**	141.84
Coeff Variation	40.1826307	**Std Error Mean**	1.83770043

Basic Statistical Measures			
Location		**Variability**	
Mean	12.10000	**Std Deviation**	4.86210
Median	11.60000	**Variance**	23.64000
Mode	.	**Range**	14.10000
		Interquartile Range	6.70000

Basic Confidence Limits Assuming Normality			
Parameter	**Estimate**	**99% Confidence Limits**	
Mean	12.10000	5.28686	18.91314
Std Deviation	4.86210	2.76539	14.48818
Variance	23.64000	7.64736	209.90732

Tests for Location: Mu0=0				
Test		**Statistic**		**p Value**
Student's t	**t**	6.584316	**Pr > |t|**	0.0006
Sign	**M**	3.5	**Pr >= |M|**	0.0156
Signed Rank	**S**	14	**Pr >= |S|**	0.0156

Tests for Normality				
Test		**Statistic**	**p Value**	
Shapiro-Wilk	**W**	0.947405	**Pr < W**	0.7060
Kolmogorov-Smirnov	**D**	0.186385	**Pr > D**	>0.1500
Cramer-von Mises	**W-Sq**	0.036686	**Pr > W-Sq**	>0.2500
Anderson-Darling	**A-Sq**	0.243937	**Pr > A-Sq**	>0.2500

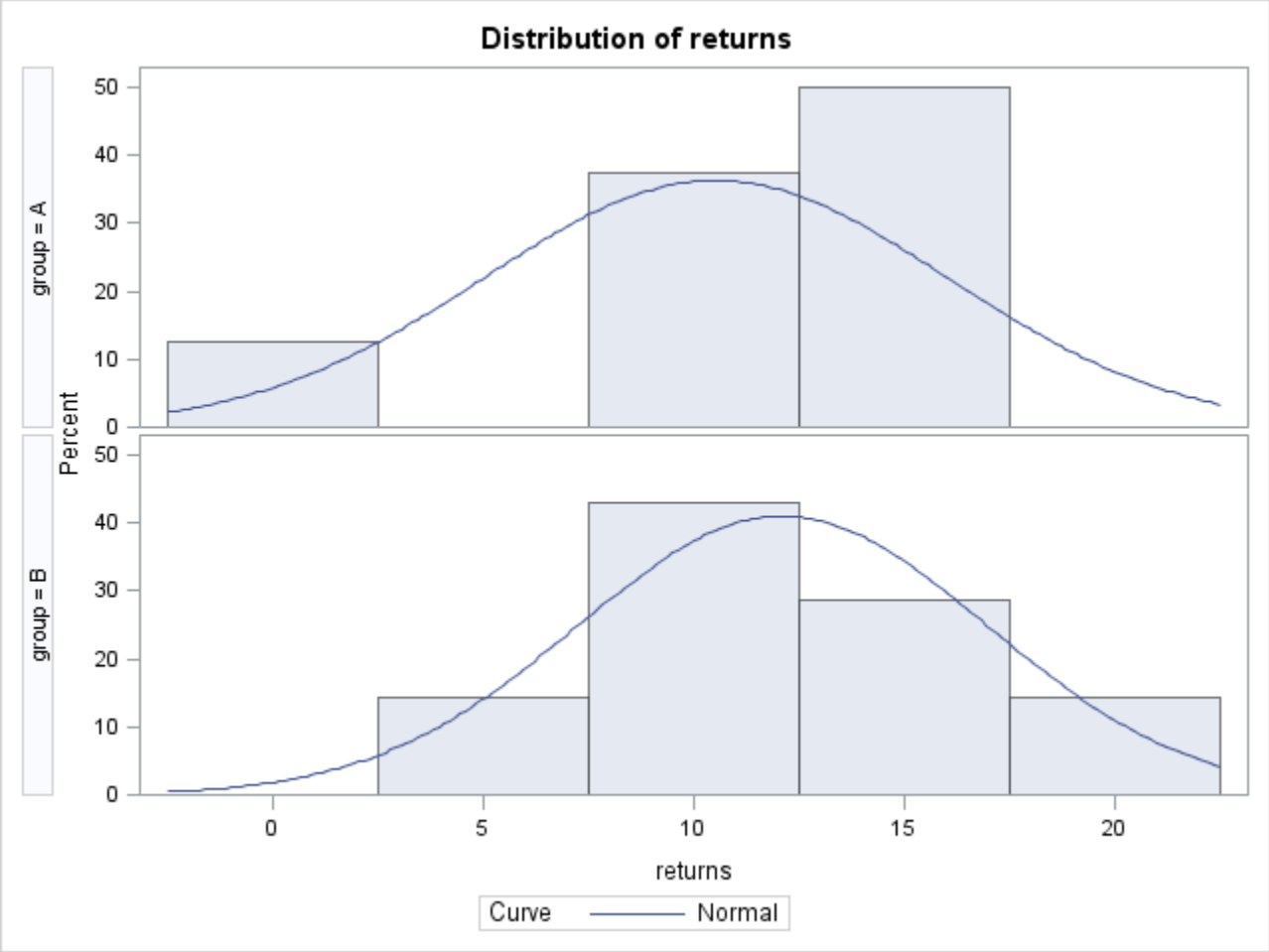
Quantiles (Definition 5)	
Quantile	**Estimate**
100% Max	18.0
99%	18.0
95%	18.0
90%	18.0
75% Q3	16.2
50% Median	11.6
25% Q1	9.5
10%	3.9
5%	3.9
1%	3.9
0% Min	3.9

