

# Chapter 4

## Normal Distributions

### Key Concepts

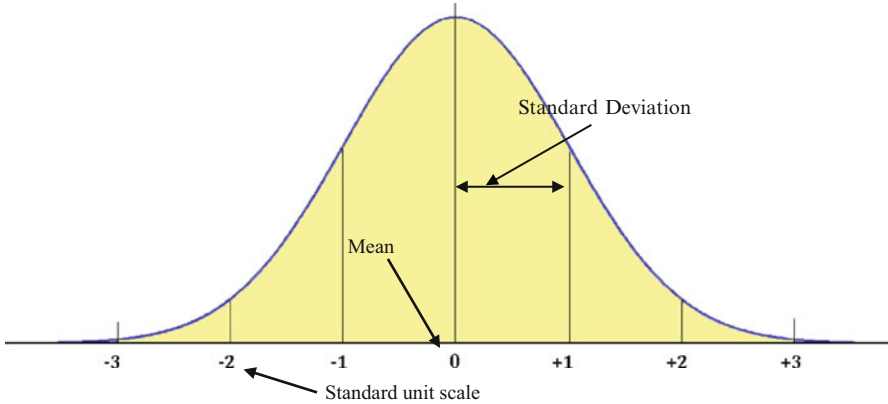
Chebyshev approximation, Normal distribution, Percentiles, Standard deviation, Standard units, and z-score.

### Discussion

This chapter will discuss one of the most common distributions in the business world, the “normal curve”. This is another example where the everyday usage of the word means something quite different from the statistical definition. In statistics the term “normal” refers to a specific mathematically defined curve that looks bell-shaped. An entire chapter is dedicated to this topic because of the normal distribution’s usefulness in many different applications throughout the business world. The area under the curve is often described as the probability and will be discussed in more detail in Chap. 8.

As a basis for comparison we provide a brief description of a non-normal technique for analysis, the Chebyshev approximation.

**Normal Distribution:** This is a bell-shaped or Gaussian distribution (Fig. 4.1).



**Fig. 4.1** Normal curve

The normal distribution is often described as a central region with a left hand tail and right hand tail. The boundary between the central region and the tails is usually defined by the particular problem you are trying to solve.

The **mean** and the **standard deviation** of the distribution describe the bell curve.

**Mean** ( $\bar{x}$ ): The arithmetic average value of the  $x$  variable. This pins down the center of the bell shape. When using the standard scale unit, the mean is always zero.

**Standard Deviation** ( $\sigma$ ,  $sd$ ,  $SD$ ): The  $sd$  defines the width of the bell shape for a sample.

$$sd = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

where:

$\Sigma$  is the capital Greek letter sigma, which indicates summing up all variables that follow it. The subscript and superscript indicate you should add the first value through the “nth” value. Often the short hand version just shows the summing sign without the sub/superscripts.

$x_i$  is the variable you are summing up the values for and the subscript is just a place holder. So if the subscript was “1”, it means the first value of  $x$ .

$\bar{x}$  is the average or mean value of  $x$

$n$  is the number of values you are summing together, more commonly referred to as the sample size

**Standard (std.) Units:** Several features of this distribution are very important.

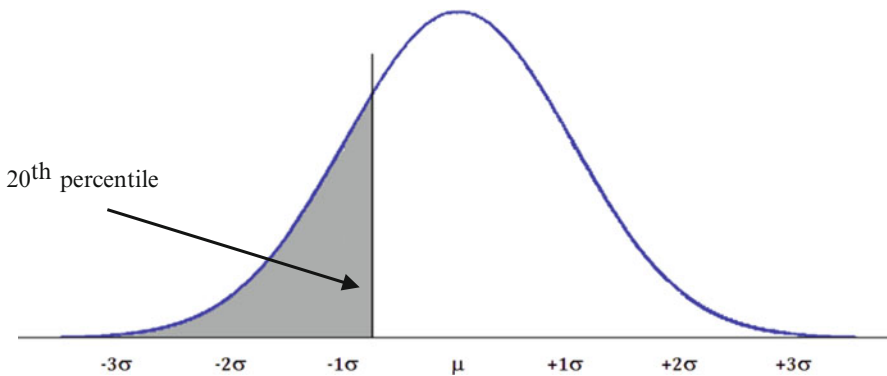
1. The graph is symmetric about 0. The part of the curve on the left side of 0 is the mirror image of the right side of the curve.
2. The total area under the curve is 100 %.

3. The area under the curve between  $-1$  and  $+1$  std. units is about 68 %.
4. The area under the curve between  $-2$  and  $+2$  std. units is about 95 %.
5. The area under the curve between  $-3$  and  $+3$  std. units is about 99 %.

An individual value is converted to standard units by seeing how many standard deviations it is above or below the average. Remember the standard deviation is a group statistic that allows you to convert individual values in the data set to standard units.

**z-score:** The z-score or standard score is found by converting a value from its measured units to standard units. In other words a z-score is the number of standard deviations a given value of  $x$  is above or below the mean. In addition, z-scores provide the key to using a table of standard normal probabilities to perform calculations related to normal distributions.

**Percentile:** The value of a variable, below which a certain percent of observations fall. For example, the 20th percentile is the value (or score) which indicates that 20 % of the observations are found below (or less than) that value.



**Chebyshev approximation:** This approximation can be used with any distribution. Not all distributions are normally distributed, or you simply may not know what the distribution looks like. To err on the side of conservative estimates, it is better to use the **Chebyshev approximation** in these cases. Because of its general application across all distributions, the answers are not as precise. By this, we mean that the answer you calculate will indicate the **MINIMUM** amount of data, in other words it defines the lowest value of how much data you should expect for your range of interest. So for example, if you want to know how much data falls between  $\pm 2$  standard deviations in your data set, the solution will indicate **at least** 75 % of the data will be located between these values, regardless of the distribution. For non-normal problems, we use the symbol  $k$  for the

standard deviation. Thanks to Chebyshev, we know that for any non-normal distribution:

At least  $(1 - 1/k^2)$  of the items in any data set will be within  $k$  standard deviations of the mean, where  $k$  is any value greater than 1.

- At least 75 % of the items must be within  $k = \pm 2$  standard deviations of the mean.
- At least 89 % of the items must be within  $k = \pm 3$  standard deviations of the mean.
- At least 94 % of the items must be within  $k = \pm 4$  standard deviations of the mean.

## Excel

When we have normally distributed data, the normal curve allows us to calculate how much data falls between any two points, and how much data exists in one or both tails. We can also easily calculate specific percentiles. But remember, we need to work in **standard units**. When you have all of the raw data, you can use Excel to calculate the standard deviation and the mean. However if you don't have the raw data you need to be given the mean and the standard deviation to use these functions.

The functions we will use in this section apply to both cases where you do not have the raw data and also to those cases where you do have all of the raw data available in Excel:

- **NORM.DIST** calculates the probability or distribution of data to the left of your value and requires the mean and standard deviation statistics. In other words, this calculates the percentile. The data is input using the following format.

*NORM.DIST (the **x value** in measured units, the **mean** in measured units, **standard deviation** in measured units, **1**)*

The argument 1 (TRUE) tells Excel to compute the normal cumulative distribution. If the last argument is 0 (FALSE), Excel returns the actual value of the normal random variable.

◇ The answer is given as a decimal so you will need to multiply by 100 to get the percentage answer.

- **NORM.INV** will convert the percentile value to measured units. The data is input using the following format.

*NORM.INV (the **percentile** to be converted to measured units and expressed as a decimal value, the **mean** in original measured units, **standard deviation** in original measured units)*

These Excel functions are summarized by way of examples in Fig. 4.2. In these examples we assume a dataset which has a normal distribution of employee ages.

Function	NORM.DIST		
	Input	Output	Required data
	Measured units	Percentile units	Mean, std dev for data
Example data	45 yrs old	60 <sup>th</sup> percentile (.60)	
Function	NORM.INV		
	Input	Output	Required data
	percentile	Measured units	Mean, std dev for data
Example data	87 <sup>th</sup> percentile(.87)	56 years old	

Fig. 4.2 Sample input/output

◇ Note that other Normal functions exist in Excel (i.e. NORM.S.DIST, NORM.S.INV) in which all input/output are in standard units. However, normal distribution analysis can be adequately handled using the NORM.DIST and NORM.INV functions.

