Chapter 7
Inferences Based on a Single Sample: Estimation with Confidence Intervals

7.1 Introduction

Chapter 7 introduces the reader to estimating population parameters with confidence intervals. Three parameters - the population mean, proportion, and variance - are studied in the chapter. The reader is also introduced to the topic of sample size determination, as it follows nicely from the estimation material presented.

XLSTAT provides the calculation of confidence intervals for all three population parameters presented in the text. While **XLSTAT** provides all three confidence intervals, it is not always easy locating these confidence intervals within the **XLSTAT** menus.

When estimating a population mean, **XLSTAT** offers the confidence interval calculations in a couple of different menus. We choose to utilize the confidence intervals found in the **Parametric Tests** tab as **XLSTAT** allows for calculations using both the z-distribution and the t-distribution in this location. The confidence interval is grouped together with the **One-sample t-test and z-test** that we will discuss in the next chapter. The confidence interval for a population proportion is similarly grouped together with the **Test for one proportion** within **XLSTAT**. Finally, the confidence interval for a population variance is contained within the **Descriptive statistics** menu that is found in the **Describing data** tab. This menu also allows us to estimate a population mean, but only gives us the t-distribution option for the calculations.

We note here that no computer work exists to aid in the determination of the sample size that is presented in the text. In addition, we note that the confidence interval material covered in *A First Course in Statistics* appears in Chapter 5 of that text. Lastly, the confidence interval for a population variance material is not covered in in *A First Course in Statistics*.

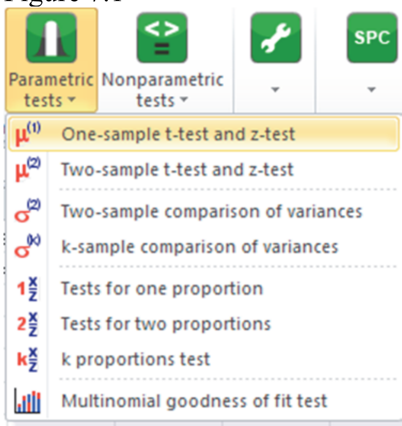
The following examples from *Statistics* illustrate the confidence interval calculations that can be found using XLSTAT in this chapter:

Excel Companion Exercise	Page	Statistics Example	Excel File Name
7.1	95	Example 7.3	AIRNOSHOWS
7.2	98	Example 7.5	PRINTHEADS
7.3	100	Example 7.7	
7.4	102	Example 7.8	
7.5	104	Example 7.11	FISHDDT

7.2 Estimation of a Population Mean Using the Z-distribution

When estimating a population mean, it is sometimes possible that the population standard deviation will be known. In such cases, it is possible to utilize the z-distribution in the calculation of the confidence interval. To use the estimation tool within **XLSTAT**, **open** a new workbook and place the cursor in the upper left cell of the worksheet. Click on the **XLSATAT Add-In** menu. Click on the **Parametric tests** option to access the **One-sample t-test and z-test** menu (see Figure 7.1).

Figure 7.1



We illustrate how to use this technique with the following exercise:

Exercise 7.1: We use Example 7.3 found in the *Statistics* text.

Problem Unoccupied seats on flights cause airlines to lose revenue. Suppose a large airline wants to estimate its average number of unoccupied seats per flight over the past year. To accomplish this, the records of 225 flights are randomly selected and the number of unoccupied seats is noted for each of the sample flights (and saved in the AIRNOSHOWS data file). Descriptive statistics are displayed in the Excel printout shown in Figure 7.2. Estimate μ, the mean number of unoccupied seats per flight during the past year using a 90% confidence interval.

Figure 7.2

NOSHOWS	
Mean	11.59556
Median	12
Mode	10
Standard Deviation	4.102588
Sample Variance	16.83123
Minimum	0
Maximum	23
Count	225

Solution:

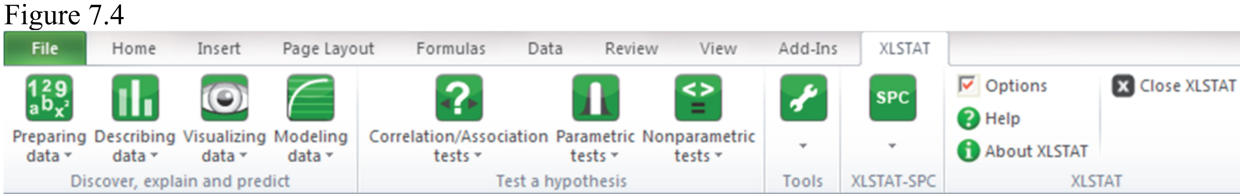
We solve Exercise 7.1 utilizing the **One-sample t-test and z-test** menu presented in **XLSTAT**. **Open** the data file **AIRNOSHOWS** by following the directions found in the preface of this manual. If done correctly, the data should appear in a workbook similar to that shown in Figure 7.3.