Extreme Observations			
Lowest		**Highest**	
Value	**Obs**	**Value**	**Obs**
3.9	9	10.0	14
9.5	15	11.6	13
10.0	14	15.5	10
11.6	13	16.2	11
15.5	10	18.0	12

Objective 2 and 3

Produced when ODS Graphics are enabled. Result of HISTOGRAM statement with NORMAL option.

The UNIVARIATE Procedure



Objective 2 and 3

The UNIVARIATE Procedure
group = A
Fitted Normal Distribution for returns

Parameters for Normal Distribution		
Parameter	**Symbol**	**Estimate**
Mean	Mu	10.5125
Std Dev	Sigma	5.495566

Goodness-of-Fit Tests for Normal Distribution				
Test		**Statistic**	**p Value**	
Kolmogorov-Smirnov	**D**	0.21862226	**Pr > D**	>0.150
Cramer-von Mises	**W-Sq**	0.06456769	**Pr > W-Sq**	>0.250
Anderson-Darling	**A-Sq**	0.43370668	**Pr > A-Sq**	0.228

Quantiles for Normal Distribution		
Percent	**Quantile**	
	Observed	**Estimated**
1.0	-1.20000	-2.27210
5.0	-1.20000	1.47310
10.0	-1.20000	3.46965
25.0	8.65000	6.80580
50.0	11.75000	10.51250
75.0	13.90000	14.21920
90.0	16.70000	17.55535
95.0	16.70000	19.55190
99.0	16.70000	23.29710

The information on this and the next page are produced by the NORMAL option on the HISTOGRAM statement. Note the overlap of information produced using the NORMAL option on the PROC UNIVARIATE statement.

Objective 2 and 3

The UNIVARIATE Procedure
group = B
Fitted Normal Distribution for returns

Parameters for Normal Distribution		
Parameter	**Symbol**	**Estimate**
Mean	Mu	12.1
Std Dev	Sigma	4.862098

Goodness-of-Fit Tests for Normal Distribution				
Test		**Statistic**	**p Value**	
Kolmogorov-Smirnov	**D**	0.18638495	**Pr > D**	>0.150
Cramer-von Mises	**W-Sq**	0.03668579	**Pr > W-Sq**	>0.250
Anderson-Darling	**A-Sq**	0.24393730	**Pr > A-Sq**	>0.250

Quantiles for Normal Distribution		
Percent	**Quantile**	
	Observed	**Estimated**
1.0	3.90000	0.78907
5.0	3.90000	4.10256
10.0	3.90000	5.86897
25.0	9.50000	8.82056
50.0	11.60000	12.10000
75.0	16.20000	15.37944
90.0	18.00000	18.33103
95.0	18.00000	20.09744
99.0	18.00000	23.41093

Objective 2 and 3

Result of the PRINT procedure.
Contents of Data Table two.