The problem types using these Excel functions include:

1. **Percentile Calculation Problems (NORM.DIST) including creating area graphs**
2. **Converting Percentiles to Measured Units (NORM.INV)**
3. **Converting Measured units to z-scores (STANDARDIZE)**
4. **Rank and Percentile (RANK, PERCENTILE)**
5. **Non-normal Distributions**

### 1. Percentile Calculation Problems (NORM.DIST)

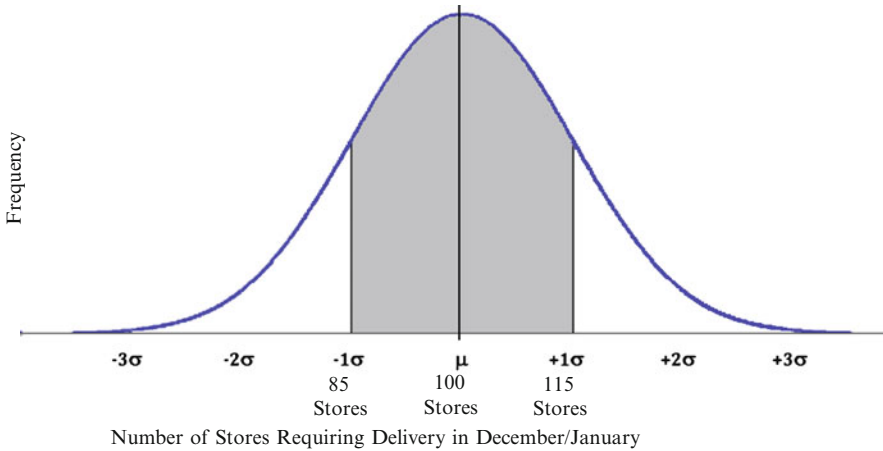
There are several distribution (area) types that use this function. These include:

- (a) Calculating the area to the left of a value
- (b) Calculating the area between two values
- (c) Calculating the area to the right of a value
- (d) Creating a normal distribution or area graph

#### Example Problem

Your transportation company provides delivery service 7 days a week to stores selling synthetic fire logs. You need to determine how many trucks and drivers to have on hand to deliver the product. The data you have collected from last year

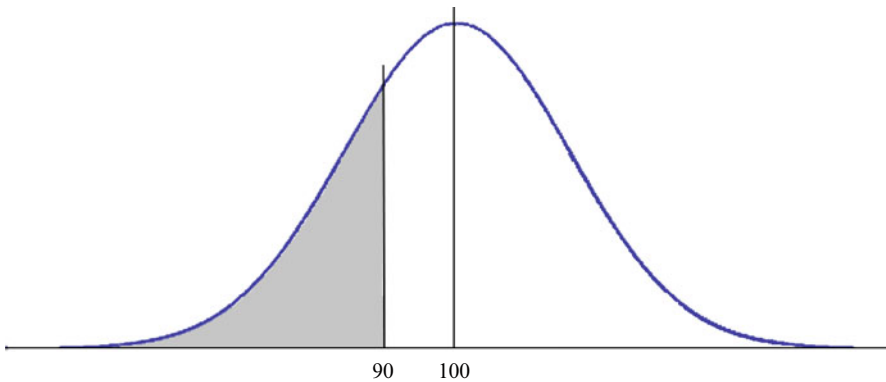
indicate the number of stores requiring deliveries on a daily basis during the months of December and January. This data seems to follow a normal distribution. The average number of stores requiring deliveries on any given day during this time is 100. The standard deviation is 15 stores.



- ◇ Note to answer these questions you don't actually need a data set as long as you know the data follows a normal distribution. Very powerful and helpful analysis tool.

(a) Calculating the area to the left of a value

*What % of the Time Did You Deliver to Less Than (<) 90 Stores During Last December and January? In Other Words How Much Data Is in the Left Hand Tail?*



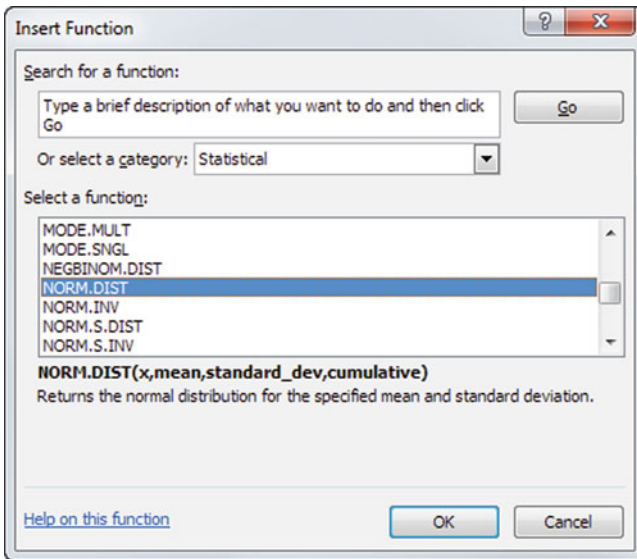
- Click on an empty cell where you want the answer to be output
- Click on the *fx* button



- Select the **Statistical** category



- Highlight the **NORM.DIST** function from the list or just type in the **NORM.DIST** function in the function bar



- Click **OK**
- Insert the value of X in original measured units. This is the upper limit of the area for which you want to calculate the distribution of data. In other words, this is the value you want to convert to a percentile



- Insert the **Mean** value



**Note:** You must already have the Mean value.

Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

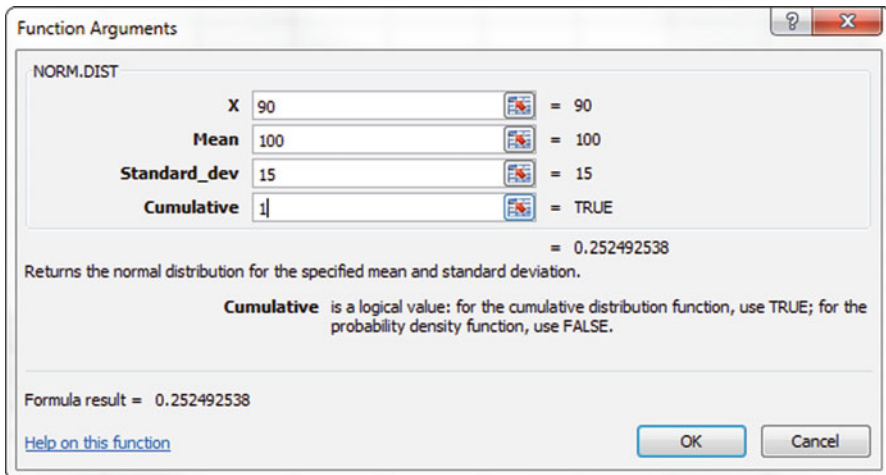
- Insert the **Standard\_dev** value



- Input the value of “**1**” or the word “**TRUE**” for normal cumulative calculations such as this one. Since this is a percentile calculation we select **1** as the last value in the function input.



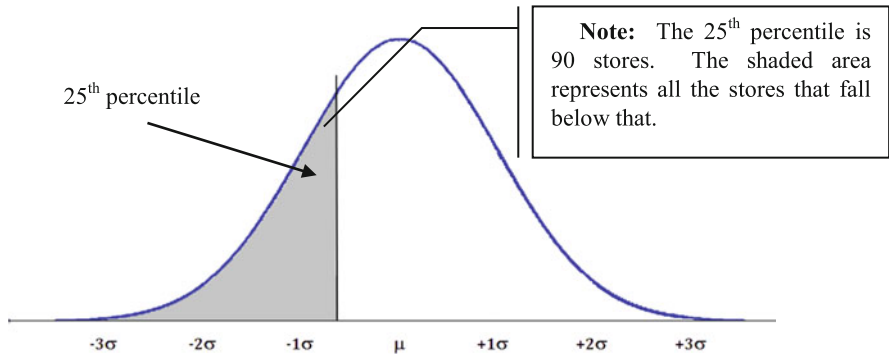
- Click **OK**



Based on this data, the answer is given as .252493 which translates to about 25 % of the time you delivered fire logs to less than 90 stores.

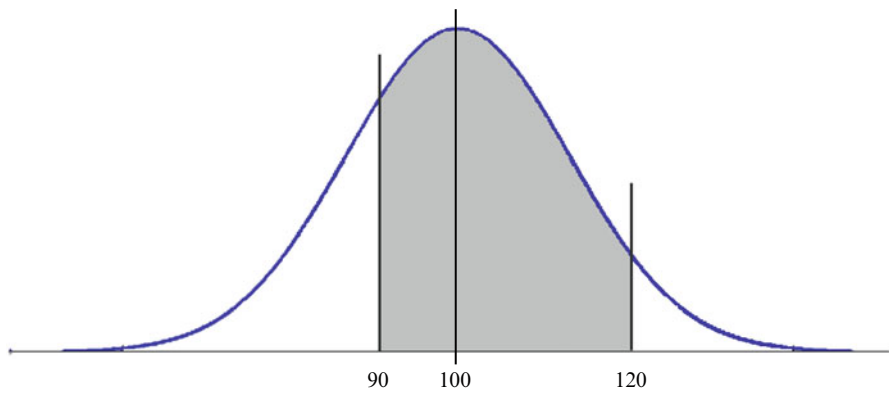


$f_x$ =NORM.DIST(90,100,15,1)			
D	E	F	G
	0.252493		



(b) Calculating the area between 2 values

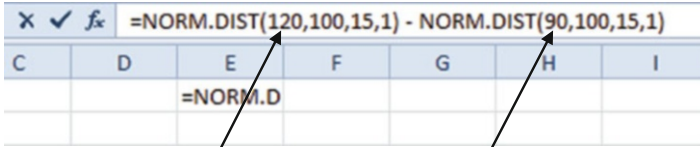
*What % of the Time Did You Deliver Fire Logs to Between 90 and 120 Stores During Last December and January?*