Figure 7.3

	A
1	NOSHOWS
2	12
3	11
4	19
5	6
6	10
7	8
8	11
9	13
10	10
11	2
12	20
13	13

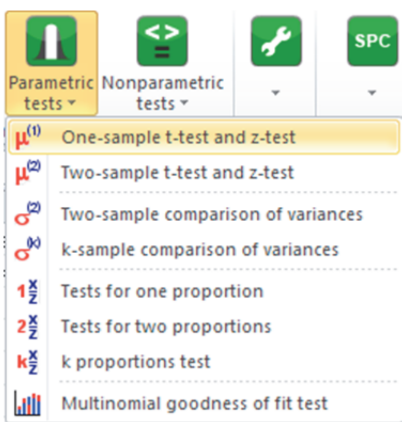
To create the desired confidence interval, we click on the **XLSTAT** tab at the top of the **Excel** workbook to access the **XLSTAT** menus shown in Figure 7.4.

Figure 7.4



To generate the confidence interval, we click on the **Parametric tests** menu and select the **One-sample t-test and z-test** option shown in Figure 7.5.

Figure 7.5



This opens the **One-sample t-test and z-test** menu shown in Figures 7.6-7.7. We need to first specify the location of the data that is to be analyzed. In our data set, the data is located in Column A, rows 2 – 226, with row 1 being the variable label. We specify the location in the **Data** box and check the **Column labels** box to indicate the first row of data represents the variable name, **NOSHOWS**. Please note that you may choose to drag the mouse over the range of the data to be included instead of typing the location in the data box as previously described. In order to utilize the z-distribution in the calculation of the confidence interval, we need to check the **z test** box in this menu.

Click on the **Options** tab (shown in Figure 7.7) to specify the **significance level** (which is the opposite of the confidence level) and to also indicate how we want **XLSTAT** to work with the population variance required in the confidence interval calculations. If this value is known, we check the **User defined** button and enter the value in the **Variance** box. If not, we check the **Estimated using samples** button. In our example, we have entered **10%** as the **Significance level** (to create a 90% confidence interval) and checked the **Estimated using samples** button. Click **OK** to create the confidence interval.

Figure 7.6

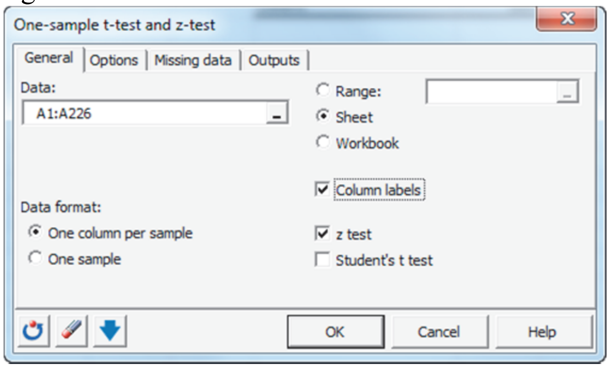
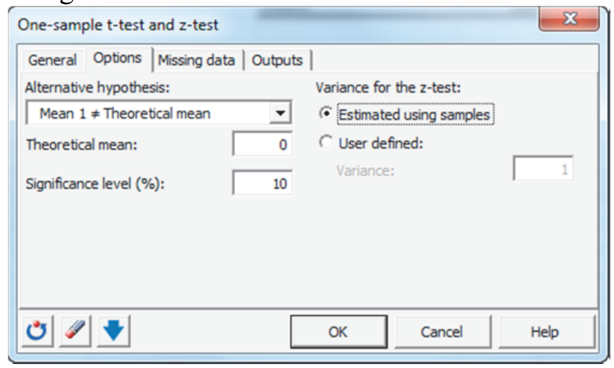


Figure 7.7



The XLSTAT output is shown in Figure 7.8.

Figure 7.8

One-sample z-test / Two-tailed test:

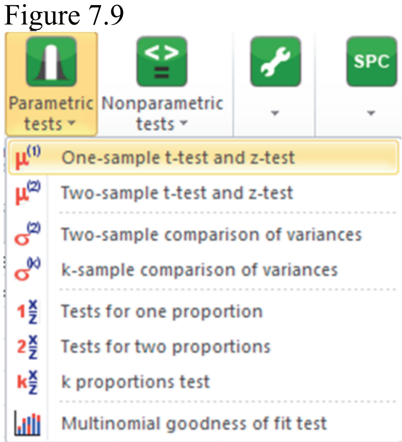
90% confidence interval on the mean:
(11.1457, 12.0454)

We compare these values to the ones shown in the text. We note that these values correspond nicely with the confidence interval endpoints found using the formula presented in the text. The MINITAB values shown in the text utilize the t-distribution in their calculations and are slightly different than the ones we found here. These small differences our common and are considered negligible.

