Chapter 5
Continuous Random Variables

5.1 Introduction

Chapter 5 introduces continuous random variables to the reader. Three continuous random variables, the normal, uniform, and exponential distribution, are introduced. In addition, a section on assessing the normality of a distribution is presented in Chapter 5.

Just as with the discrete distributions in Chapter 4, **Excel** offers various functions to calculate probabilities for both the normal and exponential continuous random variables in Chapter 5. In addition, **XLSTAT** offers the user the ability to assess the normality of a distribution of data through the use of its normal probability plot option.

It should be noted that the uniform random variable can be solved rather easily using simple mathematical techniques. Neither **Excel** nor **XLSTAT** offers any functions or techniques to calculate these probabilities. We also note that only the normal distribution and assessing the normality of a distribution are covered in *A First Course in Statistics* and these topics are covered in Chapter 4 of that text.

The following examples from *Statistics* are solved with **Excel** and **XLSTAT** in this chapter:

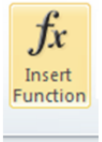
Excel Companion Exercise	**Page**	**Statistics Text**	**Excel File Name**
5.1	77	Example 5.5	
5.2	79	Example 5.9	
5.3	80	Example 5.7	
5.4	82	Example 5.11	
5.5	84	Example 5.12	EPAGAS
5.6	87	Example 5.14	

5.2 Calculating Normal Probabilities

Excel offers four different functions to work with normal random variables. The **NORM.S.DIST** and **NORM.S.INV** functions are only useful when working with standard normal distributions. The **NORM.DIST** and **NORM.INV** functions are used when working with any normal distributions. Because most normal applications utilize non-standard normal distributions, we will illustrate how to work with the **NORM.DIST** and **NORM.INV** functions. If a standard normal distribution is needed, the user can enter the values of the mean = 0 and the standard deviation = 1 and use these functions to generate the required information. As an alternative, the user could utilize the **NORM.S.DIST** and **NORMS.INV** functions when working with standard normal distributions. We will illustrate how to solve both standard normal and non-standard normal problems using the **NORM.DIST** and **NORM.INV** functions.

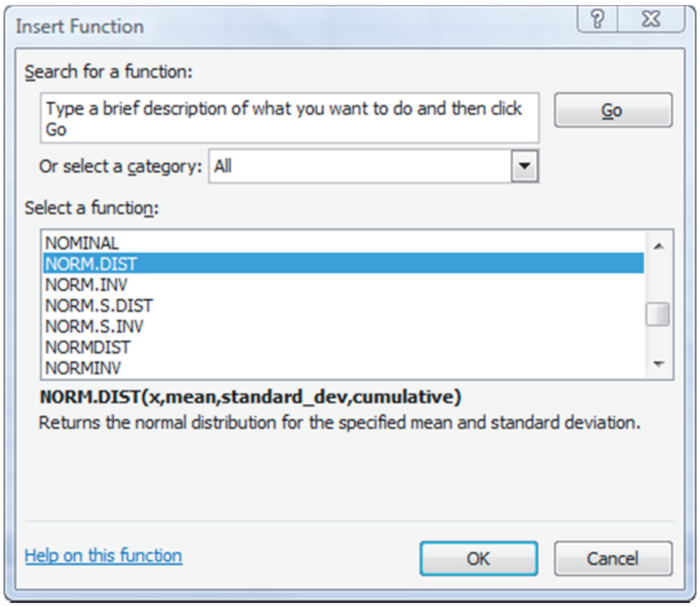
To use these functions within **Excel**, we begin by opening up **Excel** and placing the cursor on any cell in the blank worksheet. We click on the **Formulas** tab and then select the **Insert Function** icon shown in Figure 5.1.

Figure 5.1



This opens the **Insert Function** menu shown in Figure 5.2. Click on the arrow to select that **All** categories are being used and scroll down until you reach the **NORM.DIST** and **NORM.INV** functions. **Highlight** the option you wish to use and click **OK**.

Figure 5.2



The **NORM.DIST** function allows the user to determine the cumulative probability of a point in the normal distribution, given values of the mean and standard deviation of that distribution. The **NORM.INV** function allows the user to determine the point in the normal distribution at which a specified cumulative probability is achieved, given values of the mean and standard deviation of that distribution. It is up to the user to determine which function is appropriate based on the type of information is desired. We illustrate both procedures using four separate examples taken from the text.

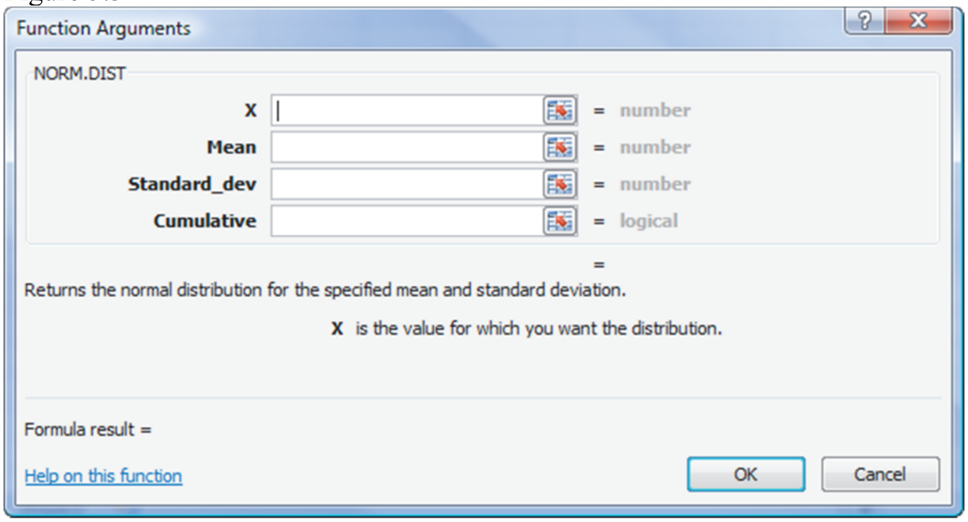
Exercise 5.1: We use Example 5.5 from the *Statistics* text:

Problem Find the probability that a standard normal random variable lies to the left of .67.