- Click on the empty cell where you want the answer
- Type “=” inside the  $f_x$  box (this prepares Excel to accept a function format)
- Type in the formula NORM.DIST (limit value, mean, sd, 1)

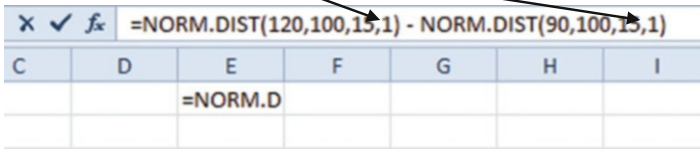
Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

- Input the upper limit value and the lower limit value of the area where you want to calculate the data distribution
- Input the mean and standard deviation from your data set inside both sets of

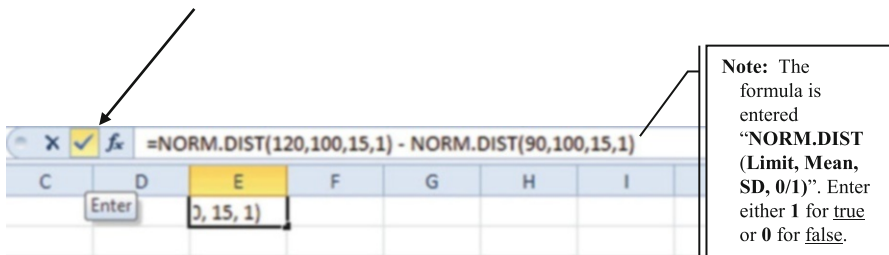


brackets as the next 2 inputs

- Input the value of **1** or the word **TRUE** for normal cumulative calculations inside both sets of brackets

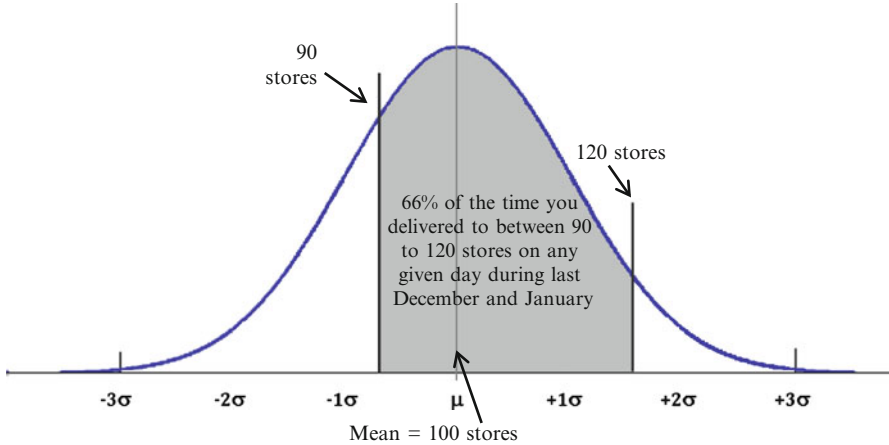


- Click the **checkmark** to the left of *fx*



The answer is given as 0.656296 which translates to during last December and January you delivered fire logs to between 90 and 120 stores about 66 % of the time.

fx =NORM.DIST(120,100,15,1) - NORM.DIST(90,100,15,1)					
D	E	F	G	H	I
	0.656296				



(c) Calculating the area to the right of a value

*What % of Time Did You Deliver to 130 or More Stores ( $\geq$ ) During Last December and January? In Other Words How Much Data Is in the Right Hand Tail?*

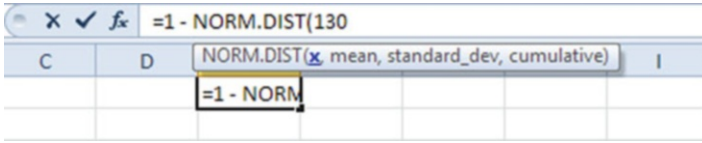
◆ Note in these type of problems the Excel answer will include when you are delivering to exactly 130 stores and when you are delivering to more than 130 stores.

- Click on an empty cell where you want the answer
- Click on the fx function

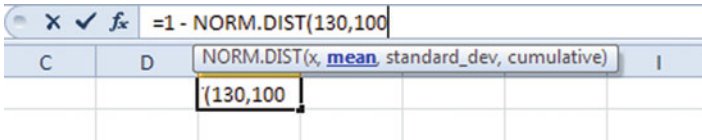


- Type the “=” in the white input box
- Type the number “1” which equate to 100 %
- Type the function or highlight the function from the drop down menu and subtract from 1
- Input the upper limit of the right hand tail

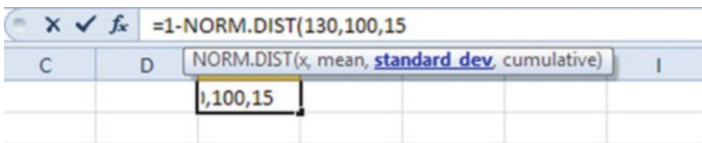
Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.



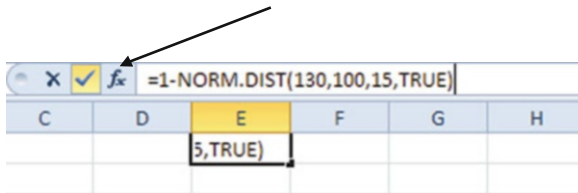
- Input the mean from your data set



- Input the standard deviation from your data set

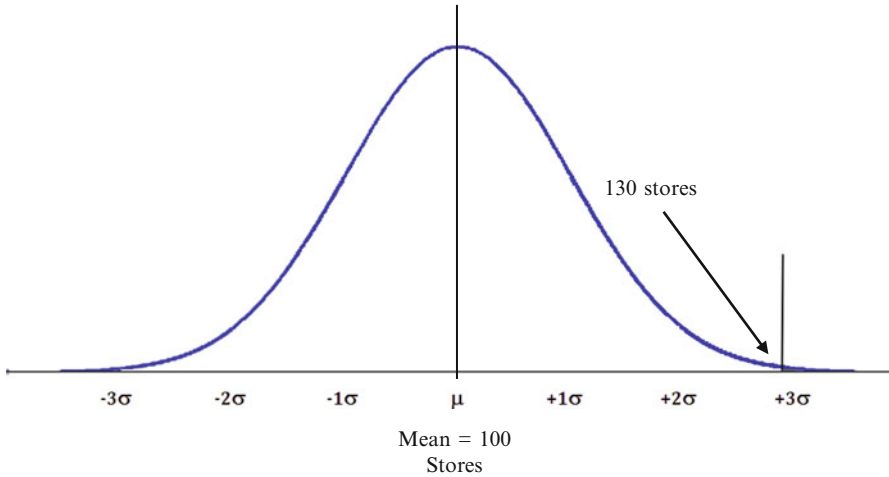


- Input the value of “1” or the word “TRUE” for the normal cumulative calculation
- Click the **checkmark** to the left of *fx*



The answer is given as 0.02275 which translates to about 2.3 % of the time you delivered fire logs to 130 or more stores on any given day. If we expect the same results this year, we could state that we will be delivering to *at least* 130 stores on any given day about 2.3 % of the time this coming December and January.

$f_x = 1 - \text{NORM.DIST}(130,100,15,1)$			
D	E	F	G
	0.02275		



(d) Graphing a normal distribution (Area Graph)

There are three steps in this process:

**Step 1. Sort the data into a sequential order (SORT)**

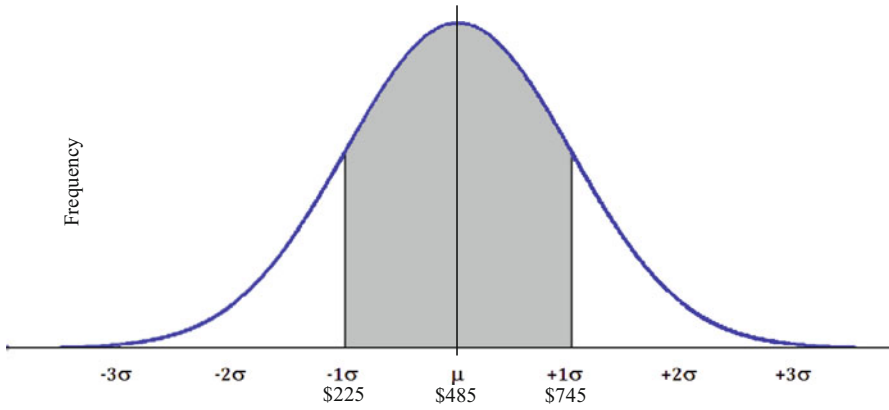
**Step 2. Calculate the height of the normal distribution for each x value (NORM.DIST)**

**Step 3. Create the graph**

**Example Problem**

You would like to graph the distribution of the cost of airplane tickets purchased by employees over the past 3 months. The average cost is \$485 and the standard deviation is \$260. You know that ticket prices are normally distributed based on historical data.

Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.



To do this in Excel will require several steps.

- Input the following ticket prices in column A

Cost of airplane tickets (\$)
505
385
65
585
825
465
785
345
865
625
265
545
905
145
745
425
185
665
225
705
305
105

Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

*Step 1*

Sort data into sequential order using the SORT function.

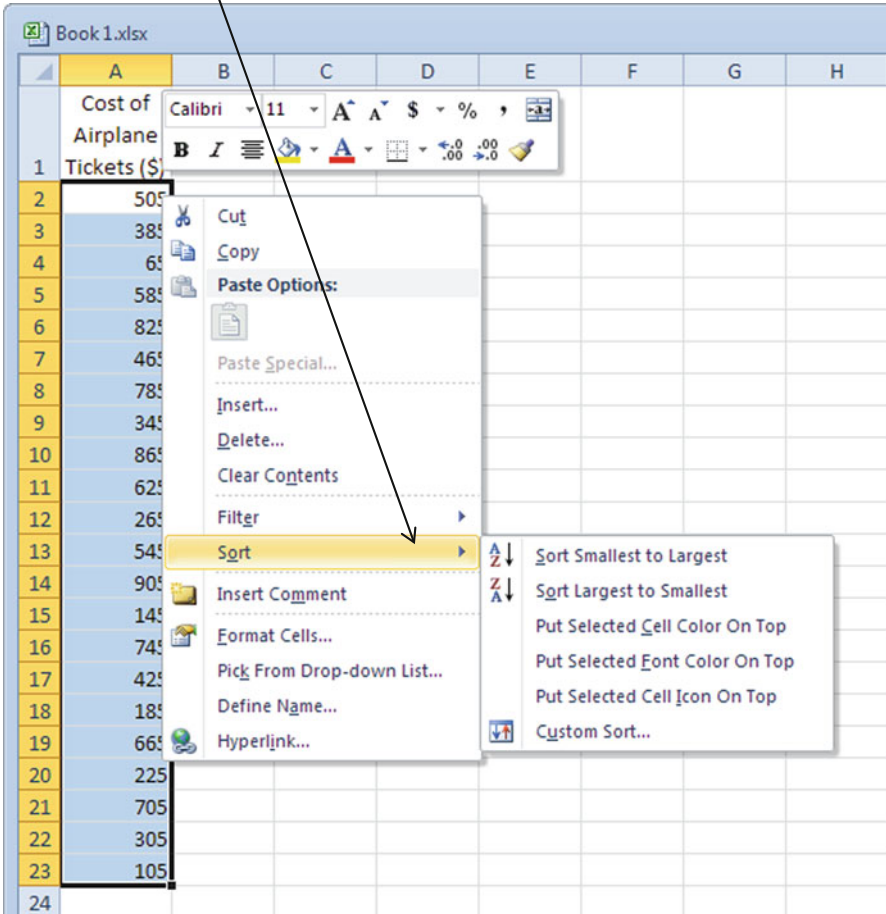
- Highlight the column of data to sort

	A	B
	Cost of Airplane	
1	Tickets (\$)	
2	505	
3	385	
4	65	
5	585	
6	825	
7	465	
8	785	
9	345	
10	865	
11	625	
12	265	
13	545	
14	905	
15	145	
16	745	
17	425	
18	185	
19	665	
20	225	
21	705	
22	305	
23	105	
24		

- ◇ If you want to preserve your original data you need to copy and paste it into another column. When Excel sorts the data, it over rides the original order of the data with the sorted order of data.

- Right click anywhere in the column

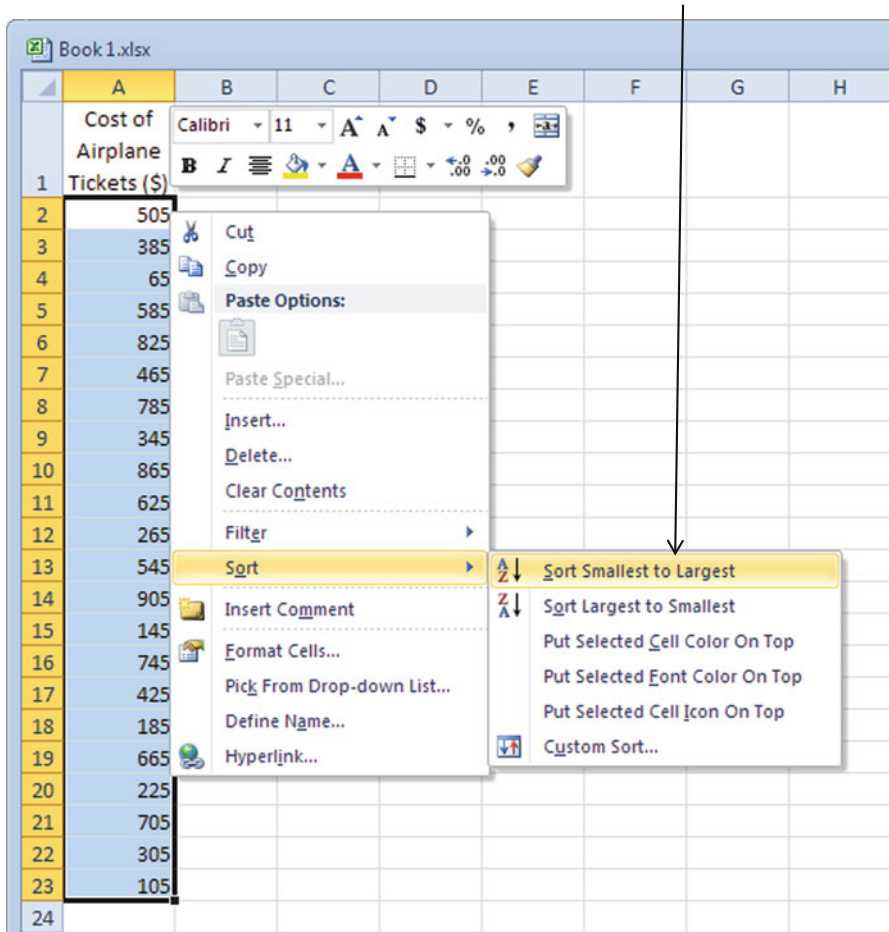
➤ Select the **SORT** function from the menu



Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.



- Select which sort option you want. In this case, choose “**Smallest to Largest**”

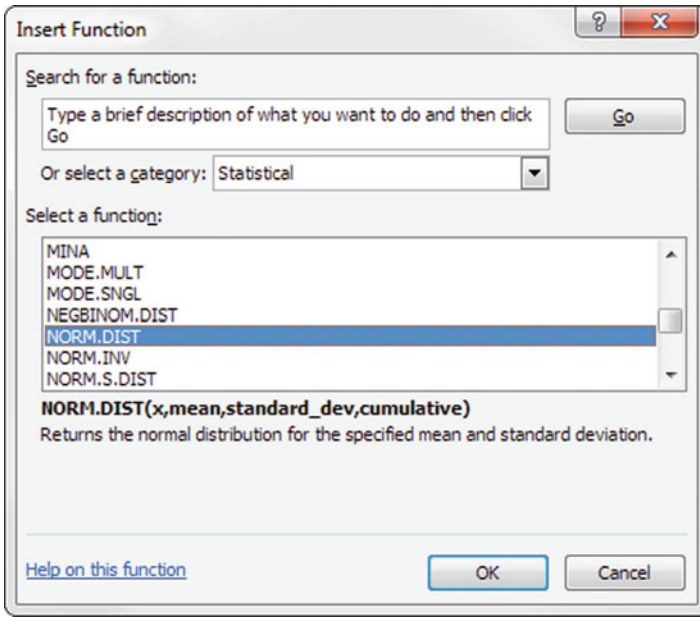


*Step 2*

Calculate the height of the normal distribution for each  $x$  value using the NORM.DIST function.

