Chapter 16: Statistical Principles

In this chapter:

Describing data (*UE* 16.1):

The annual percentage returns for 25 stock mutual funds printed in *UE*, Table 16.2, p. 522 will be used to show the EViews commands needed to calculate the descriptive statistics described in *UE* 16.1.

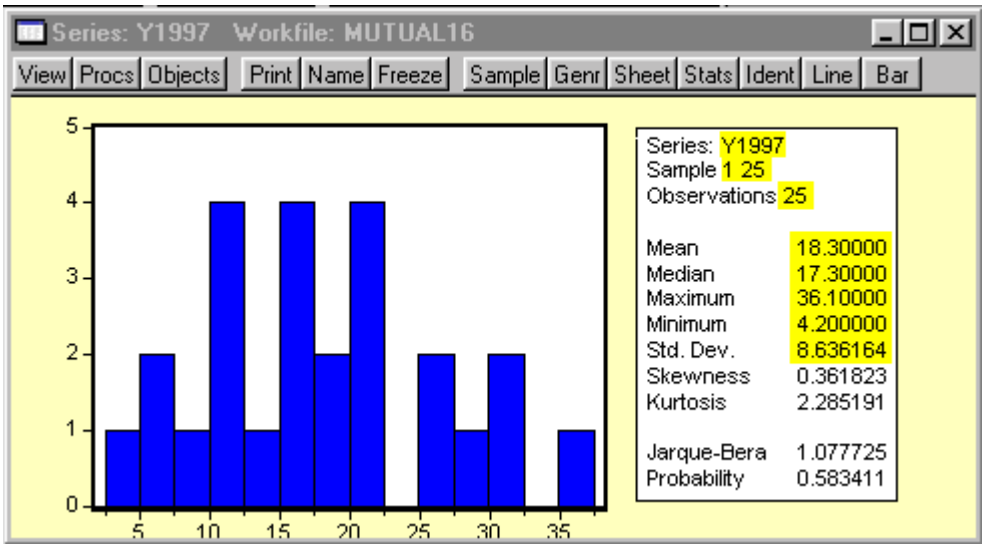
Follow these steps to view a histogram and the standard descriptive statistics for a series:

Step 1. Create a new Undated or irregular workfile with 25 observations. Use **Genr** to create a new variable named *Y1997* and enter the 1997 returns from *UE*, Table 16.2, p. 522 into the series. Repeat the process for 1998 returns (name the variable *Y1998*). Refer to Chapter 1 for help.

Step 2. Save the workfile by selecting **File/Save As…** on the EViews main menu bar, and enter *Mutual16.wf1* in the File name: window (check to make sure *workfile* is selected in the Save as type: window).

Step 3. Open the variable named *Y1997* in a new window by double clicking its icon in the workfile window.

Step 4. Select **View/Descriptive Statistics/Histogram and Stats** on the workfile menu bar to reveal the graphic shown below (statistics highlighted in yellow are described in *UE* 16.1).



The EViews output shows a histogram of the data series plus major descriptive statistics. The histogram divides the series range (the distance between the maximum and minimum values) into a number of equal length intervals or bins and displays a count of the number of observations that fall into each bin. The histogram is useful when investigating the distributional characteristics of a series (see *UE* 16.2).

The descriptive statistics window on the right of the graphic displays the standard descriptive statistics. All of the statistics are calculated using observations in the current sample. The top of the window identifies the series name, sample, and the number of observations. The descriptive statistics described in *UE* 16.1 are presented next and they are defined as follows:

1. Mean (the average value of the series, obtained by adding up the series and dividing by the number of observations)
2. Median (the middle value (or average of the two middle values) of the series when the values are ordered from the smallest to the largest)
3. Maximum (the maximum value of the series in the current sample)
4. Minimum (the minimum value of the series in the current sample)
5. Std. Dev. (standard deviation) is a measure of dispersion or spread in the series

Individual scalar values for the following descriptive statistics[1] can be obtained by typing the functions, printed in the first column of the table below, in the command window, and pressing **Enter**. The values can be viewed on the status line in the lower left portion of the screen.

Function	Name
=@obs(Y1997)	number of observations
=@mean(Y1997)	mean
=@median(Y1997)	median
=@min(Y1997)	minimum
=@max(Y1997)	maximum
=@stdev(Y1997)	sample standard deviation
=@sum(Y1997)	sum
=@sumsq(Y1997)	sum-of-squares
=@var(Y1997)	variance
=@var(Y1997)*(@obs(Y1997)/(@obs(Y1997)-1))	sample variance =@stdev(Y1997)^2

Note that EViews calculates the population variance, which is a biased measure of the population variance in samples. Since the variance is calculated as the sum of the squared deviations of a series observations about its mean divided by the number of observations (i.e., divided by *n* instead of an unbiased estimator, which divides by *n-1*), it is better to calculate the sample variance as the square of the unbiased sample standard deviation {i.e., *=@stdev(Y1997)^2*}. Due to the way EViews calculates variance, think of the variance as the square of the standard deviation instead of the other way around.

[1] The variable *Y1997* is used for demonstration purposes so you can try it out in the *Mutual16.wf1* workfile. Substitute any variable name, for *Y1997*, to calculate the statistic for that variable in a workfile. Other measures obtainable by EViews commands can be found in **Help/Function Reference** under the heading: Descriptive Statistics Functions.