Obs	group	return_size	return_mean	return_se
1	A	8	10.5125	1.94298
2	B	7	12.1000	1.83770

Testing Hypotheses and Computing Confidence Intervals

Notation:

	Population Parameters	Sample Statistics
Mean	μ	$\bar{y}$
Variance	σ^2	s^2
Standard Deviation	σ	s
		sample size = n

Hypotheses	**Test Statistic**	**Reject H_0 if**	**$(1-\alpha)100\%$ Confidence Interval for μ**
$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$ (μ_0 is specified)	$t = \dfrac{\bar{y} - \mu_0}{s/\sqrt{n}}$	$\lvert t \rvert \geq t_{\alpha/2,df}$ where df = n – 1 and $t_{\alpha/2,df}$ is the critical t-value that determines a right tail area of $\alpha/2$	$\bar{y} \pm t_{\alpha/2,df} \dfrac{s}{\sqrt{n}}$

An observed significance level or p-value is calculated for these tests rather than the critical region defined above. Recall that one would reject H_0 for small p-values. That is, reject H_0 if $p \leq \alpha$ for all tests of hypotheses, not just those about μ.

The UNIVARIATE procedure will also test the population variance and conduct a test for normality.

Hypotheses	**Test Statistic**	**Reject H_0 if**	**$(1-\alpha)100\%$ Confidence Interval for σ**
$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 \neq \sigma_0^2$ (σ_0^2 is specified)	$\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$	$\chi^2 \leq \chi^2_{1-\alpha/2,df}$ or $\chi^2 \geq \chi^2_{\alpha/2,df}$ where df = n - 1	$\dfrac{(n-1)s^2}{\chi^2_{\alpha/2,df}} \leq \sigma^2 \leq \dfrac{(n-1)s^2}{\chi^2_{1-\alpha/2,df}}$ where df = n - 1

For the test for normality, it the always the case that the null and alternative hypotheses are
H_0: pop is normally distributed
H_1: pop is not normally distributed

There are four tests produce by the UNIVARIATE procedure when the NORMAL option is included on the PROC UNIVARIATE statement. No calculations for the test statistics are given in this text. Interpret the tests based upon the observed significance levels.

In the UNIVARIATE procedure the CIBASIC option with the ALPHA=p option will generate confidence intervals for the mean, variance, and standard deviation using the above formulas. This procedure will also test the mean of the population using the t-test formula above and by two non-parametric methods: signed rank test, and sign test. The computations for those methods are not given here.

Objective 4: Modify the program in Objective 3. From the UNIVARIATE procedure test whether the mean of each of the groups is equal to 15.5. Include a HISTOGRAM statement overlaying a normal distribution with the hypothesized mean of 15.5.

This is done by adding an option to the PROC UNIVARIATE statement.

```
PROC UNIVARIATE DATA=one Mu0=15.5;
CLASS group ;
VAR returns;
HISTOGRAM returns / NORMAL (MU=15.5) ;
```

Note the change in two of the tables in the output.