Solution:

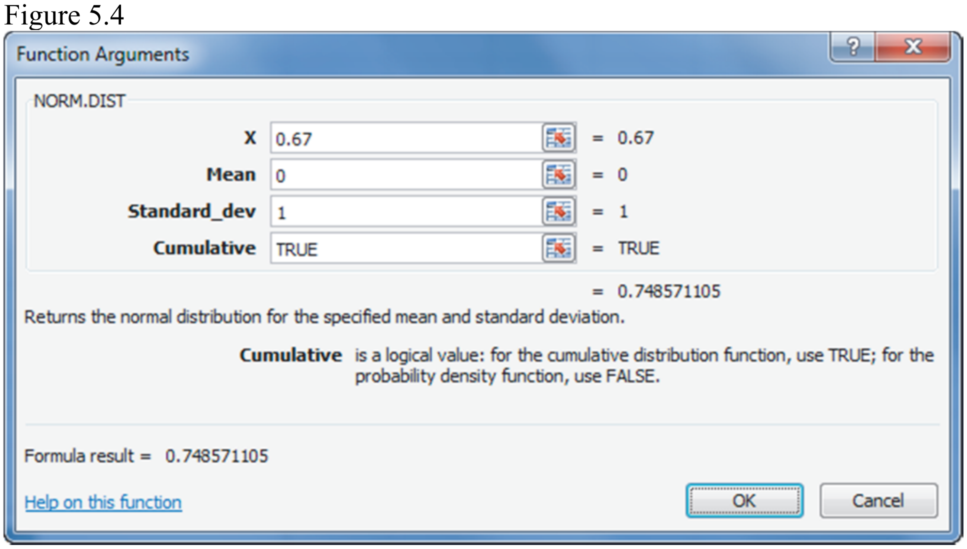
We first identify that we are working with the standard normal distribution with a mean of 0 and a standard deviation of 1. We are asked to determine the probability of selecting a value from that normal distribution that falls below the value 0.67. The **NORM.DIST** function is appropriate to use for this problem. We select the **NORM.DIST** function from our list of functions and click **OK**.

Figure 5.3



The **NORM.DIST** function requires us to identify values of the mean (**Mean=0**) and standard deviation (**Standard_dev=1**) of the standard normal distribution that we are working with. We also need to identify the value (**X=0.67**) in the distribution that we want a probability for. Lastly, we need to specify that we want to work with cumulative probabilities (probabilities that fall below) by entering **TRUE** in the **Cumulative** box in the menu. Figure 5.4 shows the completed function.

Figure 5.4



We note that the probability is shown in Figure 5.4 and then placed in our worksheet cell once we click **OK**. That probability is given as .748571105, which is the same value found using the normal table in the text.

It is important to remember that the **NORM.DIST** function will only report cumulative probabilities when the value **TRUE** is selected. While we desired to find $P(X < 0.67)$ in this problem, many problems ask us to find different types of probabilities. The user needs to be able to use the cumulative probabilities given by **Excel** to find other probabilities that are desired.

For example, if we wanted to determine $P(X > 0.67)$, we would take the cumulative probability found in **Excel**, and subtract it from the value one.

$P(X > 0.67) = 1 - P(X \leq 0.67) = 1 - 0.748571105 = 0.251428895$.

The **NORM.DIST** function can be used in this manner to determine any probabilities desired from normal distributions. To find specific values of a normal distribution that occur at a given probability, we use the **NORM.INV** function described in the following exercise.

Exercise 5.2: We use the Example 5.9 from the *Statistics* text:

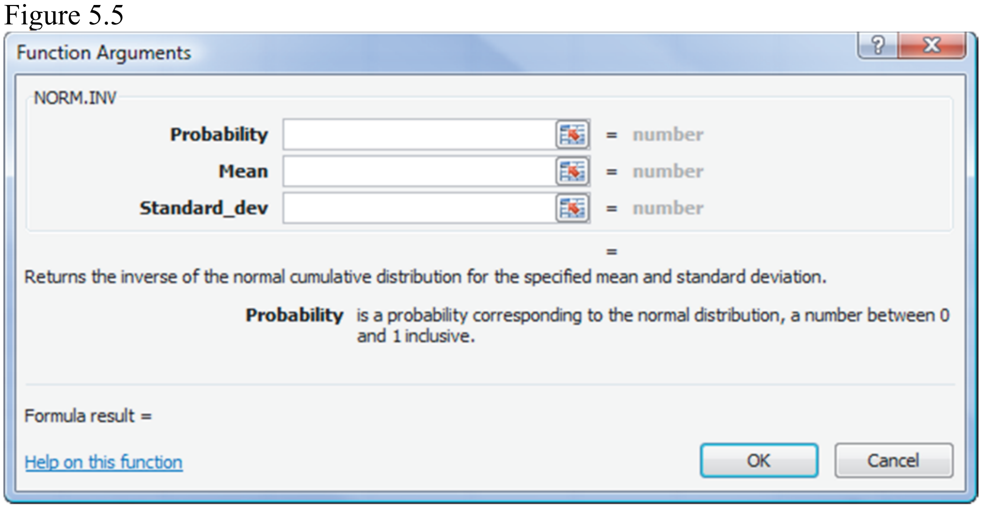
Problem Find the value of z – call it z_0 – in the standard normal distribution that will be exceeded only 10% of the time. That is, find z_0 such that $P(z \geq z_0) = .10$.