7.3 Estimation of a Population Mean Using the t-distribution

When estimating a population mean, it is likely that the value of the population standard deviation will be unknown. In such cases, the t-distribution is used in the calculation of the confidence interval. To use the estimation tool within **XLSTAT**, **open** a new workbook and place the cursor in the upper left cell of the worksheet. Click on the **XLSATAT Add-In** menu. Click on the **Parametric tests** option to access the **One-sample t-test and z-test** menu (see Figure 7.9).

Figure 7.9



We illustrate how to use this technique with the following exercise:

Exercise 7.2: We use Example 7.5 found in the *Statistics* text.

Problem Some quality control experiments require *destructive sampling* (i.e. the test to determine whether the item is defective destroys the item) in order to measure a particular characteristic of the product. The cost of destructive sampling often dictates small samples. Suppose a manufacturer of printers for personal computers wishes to estimate the mean number of characters printed before the printhead fails. The printer manufacturer tests n = 15 printheads and records the number of characters printed until failure for each. These 15 measurements (in millions of characters) are listed in Table 7.1. Form a 99% confidence interval for the mean number of characters printed before the printhead fails.

Table 7.1

1.13	0.92	1.32	1.48	1.18
1.55	1.25	0.85	1.20	1.22
1.43	1.36	1.07	1.33	1.29

Solution:

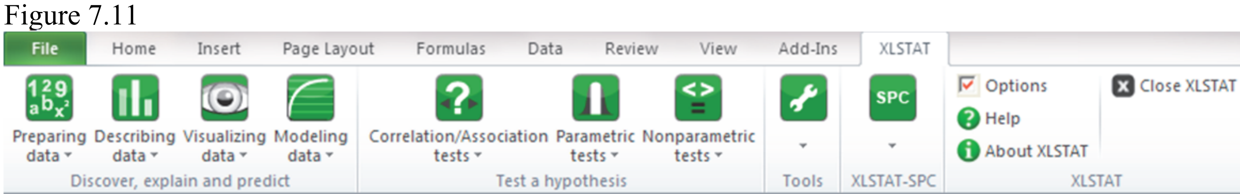
We solve Exercise 7.2 utilizing the **One-sample t-test and z-test** menu presented in **XLSTAT**. **Open** the data file **PRINTHEADS** by following the directions found in the preface of this manual. If done correctly, the data should appear in a workbook similar to that shown in Figure 7.10.

Figure 7.10

	A
1	NUMCHAR
2	1.13
3	1.55
4	1.43
5	0.92
6	1.25
7	1.36
8	1.32
9	0.85
10	1.07
11	1.48

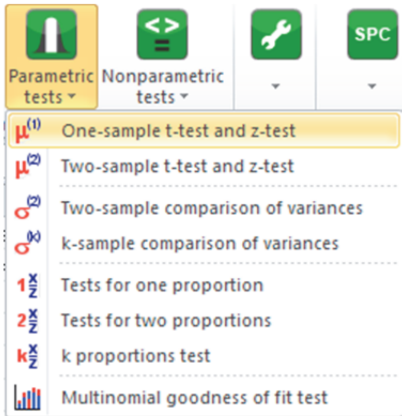
To create the desired confidence interval, we click on the **XLSTAT** tab at the top of the **Excel** workbook to access the **XLSTAT** menus shown in Figure 7.11.

Figure 7.11



To generate the confidence interval, we click on the **Parametric tests** menu and select the **One-sample t-test and z-test** option shown in Figure 7.12.

Figure 7.12



This opens the **One-sample t-test and z-test** menu shown in Figures 7.13-7.14. We need to first specify the location of the data that is to be analyzed. In our data set, the data is located in Column A, rows 2 – 16, with row 1 being the variable label. We specify the location in the **Data** box and check the **Column labels** box to indicate the first row of data represents the variable name, **NUMCHAR**. Please note that you may choose to drag the mouse over the range of the data to be included instead of typing the location in the data box as previously described. In order to utilize the t-distribution in the calculation of the confidence interval, we need to check the **Student's t test** box in this menu.

Click on the **Options** tab (shown in Figure 7.14) to specify the **significance level** (which is the opposite of the confidence level). In our example, we have entered **1%** as the **Significance level** (to create a 99% confidence interval. Click **OK** to create the confidence interval.