The UNIVARIATE Procedure
Variable: returns
group = A

Tests for Location: Mu0=15.5

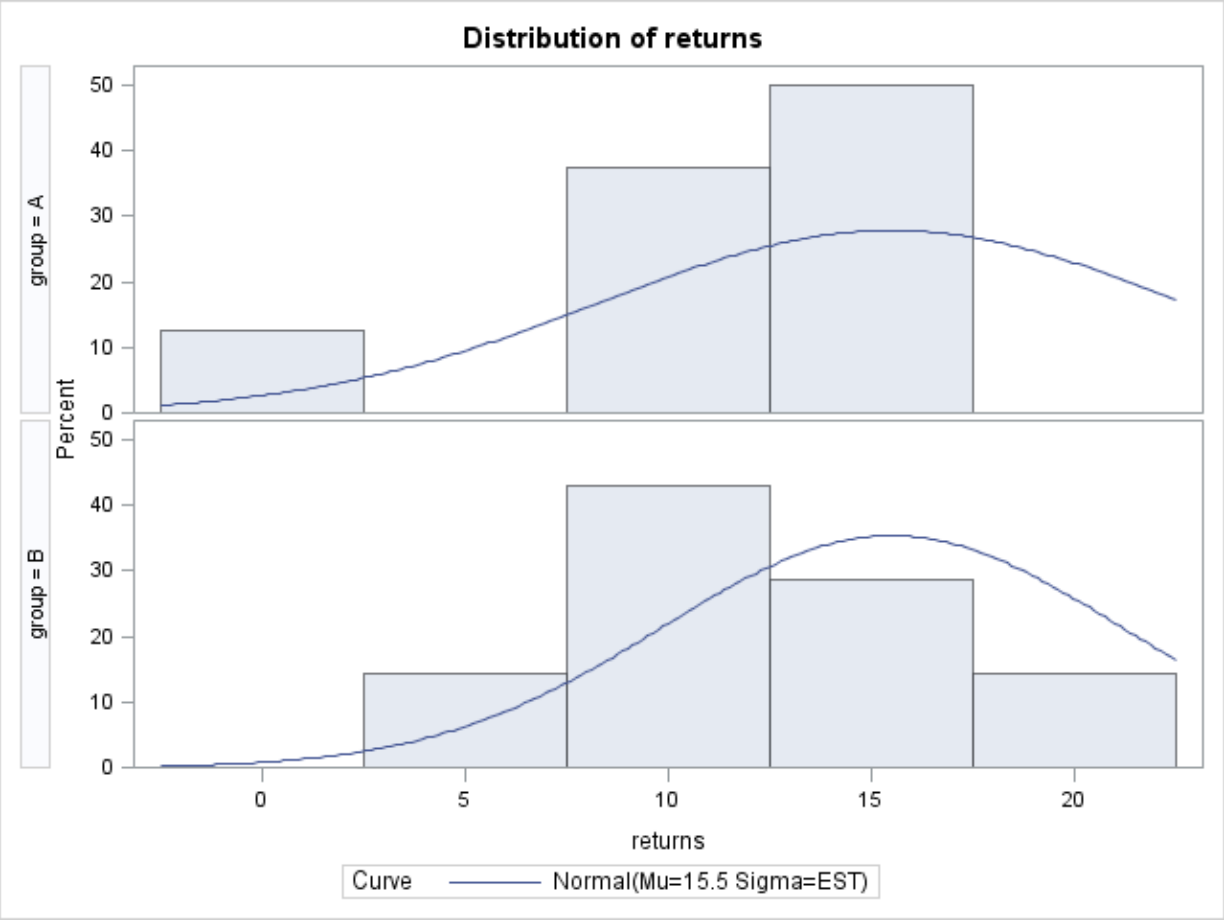
Test		Statistic		p Value
Student's t	t	-2.56694	Pr > |t|	0.0372
Sign	M	-3	Pr >= |M|	0.0703
Signed Rank	S	-16	Pr >= |S|	0.0234

The UNIVARIATE Procedure
Variable: returns
group = B

Tests for Location: Mu0=15.5

Test		Statistic		p Value
Student's t	t	-1.85014	Pr > |t|	0.1138
Sign	M	-1	Pr >= |M|	0.6875
Signed Rank	S	-7.5	Pr >= |S|	0.1563

The UNIVARIATE Procedure



One can also specify a standard deviation in addition to the mean in the HISTOGRAM statement. The syntax is

```
HISTOGRAM variable / NORMAL (MU=value  SIGMA=value) ;
```

More than one normal curve can be overlaid on the histogram produced by the UNIVARIATE procedure. Different colors can be chosen for the curves as well.