Solution:

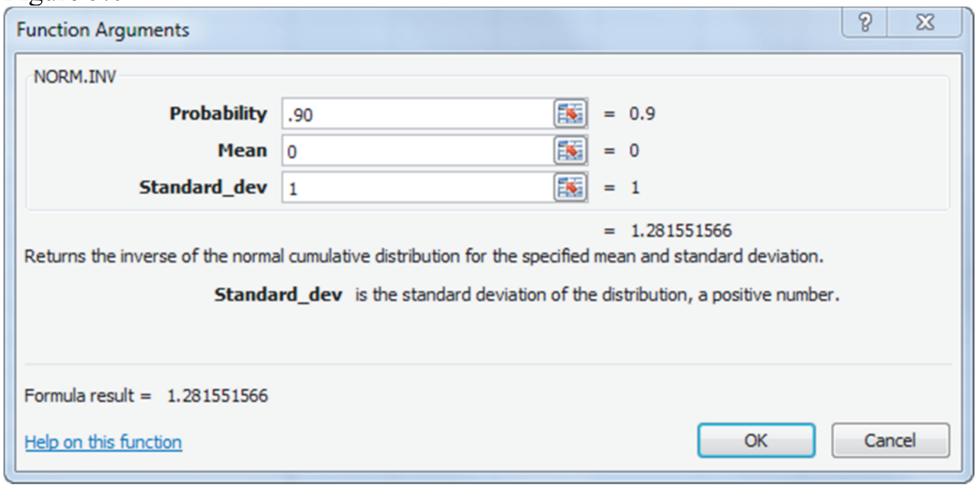
We first identify that we are working with the standard normal distribution with a mean of 0 and a standard deviation of 1. We are asked to determine the value in the distribution that has a probability of .10 to its right. In order to solve this problem, we need to identify the probability that falls below the point of interest. In this case, we use the .10 to the right to determine that the probability to the left will have to be .90.

To find the z-value we are looking for, we need to use the **NORM.INV** function available in **Excel**. We go through the steps previously detailed and select the **NORM.INV** function from our list of functions and click **OK**.

Figure 5.5



The **NORM.INV** function requires us to identify values of the mean (**Mean=0**) and standard deviation (**Standard_dev=1**) of the normal distribution that we are working with. We also need to identify the cumulative probability that applies to the point in the distribution that we are trying to find (**Probability=.90**). Figure 5.6 shows the completed function.

Figure 5.6


We note that the desired value is shown in Figure 5.6 and then placed in our worksheet cell once we click **OK**. That value is given as 1.281551566, which is the same value found using the normal table in the text.

It is important to note that the **NORM.INV** function only works with cumulative probabilities. When problems express other types of probabilities, the user must use the given probabilities to find the cumulative probabilities required in the **NORM.INV** function.

These first two examples have shown how to use the **NORM.DIST** and **NORM.INV** functions to solve standard normal probabilities. The process is exactly the same for non-standard normal distributions. We illustrate with the next two examples.

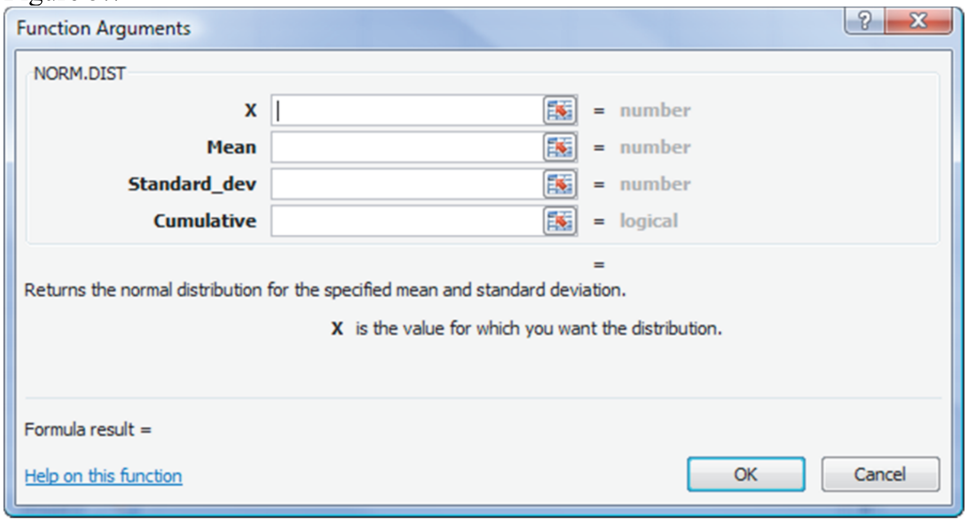
Exercise 5.3: We use Example 5.7 from the *Statistics* text:

Problem Assume that the length of time, x, between charges of a cellular phone is normal distributed with a mean of 10 hours and a standard deviation of 1.5 hours. Find the probability that the cell phone will last between 8 and 12 hours between charges.