Figure 7.13

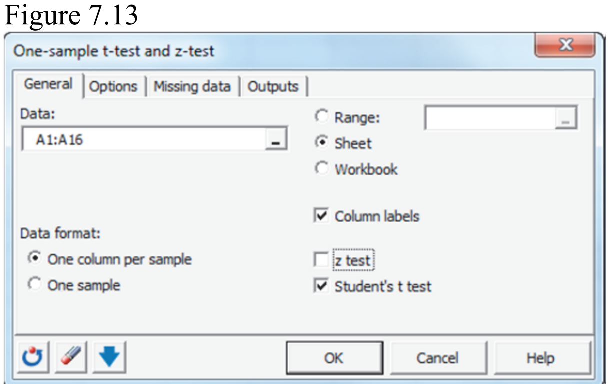
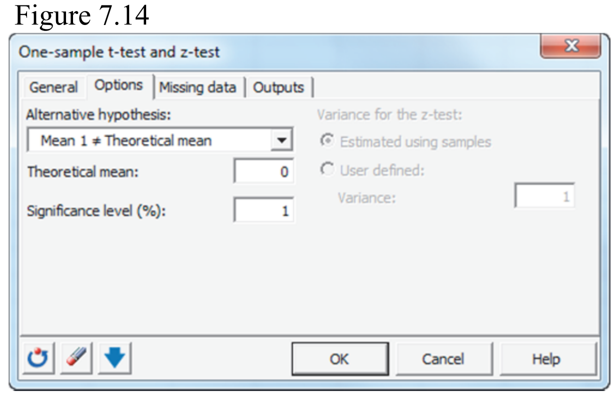


Figure 7.14



The XLSTAT output is shown in Figure 7.15.

Figure 7.15

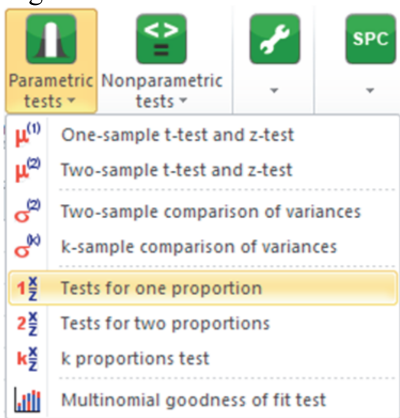
99% confidence interval on the mean:
(1.0902, 1.3871)

We compare these values to the ones shown in the text. We note that these values are exactly the same as the values shown on the MINITAB printout.

7.4 Estimation of a Population Proportion

When estimating a population proportion, XLSTAT requires the user to enter the sample size, the number of successes and the confidence level to calculate a confidence interval. To use the estimation tool within **XLSTAT**, **open** a new workbook and place the cursor in the upper left cell of the worksheet. Click on the **XLSATAT Add-In** menu. Click on the **Parametric tests** option to access the **Tests for one proportion** menu (see Figure 7.16).

Figure 7.16



We illustrate using the next exercise.

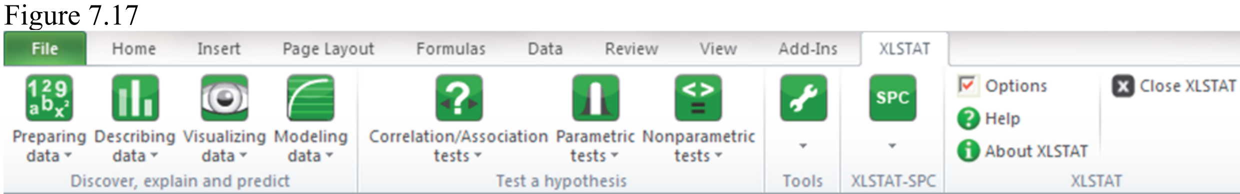
Exercise 7.3: We use Example 7.7 from the *Statistics* text.

Problem Many public polling agencies conduct surveys to determine the current consumer sentiment concerning the state of the economy. For example, the Bureau of Economic and Business Research (BEBR) at the University of Florida conducts quarterly surveys to gauge consumer sentiment in the Sunshine State. Suppose that BEBR randomly samples 484 consumers and finds that only 157 are optimistic about the state of the economy. Use a 90% confidence interval to estimate the proportion of all consumers in Florida who are optimistic about the state of the economy.

Solution:

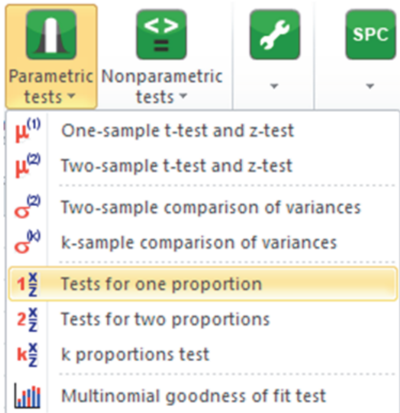
We solve Exercise 7.3 utilizing the **Tests for one proportion** menu presented in **XLSTAT**. To create the desired confidence interval, we click on the **XLSTAT** tab at the top of the **Excel** workbook to access the **XLSTAT** menus shown in Figure 7.17.

Figure 7.17



To generate the confidence interval, we click on the **Parametric tests** menu and select the **Tests for one proportion** option shown in Figure 7.18.