***Objective 5*:** Modify the programming in Objective 4 by overlaying a red normal curve centered at a mean of 12. Keep the curve requested in Objective 4.

```
PROC UNIVARIATE DATA=one Mu0=15.5;
CLASS group ;
VAR returns;
HISTOGRAM returns /     NORMAL (MU=15.5)
                        NORMAL (COLOR=red MU=12);
RUN;
```

The MEANS Procedure - This procedure does not generate as much output as the UNIVARIATE procedure, but is very useful for obtaining summary statistics for a data table.

PROC MEANS DATA=*tablename* <options>;
BY *variable;*
CLASS *variable list*;
VAR *variable1 variable2 ... variablen*;
OUTPUT OUT=*newtablename* <list options and new variable names> ;
RUN;

Some of the possible options for the PROC MEANS statement are for each variable specified in the VAR statement:

N	the number of nonmissing values
NMISS	the number of observations having missing values
MIN	the minimum value
MAX	the maximum value
RANGE	the range
SUM	the sum
MEAN	the mean
VAR	the variance
STD	the standard deviation
STDERR	the standard error of the mean
CV	coefficient of variation
T	Student's t for testing the hypothesis that the population mean is 0
PRT	the probability of a greater absolute value for the t-value above (observed significance level of the two-tailed test)
CLM	generates a two-sided 95% confidence interval, other confidence levels can be specified by using ALPHA= n option, where n is a number between 0 and 1. (1 - n)100% specifies the confidence level.
UCLM	specifies the 95% upper confidence limit, can use ALPHA=n to set confidence level.
LCLM	specifies the 95% lower confidence limit, can use ALPHA=n to set confidence level
ALPHA=n	sets a type I error level of n, where $0 < n < 1$
NOPRINT	The NOPRINT option is useful when selected items are being recovered in an output data set, and the actual output from the procedure does not need to appear in the output window. See OUTPUT statement.

It is not necessary to specify any options in the MEANS procedure. If no options are specified, then PROC MEANS prints the variable name, N, MEAN, STD, MIN, and MAX. If you specify options, only those values will be computed by the MEANS procedure.

BY statement

The UNIVARIATE procedure can operate on subgroups of the data set defined by the values of another variable. These subgroups are identified using the BY statement. This statement is optional. Data must first be sorted BY the variables (in the same order) as those appearing in a BY statement.

CLASS Statement

The CLASS statement is also an optional statement. The CLASS statement assigns a variable or variables to form subgroups. CLASS variables can be either numeric or character. The CLASS statement has an effect similar to a BY statement. The CLASS statement does NOT require the data to be sorted first.

VAR Statement

All of the numeric variables for which summary statistics are to be calculated are specified in the VAR statement. If you omit the VAR statement, the UNIVARIATE procedure will compute summary statistics for every numeric variable in your data table.

OUTPUT Statement

Some of the information produced by the MEANS procedure can be output to a new data set. This is another way to create a data table. The new data table can then be used in other SAS procedures. The syntax for the OUTPUT statement identifies a new data table name, and the statistics one wishes to recover for each variable in the VAR statement. Note the number and order of the variables in the VAR statement determine the number and order of variables to be newly created for each requested statistic. The syntax is

VAR *variable1 variable2 . . . variablen* ;
OUTPUT OUT=*newdatatablename* statistic1 = *stat1var1 stat1var2 . . stat1varn*
statistic2 = *stat2var1 stat2var2 . . . stat2varn*
:
statisticm = *statmvar1 statmvar2 . . . statmvarn* ;

The list of output statistics is identical to the list of options for the PROC MEANS statement. Specify the statistic and assign it a variable name, such as N = *size*.

As of this writing there are no ODS Graphics available in the MEANS procedure.

This SAS code will also produce mean, variance, minimum and maximum values for both genders. This method with the CLASS statement does not require the SORT procedure.