Solution:

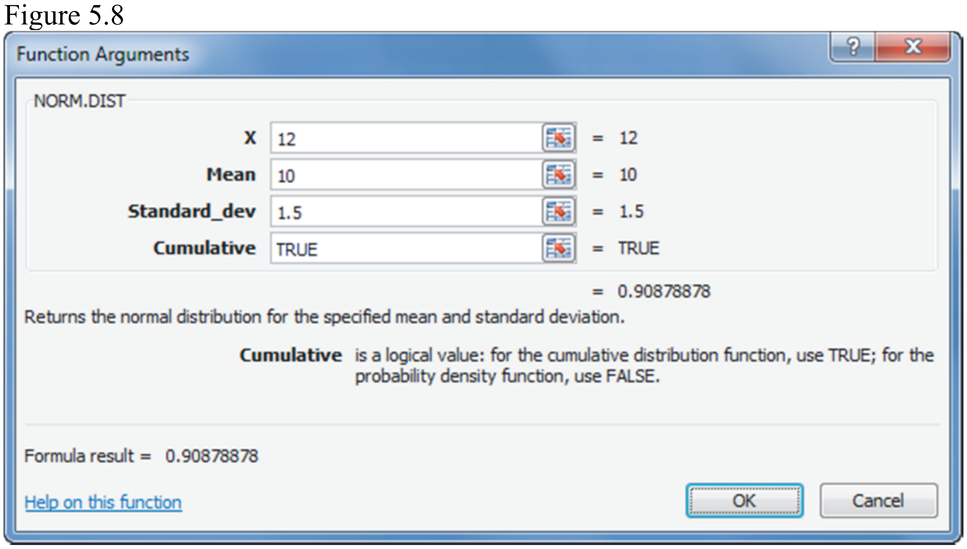
We first identify that we are working with the normal distribution with a mean of 10 and a standard deviation of 1.5. We are asked to determine the probability that the cell phone charge will last between 8 and 12 hours. To solve this problem, we will have to use the **NORM.DIST** function twice, once for a charge time of 12 hours and once for a charge time of 8 hours. We will need to subtract the resulting probabilities to get the desired answer. We begin by selecting the **NORM.DIST** function from our list of functions and click **OK**.

Figure 5.7



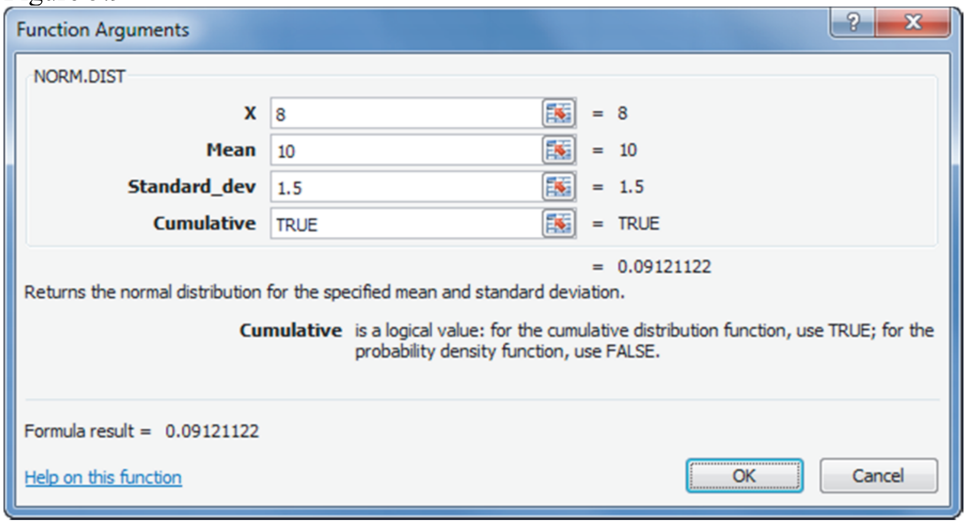
The **NORM.DIST** function requires us to identify values of the mean (**Mean=10**) and standard deviation (**Standard_dev=1.5**) of the normal distribution that we are working with. We also need to identify the value (**X=12**) in the distribution that we want probabilities for. Lastly, we need to specify that we want to work with cumulative probabilities (probabilities that fall below) by entering **TRUE** in the **Cumulative** box in the menu. Figure 5.8 shows the completed function.

Figure 5.8



We duplicate this process for the charge time of 8 hours to get the output shown in Figure 5.9.

Figure 5.9



To solve this problem, we simply subtract the two probabilities.

$P(8 < x < 12)$ = 0.90878878 - 0.09121122 = 0.81757756

We note that the probability we find here is slightly different than the probability shown in the text. This is due to slight rounding errors when the z-values in the text are calculated. The z-table in the text requires the user to round to two decimal places while **Excel** does not have that restriction.

The **NORM.DIST** function can be used in this manner to determine any probabilities desired from normal distributions. To find specific values of a normal distribution that occur at a given probability, we use the **NORM.INV** function described in the following exercise.