Figure 7.18



This opens the **Tests for one proportion** menu shown in Figures 7.19-7.20. We need to first select how the data is presented in the problem. If the data is given in terms of the number of successes, we click on the **Frequency** button. If the data is given in terms of the proportion of successes, we would click on the **Proportion** button. In this problem, we are told that 157 of the 484 consumers are optimistic about the state of the economy. We click on the **Frequency** button and enter a **Frequency** of **157** and a **Sample size** of **484**.

Click on the **Options** tab (shown in Figure 7.20) to specify the **significance level** (which is the opposite of the confidence level). In our example, we have entered **10%** as the **Significance level** (to create a 90% confidence interval. We also make sure to click on the **Wald** button to generate the confidence interval we desire. Click **OK** to create the confidence interval.

Figure 7.19

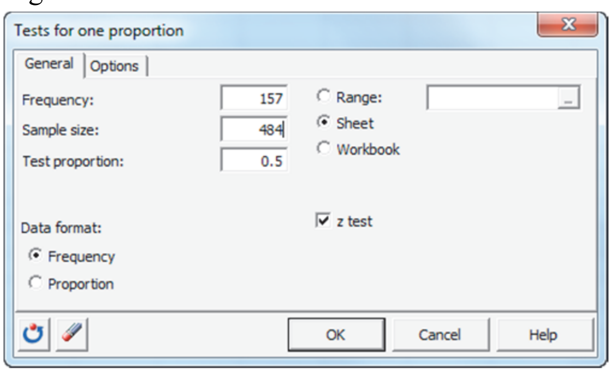
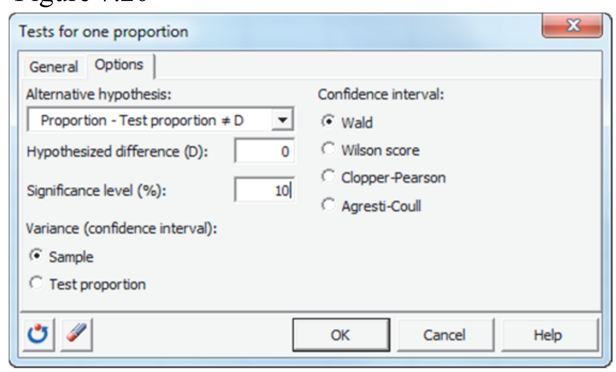


Figure 7.20



The **XLSTAT** output is shown in Figure 7.21.

Figure 7.21

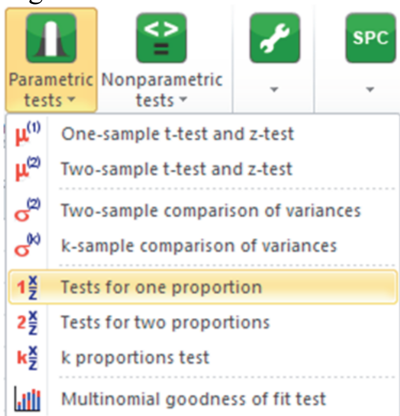
90% confidence interval on the proportion (Wald):
(0.2894, 0.3594)

We compare these values to the ones shown in the text. We note that these values are exactly the same as the values shown on the MINITAB printout.

7.5 Estimation of a Population Proportion Close to 0 or 1

When estimating a population proportion that is close to the value 0 or 1, the estimation technique has been shown to perform poorly. An alternative confidence interval technique, the Wilson technique, is used. **XLSTAT** requires the user to select this option when estimating a population proportion. To use the estimation tool within **XLSTAT**, **open** a new workbook and place the cursor in the upper left cell of the worksheet. Click on the **XLSATAT Add-In** menu. Click on the **Parametric tests** option to access the **Tests for one proportion** menu (see Figure 7.22).

Figure 7.22



We illustrate using the next exercise.

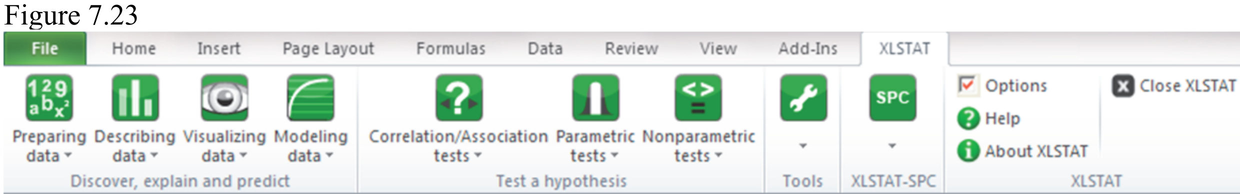
Exercise 7.4: We use Example 7.8 from the *Statistics* text.