The remaining four numbers displayed in the descriptive statistics window (see the graphic on the first page) are defined below:

1. Skewness (the skewness of a symmetric distribution, such as the normal distribution, is zero)
2. Kurtosis (the kurtosis of the normal distribution is 3)
3. Jarque-Bera test (under the null hypothesis of a normal distribution, the Jarque-Bera statistic is distributed as χ^2 with 2 degrees of freedom
4. Probability (the probability that a Jarque-Bera statistic exceeds the observed value under the null indicates that a small probability value leads to the rejection of the null hypothesis of a normal distribution)

These measures of normality are not discussed in *UE* but the normal distribution and its properties are described in *UE* 16.2.5, pp. 539 - 543.

The same descriptive statistics can be calculated for a group of variables, sans the histogram, by opening a group of variables in one window and selecting **View/Descriptive Statistics/Histogram and Stats** on the workfile menu bar.

Probability distributions (*UE* 16.2):

EViews enables you to calculate the cumulative distribution density (CDF or inverse CDF) or probability functions, cumulative distribution, and random number generators for 17 statistical distributions. We have already used EViews to calculate the critical t-value for t-tests (see Chapter 5 and Calculating a confidence interval for a population mean) and generated random numbers for the Monte Carlo Simulation to demonstrate that the estimated βs are drawn from a normal distribution (see Chapter 4). Further discussion of this topic is beyond the scope of this guide, but further a explanation of EViews capabilities relating to statistical distribution functions can be found in **Help/Function Reference**.

Standardized variables (*UE* 16.2.4):

Complete **Steps 1 & 2** of Describing data prior to attempting this section. To calculate the standardized values for *Y1997*, follow these steps:

Step 1. Open the EViews workfile named *Mutual16.wf1*.
Step 2. Type *series Y1997standized = (y1997-@mean(Y1997))/@stdev(Y1997)* in the command window and press **Enter** on the keyboard. Y1997standized successfully computed. will appear in the status line in the lower left of the screen.
Step 3. To view the standardized values for *Y1997*, double click the *Y1997standized* series icon in the workfile menu.

Calculating a confidence interval for a population mean (*UE* 16.4.6):

Complete **Steps 1 & 2** of Describing data before attempting this section. To calculate the 95%[2] confidence interval for the population mean of *Y1998*:

[2] To compute the 99% confidence interval, substitute *.995* for *.975* in **Step 2** and **Step 3** below.

Step 1. Open the EViews workfile named *Mutual16.wf1*.

Step 2. To calculate the upper confidence interval, type *scalar CI_Y1998_HIGH = @mean(Y1998)+(@qtdist(.975,@obs(Y1998)-1)*(@stdev(Y1998)/(@obs(Y1998)^.5)))* in the command window, and press **Enter**. Double click on the *CI_Y1998_HIGH* icon in the workfile to view the value on the status line in the lower left portion of the screen. Did you get Scalar CI_Y1998_HIGH = 14.1972846046?

Step 3. To calculate the lower confidence interval, type *scalar CI_Y1998_LOW = @mean(Y1998)-(@qtdist(.975,@obs(Y1998)-1)*(@stdev(Y1998)/(@obs(Y1998)^.5)))* in the command window, and press **Enter**. Double click on the *CI_Y1998_LOW* icon in the workfile to view the value on the status line in the lower left portion of the screen. Did you get Scalar CI_Y1998_LOW = 3.65871539542?

Hypothesis testing, the test statistic, and statistical significance (*UE* 16.5.3 & *UE* 16.5.4):

Complete steps 1 and 2 of Describing data prior to attempting this section. To test the hypothesis that the average return for all mutual funds in 1998 was 28.1 percent based on the sample of 25 mutual fund returns (i.e., series *Y1998*), follow these steps:

Step 1. Open the EViews workfile named *Mutual16.wf1*.

Step 2. Double click the *Y1998* series icon in the workfile window.

Step 3. Select **View/Tests for Descriptive Stats/Simple Hypothesis Tests** on the series menu bar to reveal the Series Distributions Tests dialog window. Enter *28.*1 in the Mean: window under Test Value: and click **OK** to reveal the EViews test output table shown below.[3]

Hypothesis Testing for Y1998			
Date: 07/16/00 Time: 16:29			
Sample: 1 25			
Included observations: 25			
Test of Hypothesis: Mean = 28.10000			
Sample Mean = 8.928000			
Sample Std. Dev. = 12.76537			
Method		Value	Probability
t-statistic		-7.509381	0.0000

EViews prints values for the: Test of Hypothesis: Mean, Sample mean, Sample Std. Dev., t-statistic: Value (based on the formula for t printed in *UE*, p. 558), and Probability. The t-statistic value of -7.509381 is equal to the value printed in *UE*, p. 558. This value can be compared with the critical t-value found in the *UE*, Table B-1 using the one-tailed or two-tailed test at various levels of confidence. The reported Probability is the p-value, or marginal significance level, against a two-sided alternative. If this probability value is less than the size of the test, say 0.05, we reject the null hypothesis (i.e., Mean = 28.10000). The probability value for a one-sided alternative is one-half the p-value of the two-sided test.

[3] You can enter a value for the series standard deviation in the window under Mean Test Assumption:, if it is known. If unknown, leave the window blank and EViews will use the sample standard deviation in the test calculation.

Exercises:

3. Follow the steps in Describing data to calculate the mean and standard deviation.

10. Follow the steps in Describing data to create a workfile for the problem. Then follow the steps in Calculating a confidence interval for a population mean to calculate the 99% confidence interval (check footnote 2). You could also answer this problem by following the procedures in Hypothesis testing, the test statistic, and statistical significance.

13. Follow the steps in Describing data to create a workfile for the problem. Then follow the steps in Hypothesis testing, the test statistic, and statistical significance to calculate the t-value and probability that the sample is drawn from a population with a mean value of 11.2.