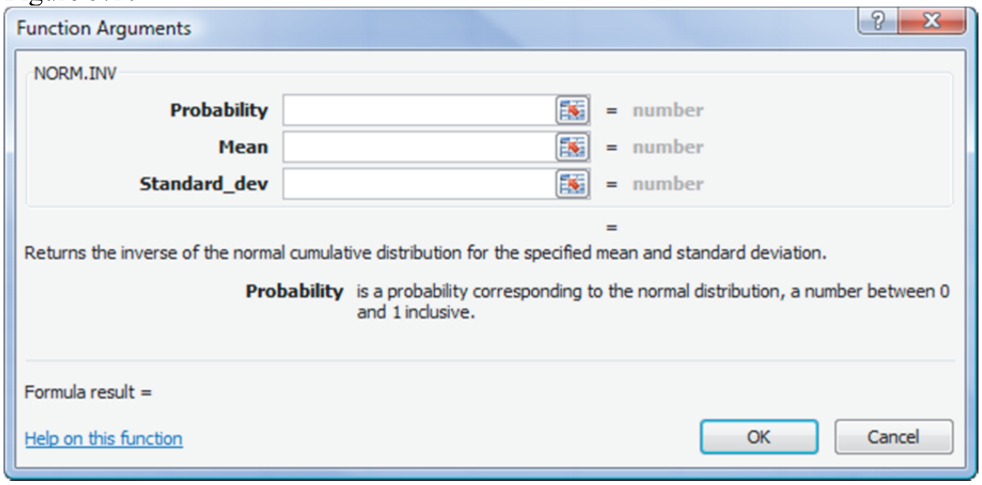
Exercise 5.4: We use the Example 5.11 from the *Statistics* text:

Problem Suppose the scores x on a college entrance examination are normally distributed with a mean of 550 and a standard deviation of 100. A certain prestigious university will consider for admission only those applicants whose score exceed the 90th percentile of the distribution. Find the minimum score an applicant must achieve in order to receive consideration for admission to the university.

Solution:

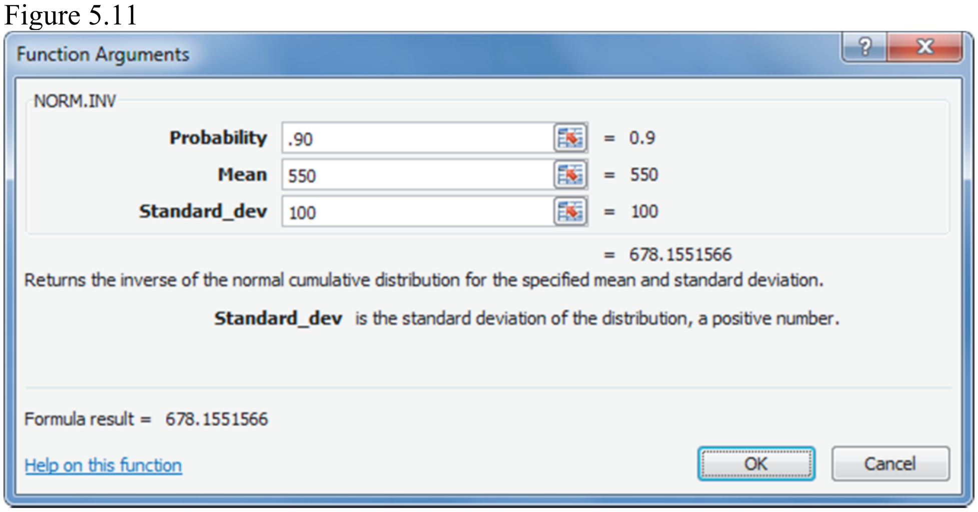
We first identify that we are working with the normal distribution with a mean of 550 and a standard deviation of 100. We are asked to determine the value in the distribution that has a probability of .90 to its left. The **NORM.INV** function is appropriate to use for this problem. We select the **NORM.INV** function from our list of functions and click **OK**.

Figure 5.10



The **NORM.INV** function requires us to identify values of the mean (**Mean=550**) and standard deviation (**Standard_dev=100**) of the normal distribution that we are working with. We also need to identify the cumulative probability that applies to the point in the distribution that we are trying to find (**Probability=.90**). Figure 5.11 shows the completed function.

Figure 5.11



We note that the desired value is shown in Figure 5.11 and then placed in our worksheet cell once we click OK. That value is given as 678.1551566, which is the same value found using the normal table in the text.

5.3 Assessing the Normality of a Data Set

XLSTAT offers the user a method of assessing whether a data set possesses a normal distribution. The **Normal Probability plot** utility creates a plot that enables the reader to determine the shape of the data. **The XLTAT Add-In** offers an easy method of creating this plot within the **Visualizing data** menu. We will utilize the **Univariate plots** option within this menu to solve the following problem.

Exercise 5.5: We utilize the EPA Gas Mileage Ratings for 100 Cars that is given in Example 5.12 of the *Statistics* text. The data is shown in Table 5.1. Construct a normal probability plot of the data and assess the shape of the EPA gas mileage ratings.

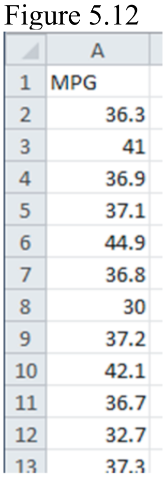
Table 5.1

EPA Gas Mileage Ratings for 100 Cars (miles per gallon)									
36.3	32.7	40.5	36.2	38.5	36.3	41.0	37.0	37.1	39.9
41.0	37.3	36.5	37.9	39.0	36.8	31.8	37.2	40.3	36.9
36.9	41.2	37.6	36.0	35.5	32.5	37.3	40.7	36.7	32.9
37.1	36.6	33.9	37.9	34.8	36.4	33.1	37.4	37.0	33.8
44.9	32.9	40.2	35.9	38.6	40.5	37.0	37.1	33.9	39.8
36.8	36.5	36.4	38.2	39.4	36.6	37.6	37.8	40.1	34.0
30.0	33.2	37.7	38.3	35.3	36.1	37.0	35.9	38.0	36.8
37.2	37.4	37.7	35.7	34.4	38.2	38.7	35.6	35.2	35.0
42.1	37.5	40.0	35.6	38.8	38.4	39.0	36.7	34.8	38.1
36.7	33.6	34.2	35.1	39.7	39.3	35.8	34.5	39.5	36.9