Objective 6: Run the MEANS procedure for the same data table the UNIVARIATE procedure was analyzing. Examine the default output.

```
PROC MEANS DATA=one  ;
TITLE 'Objective 6';
RUN;
```

The MEANS Procedure

Analysis Variable : returns				
N	Mean	Std Dev	Minimum	Maximum
15	11.2533333	5.0896066	-1.2000000	18.0000000

Notice the effect of not writing the VAR statement. All numeric variables will be analyzed.

Objective 7: Compute the minimum, maximum, mean, standard deviation, sample size, and the coefficient of variation for each of the groups. (PROC MEANS with a CLASS statement.)

```
PROC MEANS DATA=one MIN MAX MEAN STD N CV;
CLASS group ;
VAR returns ;
TITLE 'Objective 7';
RUN;
```

The MEANS Procedure

Analysis Variable : returns							
group	N Obs	Minimum	Maximum	Mean	Std Dev	N	Coeff of Variation
A	8	-1.2000000	16.7000000	10.5125000	5.4955664	8	52.2764936
B	7	3.9000000	18.0000000	12.1000000	4.8620983	7	40.1826307

Notice that the output contains the statistics requested in the PROC MEANS statement in the order that they were specified.

What happens if you listing only the CV option? Try it.

Objective 8: Rerun Objective 7. Recover the requested statistics in data table named in an OUTPUT statement. Examine the contents of the data table by adding a PRINT procedure. (You do not have to recover all of the statistics requested in the PROC MEANS statement, in practice; just those statistics you'll need further.)

```
PROC MEANS DATA=one MIN MAX MEAN STD N CV;
CLASS group ;
VAR returns ;
OUTPUT OUT=eight MIN=rmin MAX=rmax MEAN=rmean STD=rstd N=rn CV=rcv ;
TITLE 'Objective 8';
PROC PRINT DATA=eight ;
RUN;
```

Objective 8

The MEANS Procedure

Analysis Variable : returns

group	N Obs	Minimum	Maximum	Mean	Std Dev	N	Coeff of Variation
A	**8**	-1.2000000	16.7000000	10.5125000	5.4955664	8	52.2764936
B	**7**	3.9000000	18.0000000	12.1000000	4.8620983	7	40.1826307

Objective 8

Obs	group	_TYPE_	_FREQ_	rmin	rmax	rmean	rstd	rn	rcv
1		0	15	-1.2	18.0	11.2533	5.08961	15	45.2275
2	A	1	8	-1.2	16.7	10.5125	5.49557	8	52.2765
3	B	1	7	3.9	18.0	12.1000	4.86210	7	40.1826

Run this first with the OUTPUT statement as shown above, and a second time with the NOPRINT option on the PROC MEANS statement.

Note: *This data set has only one response variable to be analyzed. If multiple variables are analyzed in the VAR statement, one would also have to provide multiple variable names in the OUTPUT statement. The names you select must correspond to the order of the variables in the VAR statement. For example, if additionally there were variables x, y and z in the analysis:*

```
PROC MEANS DATA=one MIN MAX MEAN STD N CV;
CLASS group ;
VAR returns x y z;
OUTPUT OUT=eight  MIN=rmin xmin ymin zmin      MAX=rmax xmax ymax zmax
                  MEAN=rmean xmean ymean zmean STD=rstd xstd ystd zstd
                  N=rn xn yn zn                CV=rcv xcv ycv zcv;
RUN;
```

Objective 9: Compute the default summary statistics for each group using a BY statement rather than a CLASS statement. (PROC MEANS, BY group – Note the data are sorted by group first.)

```
PROC SORT DATA=one; BY group;
PROC MEANS DATA=one; BY group;
TITLE 'Objective 9';
RUN;
```

Objective 10: Repeat Objective 9 adding the OUTPUT statement and the PRINT procedure to view the contents of the output data table. How does this data table compare to that output when using a CLASS statement.

How do both of these compare to the output data tables from the UNIVARIATE procedure demonstrated in Objective 3?