- Click on an empty cell where you want the output for the first value

- Click on the *fx* and select **NORM.DIST**



- Click **OK**
- Input the location (cell) of the first x value

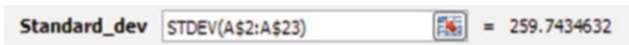


- Calculate the mean using the **AVERAGE** function



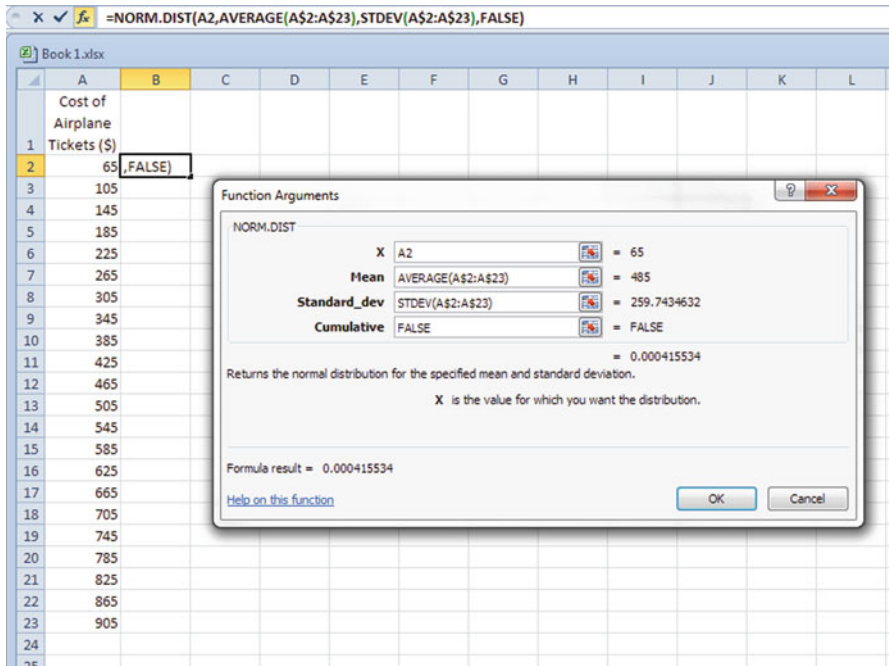
- ◆ Make sure to add "\$" with each cell location to treat these cells as constant values when you cut and paste them for use with the other data values.

- Calculate the std deviation using the **STDEV** function



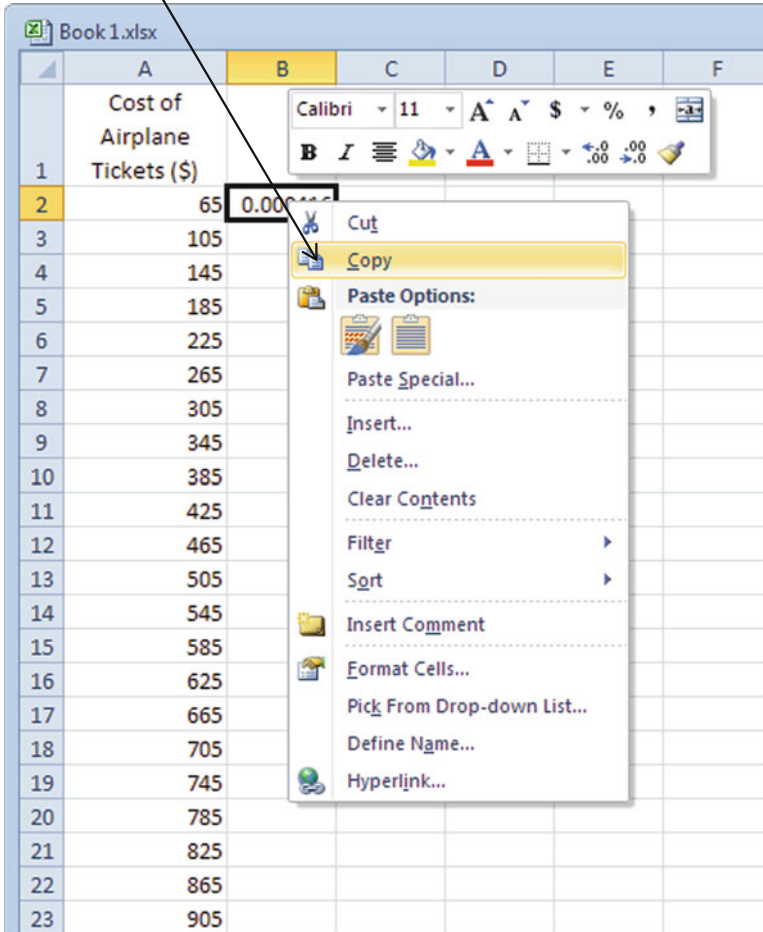
Examine the Cell Reference section in Chapter 2 for a refresher on “4” references.

- Input **0** (or **FALSE**) to calculate the height of the distribution for each ticket value in this normal distribution



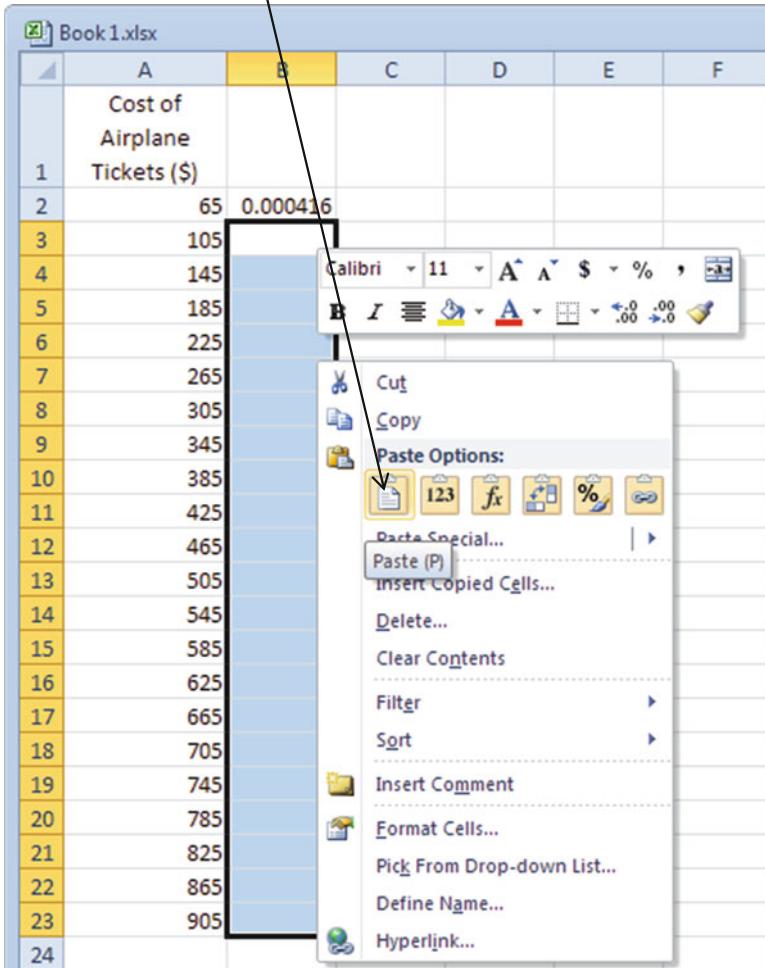
- Click **OK**

- Highlight cell B2 (or wherever your first calculation is located) and right click
- Select the **Copy** function



Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

- Highlight the empty cells in this same column across from the other ticket prices, right click, and select the **Paste** function



*Step 3*

Create the Graph.

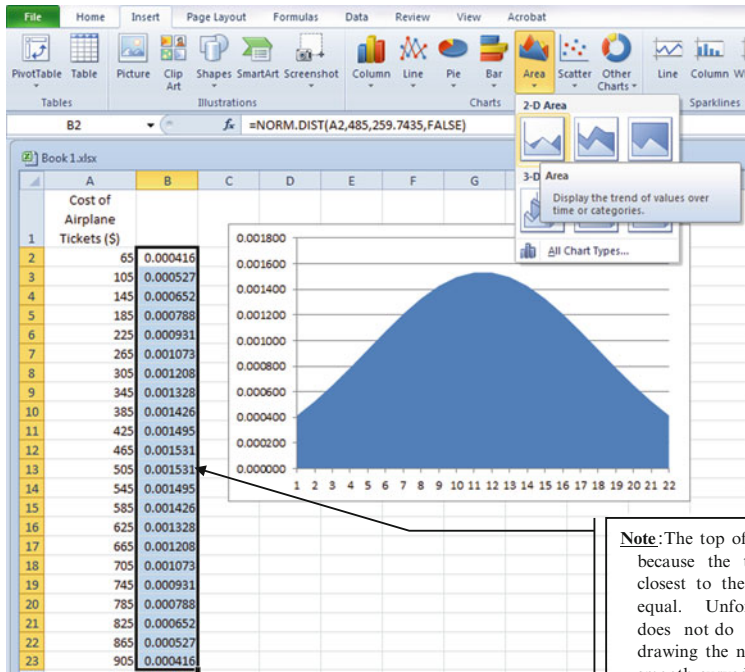
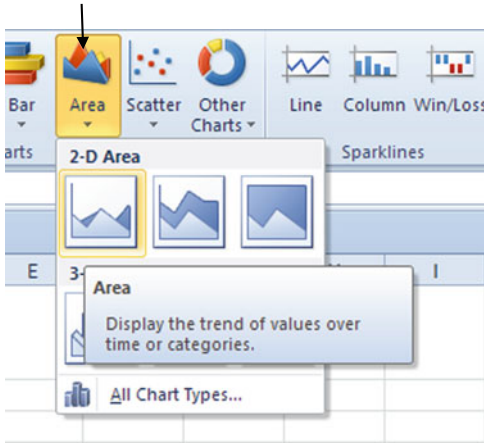
- Highlight the second column of values

	A	B	C
	Cost of Airplane Tickets (\$)		
1			
2	65	0.000416	
3	105	0.000527	
4	145	0.000652	
5	185	0.000788	
6	225	0.000931	
7	265	0.001073	
8	305	0.001208	
9	345	0.001328	
10	385	0.001426	
11	425	0.001495	
12	465	0.001531	
13	505	0.001531	
14	545	0.001495	
15	585	0.001426	
16	625	0.001328	
17	665	0.001208	
18	705	0.001073	
19	745	0.000931	
20	785	0.000788	
21	825	0.000652	
22	865	0.000527	
23	905	0.000416	

- Click the Insert tab



➤ Select the **Area** graph



**Note:** The top of the graph is flat because the two data values closest to the mean (485) are equal. Unfortunately Excel does not do a great job of drawing the normal curve with smooth curved lines.

Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

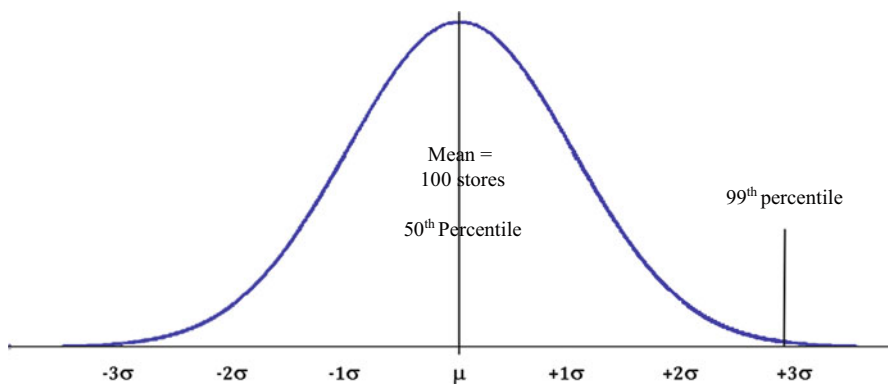
## 2. Converting Percentiles to Measured Units (NORM.INV)

Excel can convert percentile values in a normally distributed data set into measured units.

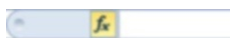
### Example Problem

Your distribution company provides delivery service 7 days a week to stores selling synthetic fire logs. You need to determine how many trucks and drivers to have on hand to deliver the product. The data you have collected from last year indicates the number of stores requiring deliveries on a daily basis during the months of December and January. This data seems to follow a normal distribution. The average number of stores requiring deliveries on any given day during this time is 100. The standard deviation is 15 stores. Let's analyze the data.

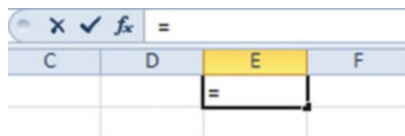
Calculate the Number of Stores Corresponding with the 99th Percentile



- Click on an empty cell where you want the answer
- Click on the **fx** function



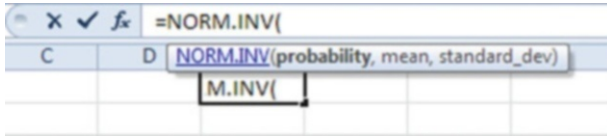
- Type "=" in the white input box



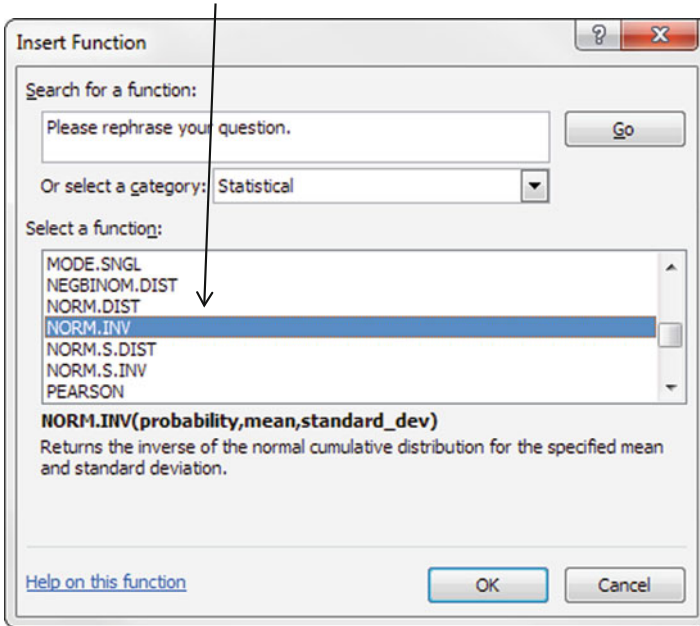
Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.



- Type the function



OR highlight the function from Insert Function window

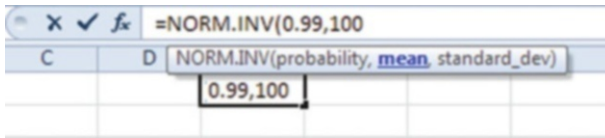


- Input the percentile value that you want to convert. **Input as a decimal value.** This is referred to as the probability by Excel.

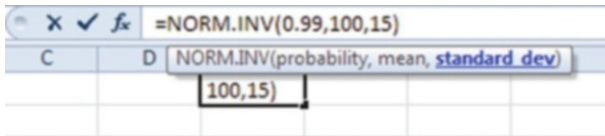


Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

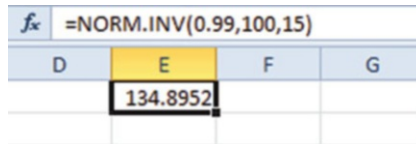
- Input the **mean** from your data set



- Input the **standard deviation** from your data set



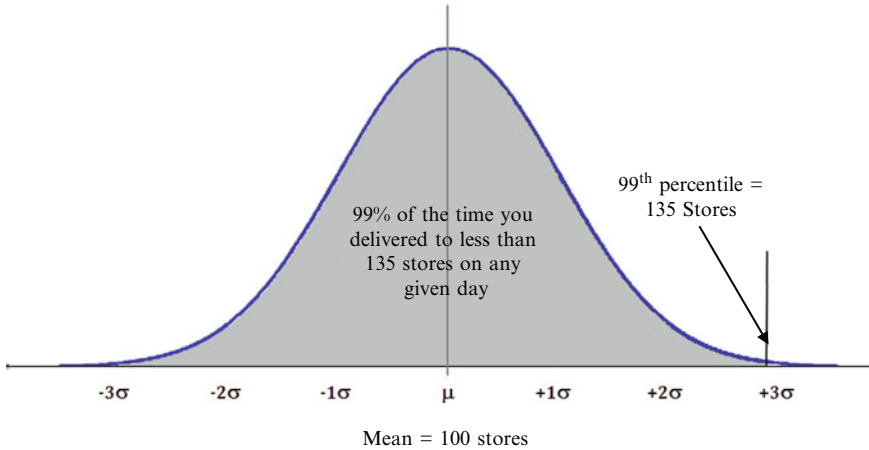
- Click the checkmark to the left of *fx*



The answer is given as 134.8952 which means that the 99th percentile for this data set is about 135 stores. In other words, about 135 stores or less required deliveries on any given day 99 % of the time period in December and January of last year.

- ◇ No matter what the decimal indicates, we need to round up, because we cannot have part of a store.

fx =NORM.INV(0.99,100,15)			
D	E	F	G
	134.8952		



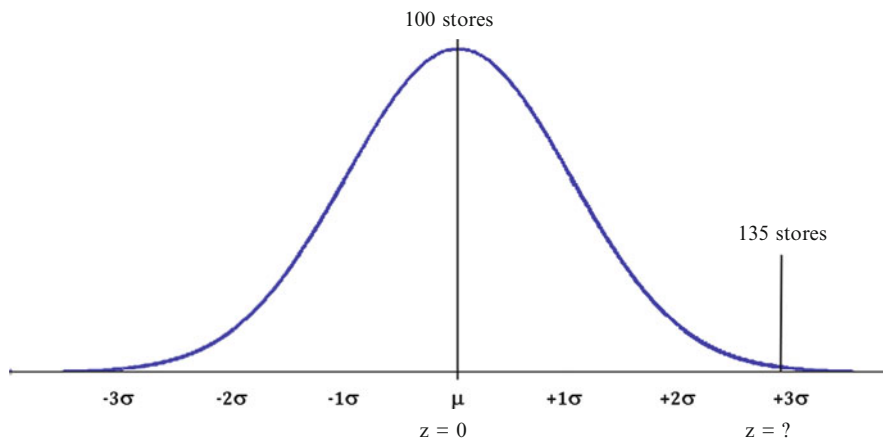
### 3. Converting Measured Units to z-Scores (STANDARDIZE)

#### Example Problem

Your distribution company provides delivery service 7 days a week to stores selling synthetic fire logs. You need to determine how many trucks and drivers to have on hand to deliver the product. The data you have collected from last year indicates the number of stores requiring deliveries on a daily basis during the months of December and January. This data seems to follow a normal distribution. The average number of stores requiring deliveries on any given day during this time is 100. The standard deviation is 15 stores.

Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

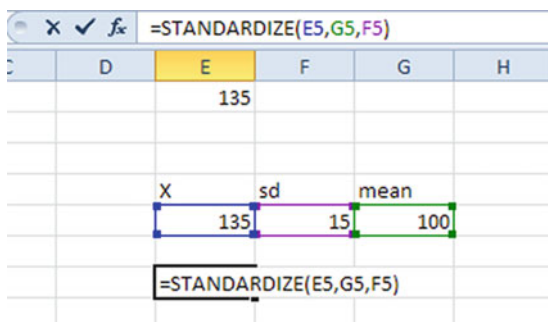
### Convert the Measured Value of 135 Stores to a z-Score



- Input in the **mean** number of stores as **100** into cell G5
  - Input the **standard deviation** as **15** stores into cell F5.
  - Input the value that we want to convert of **135** stores into cell E5.
- ◇ This problem could also be solved by directly typing in these required input values within the brackets of the function. The **STANDARDIZE** function uses the following format:

*STANDARDIZE(observed value in measured units to be converted to a z-score, mean of the data set, standard deviation of the data set)*

- Click on an empty cell where you want the answer
- Input “=” inside the **fx** formula box
- Type in the formula using cell locations for the observed value (x), mean and standard deviation. The formula divides the z-score by the sd and adds it back to the mean.



➤ Click on the checkmark



So in this case, 135 stores equates to a z- score of 2.326347 standard units. We should round the z-score to a value of approximately 2.33.

fx =STANDARDIZE(E5,G5,F5)				
D	E	F	G	H
	135			
	X	sd	mean	
	135	15	100	
	2.333333			

#### 4. Calculate Rank and Percentile (Rank and Percentile)

##### Example Problem

You would like to calculate the rank and percentile costs of airplane tickets purchased by employees over the past 3 months. The average cost is \$485 and the standard deviation is \$260. You know that the ticket prices are normally distributed based on historical data. Input the following ticket prices that are sorted from smallest to largest

Cost of airplane tickets (\$)
65
105
145
185
225
265
305
345
385
425
465
505
545

(continued)

(continued)

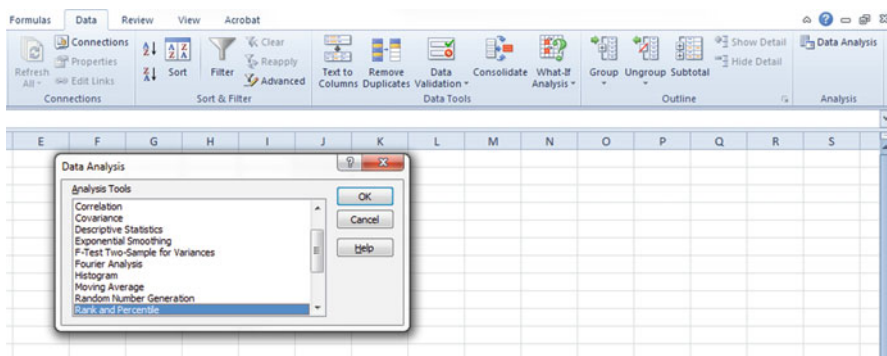
Cost of airplane tickets (\$)

585  
625  
665  
705  
745  
785  
825  
865  
905

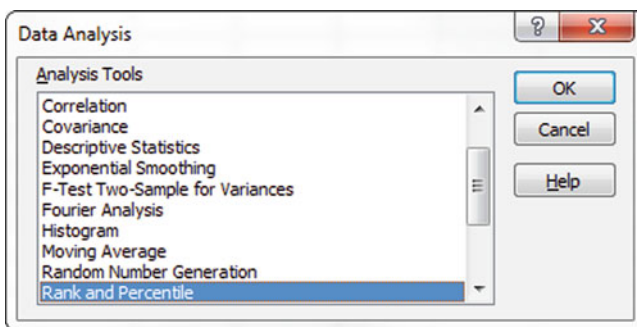
- Select the **Data** tab



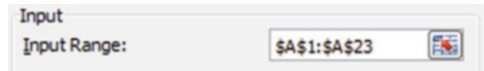
- Open the **Data Analysis** function



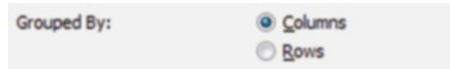
- Select the **Rank and Percentile** function



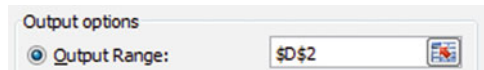
- Type in your column of x data



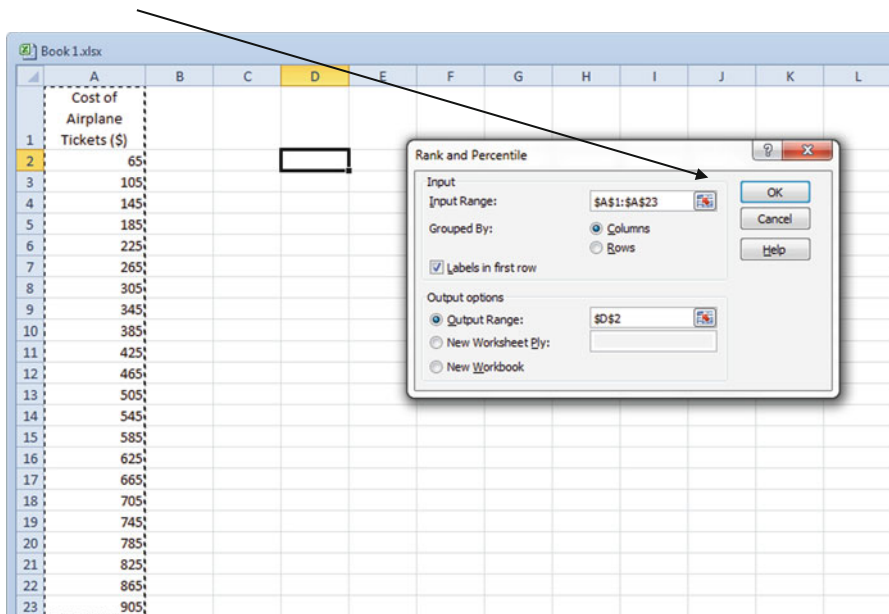
- Select the how your data is grouped: **rows** or **columns**



- Select an output location on your Excel spreadsheet



- Click **OK**



*Output*

- The first column of data is just the rank order from highest to lowest.
- The second column includes the actual data value.
- The third column has the **Rank** from lowest to highest.

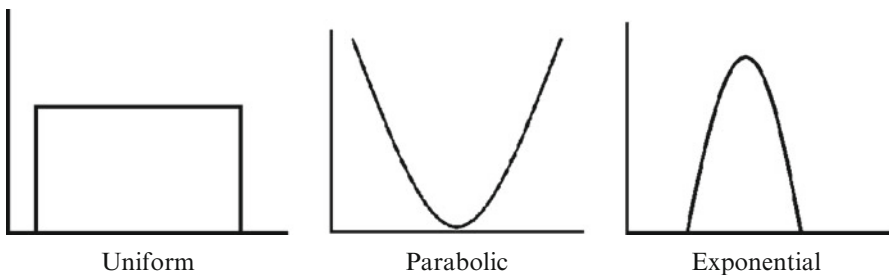
Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

- The final column is the **Percentile** moving from the highest value (right hand side of the distribution) to the lowest x value (left hand side of the distribution) in the distribution.

Point	Cost of airplane tickets (\$)	Rank	Percent
22	905	1	100.00 %
21	865	2	95.20 %
20	825	3	90.40 %
19	785	4	85.70 %
18	745	5	80.90 %
17	705	6	76.10 %
16	665	7	71.40 %
15	625	8	66.60 %
14	585	9	61.90 %
13	545	10	57.10 %
12	505	11	52.30 %
11	465	12	47.60 %
10	425	13	42.80 %
9	385	14	38.00 %
8	345	15	33.30 %
7	305	16	28.50 %
6	265	17	23.80 %
5	225	18	19.00 %
4	185	19	14.20 %
3	145	20	9.50 %
2	105	21	4.70 %
1	65	22	0.00 %

From the output we can determine that 47.6 % of the prices were at the mean (\$485) or less. Similarly we note that about 90 % are \$825 or less. The 10th percentile is approximately \$145. The ticket price of \$705 is the sixth most expensive ticket. There are numerous analytical conclusions that can be drawn from this simple data output.