Solution:

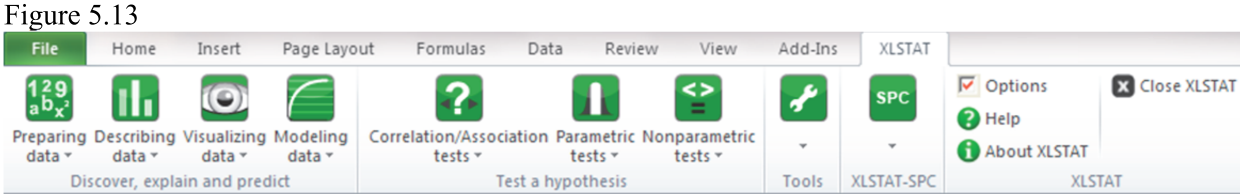
We solve this problem by using the **Visualizing data** menu within the **XLSTAT** program. Before we begin, we must access the data set for this example. **Open** the Data File **EPAGAS** by following the directions found in the preface of this manual. If done correctly, the data should appear in a workbook similar to that shown in Figure5.12. Use the mouse to select the data shown in the workbook.

Figure 5.12



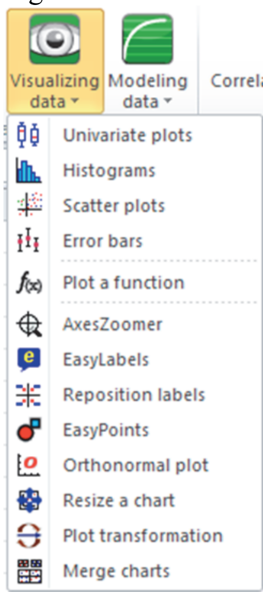
To create the desired displays, we click on the **XLSTAT** tab at the top of the **Excel** workbook to access the **XLSTAT** menus shown in Figure 5.13.

Figure 5.13



To generate the desired plots, we click on the **Visualizing data** menu and select the **Univariate plots** option shown in Figure 5.14.

Figure 5.14



This opens the **Univariate plots** menu shown in Figures 5.15 and 5.16. 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 **Univariate plots** 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 A, rows 2 – 101, 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, **MPG**. 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.

We click on the **Charts (1)** tab to select the type of chart to create. To create the normal probability plots presented in the text, we check the **Normal Q-Q plots** box. We click **OK** to generate the desired output shown in Figure 5.17.

Figure 5.15

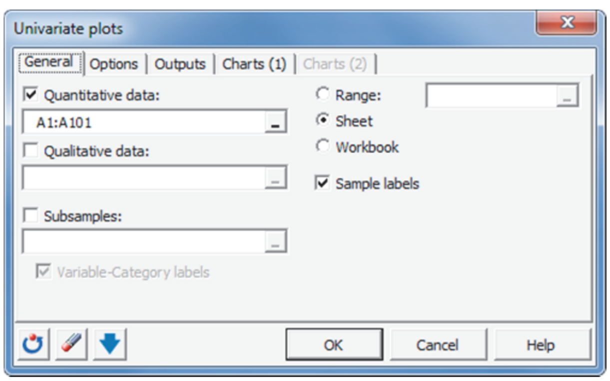


Figure 5.16

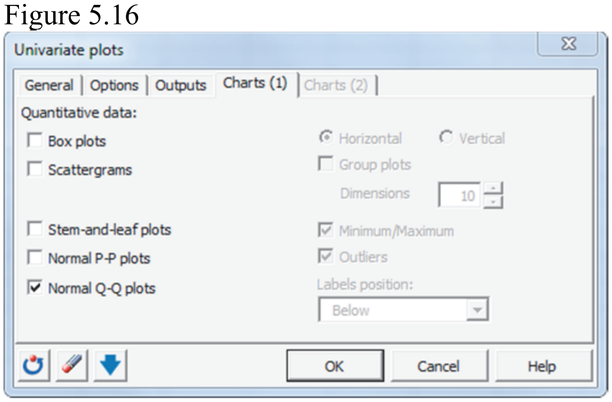
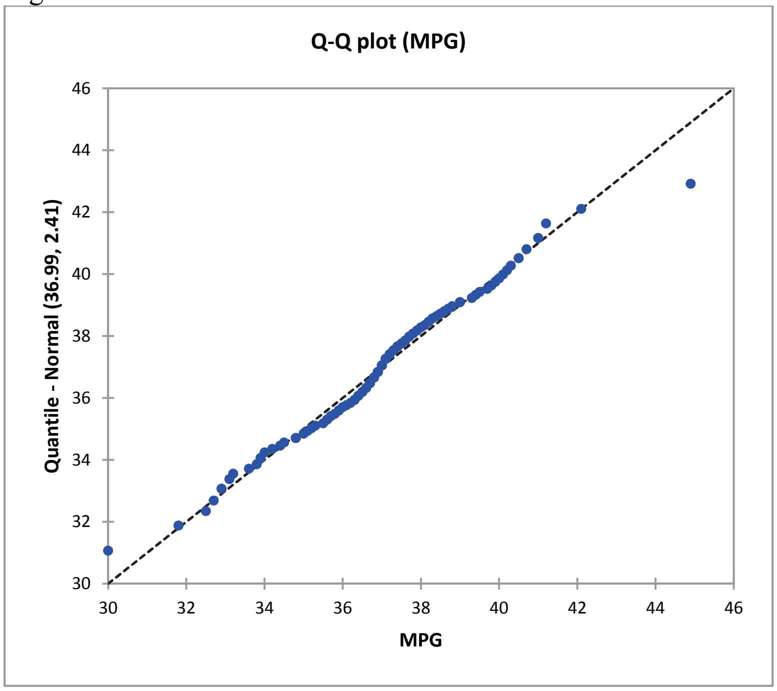


Figure 5.17

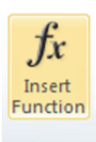


Compare this plot to the one shown in the text. The straight line shown in the plot indicates that the data are extremely normal.