Problem According to True Odds: How Risk Affects Your Everyday Life (Walsh, 1997), the probability of being the victim of a violent crime is less than .01. Suppose that, in a random sample of 200 Americans, 3 were victims of a violent crime. Use a 95% confidence interval to estimate the true proportion of Americans who were victims of a violent crime.

Solution:

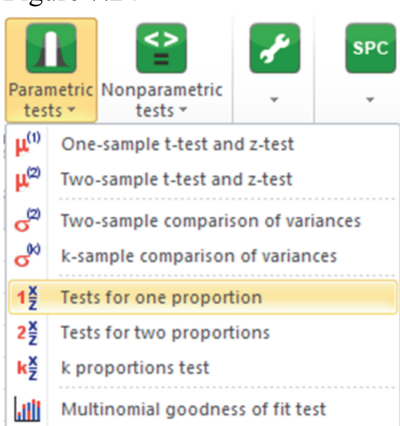
We solve Exercise 7.4 utilizing the **Tests for one proportion** menu presented in **XLSTAT**. To create the desired confidence interval, we click on the **XLSTAT** tab at the top of the **Excel** workbook to access the **XLSTAT** menus shown in Figure 7.23.

Figure 7.23



To generate the confidence interval, we click on the **Parametric tests** menu and select **Tests for one proportion** option shown in Figure 7.24.

Figure 7.24



This opens the **Tests for one proportion** menu shown in Figures 7.25-7.26. We need to first select how the data is presented in the problem. If the data is given in terms of the number of successes, we click on the **Frequency** button. If the data is given in terms of the proportion of successes, we would click on the **Proportion** button. In this problem, we are told that 3 of the 200 consumers are optimistic about the state of the economy. We click on the **Frequency** button and enter a **Frequency** of **3** and a **Sample size** of **200**.

Click on the **Options** tab (shown in Figure 7.26) to specify the **significance level** (which is the opposite of the confidence level). In our example, we have entered **5%** as the **Significance level** (to create a 95% confidence interval. We also make sure to click on the **Wilson score** button to generate the confidence interval we desire. Click **OK** to create the confidence interval.

Figure 7.25

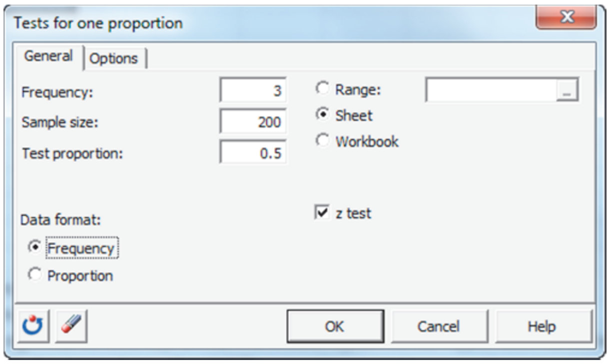
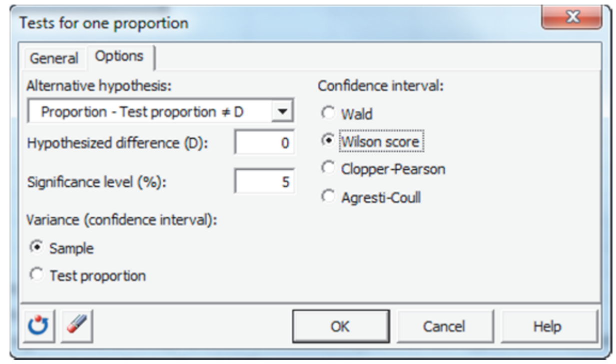


Figure 7.26



The XLSTAT output is shown in Figure 7.27.

Figure 7.27

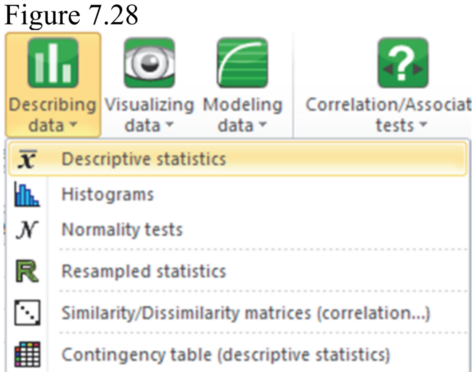
95% confidence interval on the proportion (Wilson score):
(0.0039, 0.0468)

We note that these values are exactly the same as the values shown in the text.

7.6 Estimation of a Population Variance

Finding where to estimate a population variance within **XLSTAT** can be difficult. It is hidden within a different menu than the menus we covered earlier in this chapter. To use the estimation tool within **XLSTAT**, **open** a new workbook and place the cursor in the upper left cell of the worksheet. Click on the **XLSATAT Add-In** menu. Click on the **Describing data** option to access the **Descriptive statistics** menu (see Figure 7.28).

Figure 7.28



We illustrate using the next exercise.