### 5. Non-normal Distributions



Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.



Not all distributions are normally distributed, or you simply may not know what the distribution looks like. To err on the side of conservative estimates, it is better to use the Chebyshev approximation in these cases. At least  $(1 - 1/k^2)$  of the items in any data set will be within  $k$  standard deviations of the mean, where  $k$  is any value greater than 1.

- At least 75 % of the items must be within  $k = \pm 2$  standard deviations of the mean.
- At least 89 % of the items must be within  $k = \pm 3$  standard deviations of the mean.
- At least 94 % of the items must be within  $k = \pm 4$  standard deviations of the mean.

There are two steps in this type of problem:

**Step 1. Convert measured units to standard units (STANDARDIZE)**

**Step 2. Use Chebyshev approximation ( $fx$ )**

**Example Problem**

The data set of apartment rents has a mean of \$490.80 and a standard deviation of \$54.75. The data is not normally distributed.

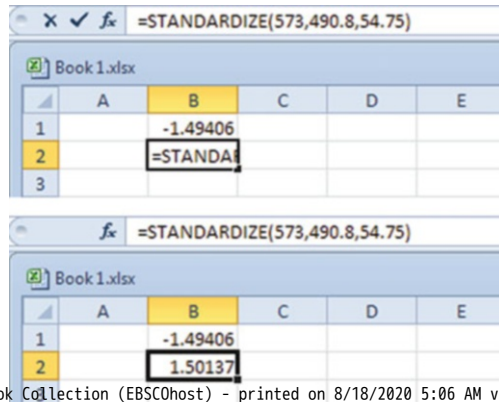
Calculate What Percentage of Rents Fall Between \$409 and \$573

- With this approximation, we can only calculate the minimum amount of data that would fall in the range of interest.

*Step 1: Convert the Measured Values to Standard Units*

For the first boundary value of 409

- Click on an empty cell where you want the answer
- Input “=” inside the  $fx$  formula box
- Type in the formula using cell locations for the observed value ( $x$ ) of 409, mean and standard deviation



Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

The resulting standardized value for 409 is  $-1.49406$  or approximately 1.5

For the second boundary value of 573

- Click on an empty cell where you want the answer
- Input “=” inside the *fx* formula box
- Type in the formula using cell locations for the observed value ( $x$ ) of 573, mean and standard deviation

The resulting standardized value for 573 is 1.50137 or 1.5

*Step 2: Use Chebyshev Approximation =  $1 - (1/(k)^2)$  where  $k$  is the boundary value in standard units*

- Click on an empty cell where you want the answer
- Input “=” inside the *fx* formula box
- Type in the formula using cell locations for the observed value ( $x$ ) in standard units

In this example the formula =  $1 - (1/(k)^2)$  is input to Excel as  
 $= 1 - (1/((-1.5)^2))$

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D
1		-1.49406	0.55	
2		1.50137		
3				
4				

The formula bar at the top shows the formula:  $=1-(1/(B1^2))$

Answer: At least 55 % of the rents fall between \$409 and \$573.

Note in this case the standard units were equal on either side of zero. For non-symmetric problems where the two  $z$  scores are different distances away from the mean you can perform a series of subtractions just as you would if the distribution was normal. For example if the lower  $z$  score was +1 and the upper  $z$  score was +2, calculate how much data falls between  $\pm 1$  using the Chebyshev formula and subtract that amount from what falls between  $\pm 2$ . Then divide by 2 so you have just the part greater than the mean. But this approximation only indicates the minimum amount of data that exists in this region.

## *Common Pitfalls*

- ◆ Many of the Excel functions only differ by one letter, so make sure you have the right normal distribution function.
- ◆ Remember, Excel gives percentages as decimals. Multiply by 100 to get a whole number percentage.
- ◆ Some versions of Excel may have a period in the function name so make sure you include this to get the right function.
- ◆ Chebyshev approximation does not work with small standard deviations (e.g., the sd must be greater than 1).
- ◆ Remember the Chebyshev approximation only calculates the minimum amount of data (at least. . .).
- ◆ NORM.INV does not provide the percent of distributions nor frequency on the y-axis. It is used for graphing purposes only.

## **Final Thoughts and Activities**

### *Practice Problems*

1. The data in the file **Amusementparks** contain the starting admission price (in \$) for one-day passes to 15 amusement parks in the United States.
  - (a) Compute the mean, range, variance, standard deviation.
  - (b) Compute the z-scores.
  - (c) Based on the results from (a) and (b), what conclusions can you reach concerning the starting admission price for one-day passes.
2. The data in the file **Candy** represent the cost per pound (\$) for a sample of candy.
  - (a) Compute the mean, range, standard deviation.
  - (b) Compute the z-scores.
  - (c) Is the data skewed? If so, how?
3. You are an operations analyst for AT&T. The length of long distance telephone calls may be normally distributed but you are not certain. You select a random sample of 25 calls. The average length is 8 min and the SD is 2 min. What % of these calls should be between 4.8 and 11.2 min?

### *Discussion Boards*

1. IQ scores are normally distributed with a mean of 100 and a standard deviation of 15. MENSA is an organization for people with high IQ's; eligibility requires

an IQ about 131.5. New York Times reporter Trish Hall notes that people are scoring substantially higher on IQ tests than in the past. Discuss.

2. Discuss some of the cultural and political factors that could affect a normal distribution of voter ages in government election results in different countries around the world.

### *Group Activity*

1. Queuing Theory is a branch of statistics that is very important to businesses such as McDonald's, Safeway and Disney. Disney conducts extensive studies of lines at its amusement parks to keep patrons happy and to assist with expansion plans. Discuss how all of this relates to the normal distribution.
2. Ergonomics is the study of problems associated with people adjusting to their environments. A case in point was faced by the US Air Force in studying the cockpits of fighter jets. These were originally designed for men but of course now we also have women pilots. Research information on the web in conjunction with a normal distribution approach to address this issue.

### **Parting Thought**

Smoking is one of the leading causes of statistics.

### **Problem Solutions**

1. The data in the file **Amusementparks** contain the starting admission price (in \$) for one-day passes to 15 amusement parks in the United States.
  - (a) Compute the mean, median, range, variance, standard deviation.

Admission	
Mean	49.80
Median	45.50
Standard deviation	11.10
Sample variance	123.29
Range	34.00
Count	10.00

(b) Compute the z-scores.

Location	Admission (\$)	z-score
Boo's Gardens	61	1.01
Animal Adventures	66	1.46
Dolly's Wonderland	44	-0.52
Chocolate Heaven Park	45	-0.43
Bob's Children's Park	32	-1.60
Mouse's Great Adventure	53	0.29
Fairy Lake Park	65	1.37
Silver Lining Park	46	-0.34
Adventure City Rides	43	-0.61
Mountain Explorer's Park	43	-0.61

(c) Based on the results from (a) and (b), what conclusions can you reach concerning the starting admission price for one-day passes?

The admission price for a one-day pass is slightly skewed to the right because the mean is slightly greater than the median.

2. The data in the file **Candy** represent the cost per pound of (\$) for a sample of candy.

(a) Compute the mean, median, range, standard deviation.

Cost(\$)	
Mean	1.68
Median	1.63
Standard Deviation	0.33
Sample Variance	0.11
Range	0.96
Count	14.00

(b) Compute the z-scores. Are there any outliers? Explain

Cost(\$)	z-score
1.43	-0.75
1.47	-0.63
1.67	-0.02
1.89	0.65
2.17	1.51
1.69	0.04
1.52	-0.48
1.32	-1.09
2.26	1.79
1.32	-1.09
1.30	-1.15
1.61	-0.20
2.16	1.48
1.65	-0.08

Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

There are no outliers because none of the Z-Scores has an absolute value that is greater than 3.0.

(c) Are the data skewed? If so, how?

The data appears to be skewed to the right because the mean is greater than the median.

3. Given: avg = 8 min sd = 2 min

- Convert 11.2 and 4.8 min to std units

Enter “=STANDARDIZE(11.2,8,2)” in a cell

Answer: 1.6

And enter “=STANDARDIZE(4.8,8,2)” into another cell

Answer: -1.6

- If  $k = 1.6$  AT LEAST  $\{1 - (1/k^2)\}$  will be within  $\pm 1.6$  standard deviations of the mean

Enter “=1-(1/(1.6^2))” in one cell

=1-(1/2.56) = .61 or AT LEAST 61 % of the calls will be between 4.8 and 11.2 min

The screenshot shows an Excel spreadsheet titled 'Book 1.xlsx' with a grid of cells. The grid has 4 columns labeled A, B, C, and D, and 3 rows labeled 1, 2, and 3. The values in the cells are as follows:

	A	B	C	D
1		-1.6	0.61	
2		1.6	0.61	
3				