5.4 Calculating Exponential Probabilities

To use the exponential probability tool within **Excel**, we begin by opening up **Excel** and placing the cursor on any cell in the blank worksheet. We click on the **Formulas** tab and then select the **Insert Function** icon shown in Figure 5.18

Figure 5.18



This opens the **Insert Function** menu shown in Figure 5.19. Click on the arrow to select that **All** categories are being used and scroll down until you reach the **EXPON.DIST** function. **Highlight** the function and click **OK**.

Figure 5.19

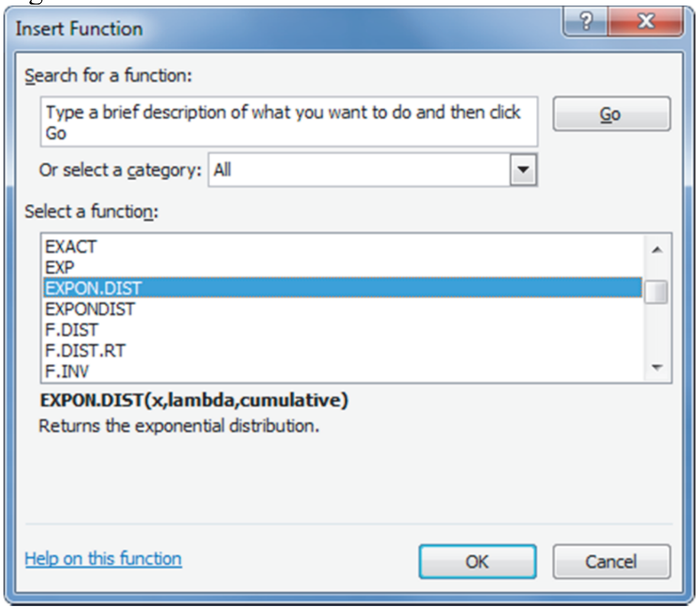
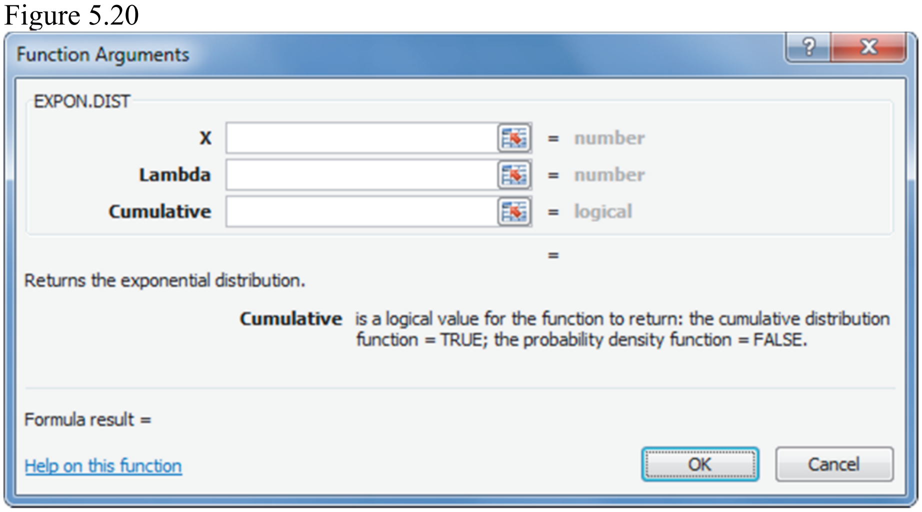


Figure 5.20



The **EXPON.DIST** function requires the user to enter a value of X to find a probability for, (**X**), the value of lambda (where $lambda = {}^{1}\!/_{mean}$) for the distribution (**Lamda),** and the type of probability desired (**Cumulative – either True or False**). For most applications, the cumulative probability option should be selected (**Cumulative = True**) in order to maximize the information that **Excel** will offer. **Click OK** to finish. We illustrate with the next example.

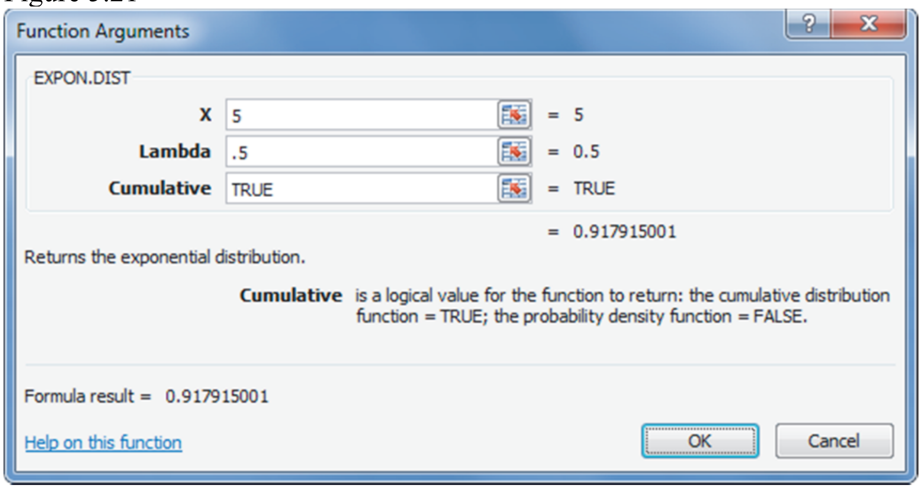
Exercise 5.6: As an example, we turn to Example 5.14 from the *Statistics* text.