Exercise 7.5: We use Example 7.11 from the *Statistics* text.

Problem Refer to U.S. Army Corps of Engineers study of contaminated fish in the Tennessee River. The Corps of Engineers have collected data for a random sample of 144 fish contaminated with DDT. (The Engineers made sure to capture contaminated fish in several different randomly selected streams and tributaries of the river.) The fish weights (in grams) are saved in the FISHDDT file. The Army Corps of Engineers wants to estimate the true variation in fish weights in order to determine whether the fish are stable enough to allow further testing for DDT contamination. Use the sample data to find a 95% confidence interval for the parameter of interest.

Solution:

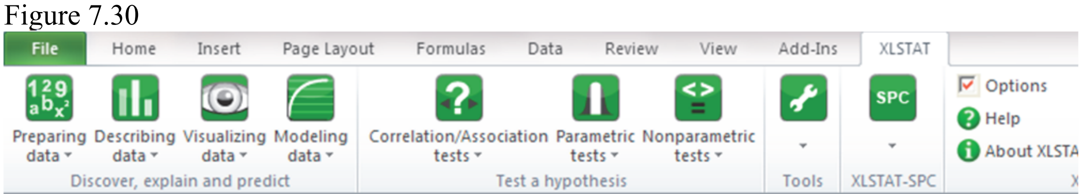
Figure 7.29

E
WEIGHT
732
795
547
465
1252
1255
741
1151
1186
754
679
985

We solve this problem by using the **Describing data** menu within the **XLSTAT** program. Before we begin, we must access the data set for this example. **Open** the data file **FISHDDT** by following the directions found in the preface of this manual. If done correctly, the data should appear in a workbook similar to that shown in Figure 7.29. Use the mouse to select the data shown in the workbook.

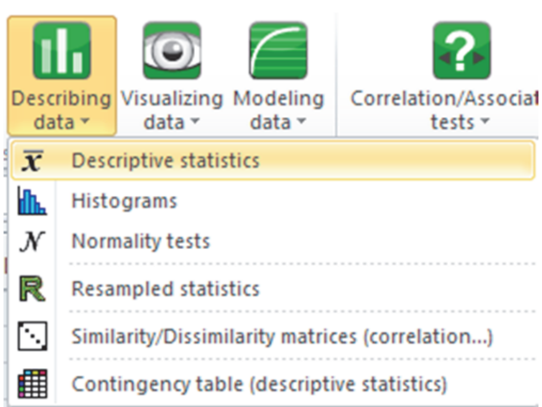
To create the desired displays, we click on the **XLSTAT** tab at the top of the **Excel** workbook to access the **XLSTAT** menus shown in Figure 7.30.

Figure 7.30



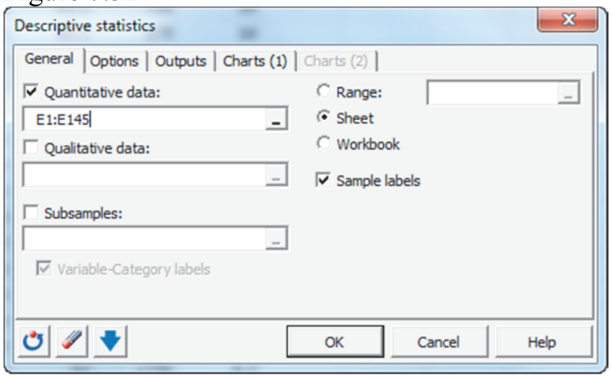
To calculate the desired confidence interval, we click on the **Describing data** menu and select the **Descriptive statistics** option shown in Figure 7.31.

Figure 7.31



This opens the **Descriptive statistics** menu shown in Figures 7.32 – 7.34. In order to work with quantitative data, we need to make sure that the **Quantitative data** box is checked in the **General** tab on the **Descriptive statistics** menu. In addition, we need to specify the location of the data that is to be analyzed. In our data set, the data is located in Column E, rows 1 - 145, with row 1 being the variable label. We specify the location in the **Quantitative data** box and check the **Sample labels** box to indicate the first row of data represents the variable name, **WEIGHT**. Please note that you may choose to drag the mouse over the range of the data to be included instead of typing the location in the data box as previously described.

Figure 7.32



We next click on the **Options** tab (Figure 7.33) to make sure that the **Descriptive Statistics** box is checked. We also need to specify the confidence level in the **Confidence interval (%)** box. In this example, we enter **95** in this box. Lastly, we click on the **Outputs** tab (Figure 7.34) to specify that we want to create a confidence interval for the population variance. We scroll down the options available in the **Quantitative data** box until we locate both the **Lower bound on variance (95%)** and the **Upper bound on variance (95%).** We make sure each box is checked. Click OK to create the confidence interval.