Problem Suppose the length of time (in hours) between emergency arrivals at a certain hospital is models as an exponential distribution with a mean of 2. What is the probability that more than 5 hours pass without an emergency arrival?

Solution:

We utilize the **EXPON.DIST** function to solve this problem. We identify in the problem that the mean of the distribution is the value 2. Therefore we enter the value of lambda as ½ or .50 (**Lambda=.5)**. The value of X we want to determine a probability for is the value 5, so we enter **X**=5. Lastly, we enter **True** as the option in the Cumulative box to signify that we want a cumulative probability.

Figure 5.21



We note that the probability is shown in Figure 5.26 and then placed in our worksheet cell once we click OK. That probability is given as 0.917915. We note that this is a cumulative probability. In the problem, we wanted to find out the probability that more than 5 hours pass without an emergency. We use the computed probability to find:

$$P(X > 5) = 1 - P(X \leq 5) = 1 - .917915 = .082085$$

Compare this probability to the one shown in the text.

5.5 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. **Tomato as a taste modifier.** Miraculin – a protein naturally produced in a rare tropical fruit – can convert a sour taste into a sweet taste. Consequently, miraculin has the potential to be an alternative low-calorie sweetener. In *Plant Science* (May, 2010), a group of Japanese environmental scientists investigate the ability of a hybrid tomato plant to produce miraculin. For a particular generation of the tomato plant, the amount x of miraculin produced (measured in micro-grams per gram of fresh weight) had a mean of 105.3 and a standard deviation of 8.0. Assume that x is normally distributed.

   a. Find $P(X > 120)$
   b. Find $P(100 < x < 110)$
   c. Find the value a for which $P(X < a) = .25$

   Excel Output

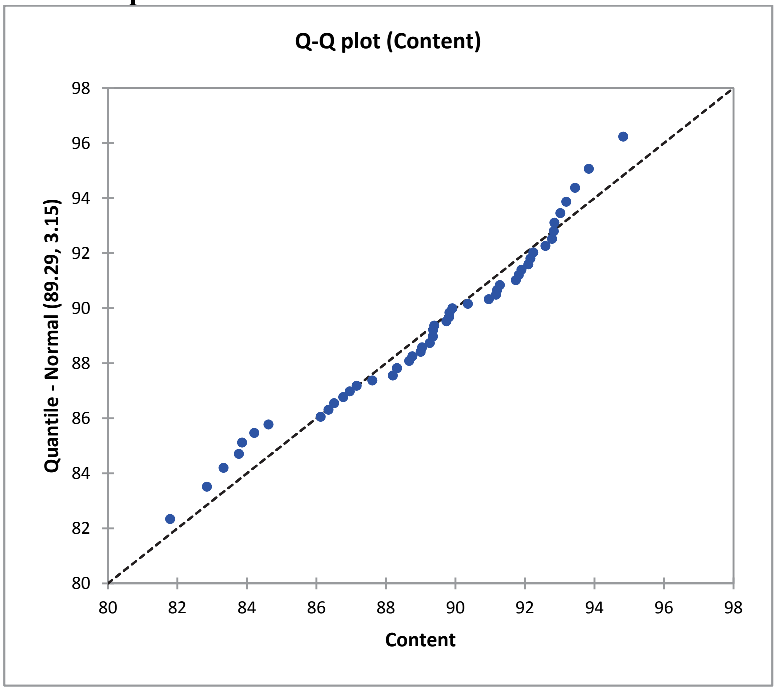
	A	B
1	1 a	0.033068
2	1 b	0.467741
3	1 c	99.90408

2. **Drug content assessment.** Scientists at GlaxoSmithKline Medicines Research Center used high-performance liquid chromatography (HPLC) to determine the amount of drug in a tablet produced by the company (*Analytical Chemistry*, Dec. 15, 2009). Drug concentrations (Measured as a percentage) for 50 randomly selected tables are listed in the accompanying table (data file **DRUGCON**).

91.28	88.32	89.39	88.32	88.2
92.83	91.17	89.82	88.76	92.78
89.35	83.86	89.91	89.26	86.35
91.9	89.74	92.16	90.36	93.84
82.85	92.24	88.67	87.16	91.2
94.83	92.59	89.35	91.74	93.44
89.83	84.21	86.51	86.12	86.77
89	89.36	89.04	92.1	83.77
84.62	90.96	91.82	83.33	93.19
86.96	92.85	93.02	87.61	81.79

   Create a normal probability plot for the data set and assess the normality of the data.

Excel Output



3. **Preventative maintenance tests.** The optimal scheduling of preventative maintenance tests of some (but not all) of n independently operating components was developed *in Reliability Engineering and System Safety* (Jan., 2006). The time (in hours) between failures of a component was approximated by an exponentially distribution with mean θ.

   a. Suppose θ =1,000 hours. Find the probability that the time between component failures ranges between 1,200 and 1,500 hours.
   b. Suppose θ=1,000 hours. Find the probability that the time between component failures is at least 1,200 hours.
   c. Given that the time between failures is at least 1,200 hours, what is the probability that the time between failures is less than 1,500 hours?

Excel Output

	A	B
1	3 a	0.078064
2	3 b	0.301194
3	3 c	0.713495