Figure 7.33

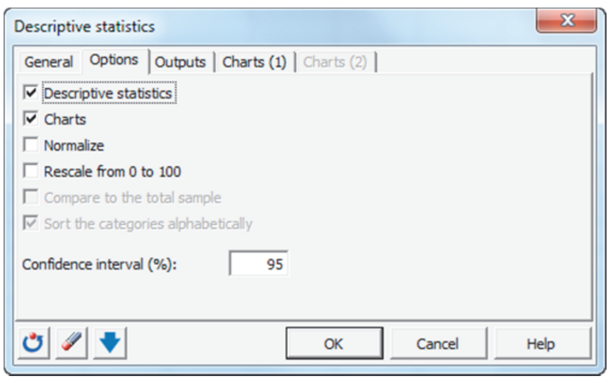
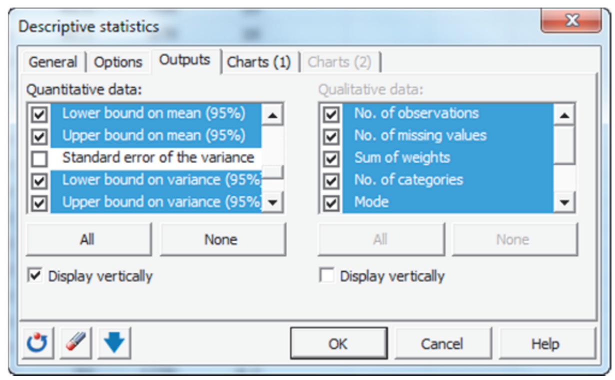


Figure 7.34



The confidence interval is shown below.

Figure 7.21

Statistic	WEIGHT
No. of observations	144
Mean	1049.7153
Variance (n-1)	141786.9743
Standard deviation (n-1)	376.5461
Lower bound on mean (95%)	987.6890
Upper bound on mean (95%)	1111.7416
Lower bound on variance (95%)	113908.9008
Upper bound on variance (95%)	181376.6013

We compare these values to the ones shown in the text. We note that the 95% confidence interval for the population variance is (113,908.9, 181,376.6). These values are different than the ones shown in the text. The reason for the difference is due to the estimation of the degrees of freedom in the chi-square table that was necessary for the calculations shown in the text (150 degrees of freedom were used in the table, when the real number should have been 143). The computer is able to use the real degrees of freedom, 143, and the result is that the numbers shown here are more accurate than the estimated values shown in the text.

7.7 Technology Lab

The Technology Lab consists of problems for the student to practice the techniques presented in each lesson. Each problem is taken from the homework exercises within the *Statistics* text and includes an **Excel** data set (when applicable) that should be used to create the desired output. The completed output has been included with each problem so that the student can verify that he/she is generating the correct output.

1. **Lipid profiles of hypertensive patients**. Hypertension is diagnosed if a patient's systolic blood pressure exceeds 140 mmHg and diastolic blood pressure exceeds 90 mmHg. A study of the lipid provides of hypertensive patients was carried out and the results published in Biology and Medicine (Vol. 2, 2010). Data on fasting blood sugar (milligrams/deciliter) and magnesium (milligrams/deciliter) in blood specimens collected from 50 patients diagnosed with hypertension were collected and are stored in the HYPER file. Biochemists used these data to establish a benchmark for fasting blood sugar (FBS) and magnesium (MAG levels in hypertensive patients. Create a 90% confidence interval for the mean fasting blood sugar and mean magnesium level using the z-distribution in the calculation of the intervals.

   Excel Output

90% confidence interval on the mean:	
FBS	(93.6805, 109.5595)
MAG	(1.9280, 1.9540)

2. **Characteristic of a rock fall**. . Refer to the Environmental Geology (Vol. 58, 2009) simulation study of how far a block from a collapsing rock wall will bounce down a soil slope. Rebound lengths (in meters) were estimated for 13 rock bounces. The data are repeated in the table and saved in the ROCKFALL data file.

   a. Create a 95% confidence interval for the population variance of all rebound lengths in the population.
   b. Create a 95% confidence interval for the population mean rebound length

   Excel Output

Statistic	REB-LENGTH
No. of observations	13
Lower bound on mean (95%)	7.2425
Upper bound on mean (95%)	12.1913
Lower bound on variance (95%)	8.6217
Upper bound on variance (95%)	45.6883

95% confidence interval on the mean:
(7.2425, 12.1913)

3. **Is Starbucks coffee overpriced?** The *Minneapolis Star Tribune* (August 12, 2008) reported that 73% of Americans say that Starbucks coffee is overpriced. The source of this information was a national telephone survey of 1,000 American adults conducted by Rasmussen Reports. Use this information to create a 95% confidence interval for the true proportion of all American adults who think that Starbucks coffee is overpriced.

Excel Output

```
95% confidence interval on the proportion (Wald):
                ( 0.7025,    0.7575 )
```