Statistical Analysis

JOHNS HOPKINS
CAREY BUSINESS SCHOOL

Some of the slides in this presentation are from Bowerman, B. L., O'Connell, R. T., & Murphree, E.S., (2010). *Business Statistics in Practice* (6th Ed.), Copyright © McGraw-Hill Education.

Continuous Random Variables

1. Continuous Probability Distributions
2. The Uniform Distribution
3. The Normal Probability Distribution
4. The Exponential Distribution

Continuous Probability Distributions

A continuous random variable may assume any numerical value in one or more intervals

Use a continuous probability distribution to assign probabilities to intervals of values

Continuous Probability Distributions *continued*

The curve $f(x)$ is the continuous probability distribution of the random variable x if the probability that x will be in a specified interval of numbers is the area under the curve $f(x)$ corresponding to the interval

Other names for a continuous probability distribution:

- Probability curve
- Probability density function

Properties of Continuous Probability Distributions

Properties of $f(x)$: $f(x)$ is a continuous function such that

1. $f(x) \geq 0$ for all x
2. The total area under the $f(x)$ curve is equal to 1

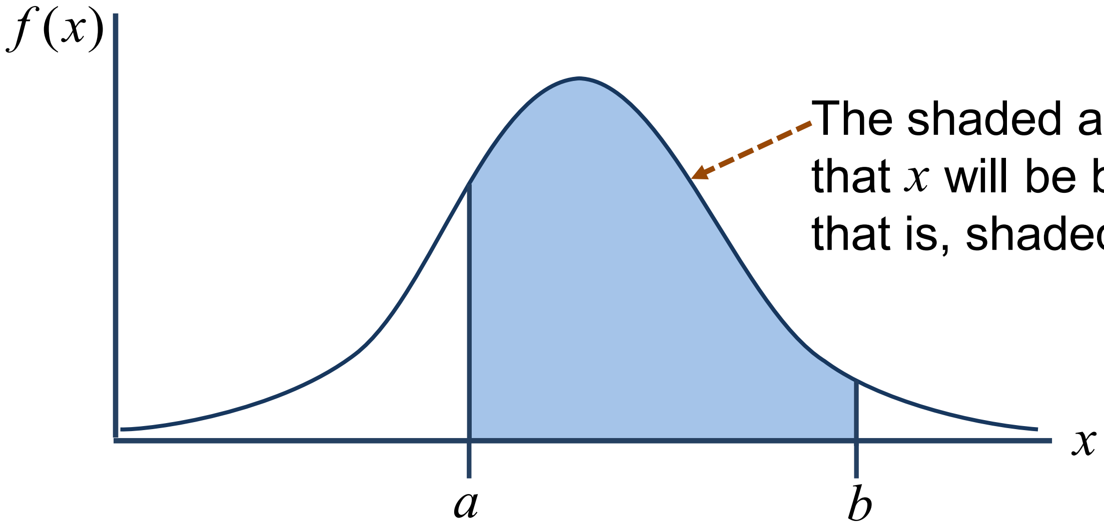
Essential point: An area under a continuous probability distribution is a probability

Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Area and Probability

The blue area under the curve $f(x)$ from $x = a$ to $x = b$ is the probability that x could take any value in the range a to b

- Symbolized as $P(a \leq x \leq b)$
- Or as $P(a < x < b)$, because each of the interval endpoints has a probability of 0

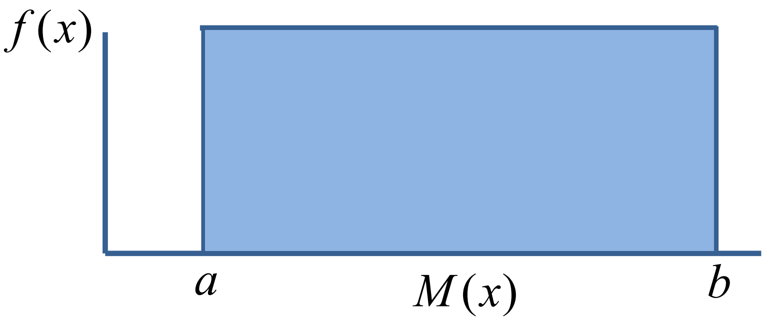


The shaded area is the probability that x will be between a and b, that is, shaded area $= P(a \leq x \leq b)$

Distribution Shapes

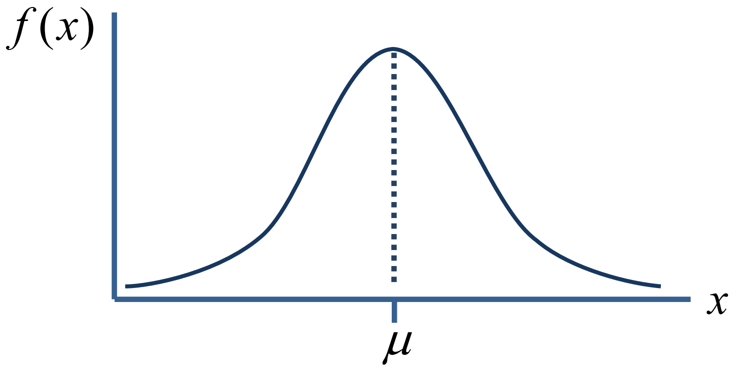
Symmetrical and rectangular

- The uniform distribution



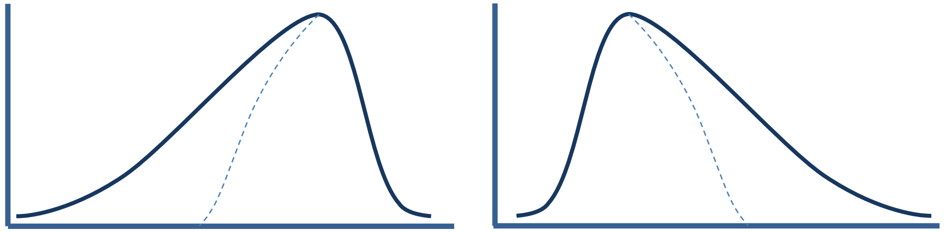
Symmetrical and bell-shaped

- The normal distribution



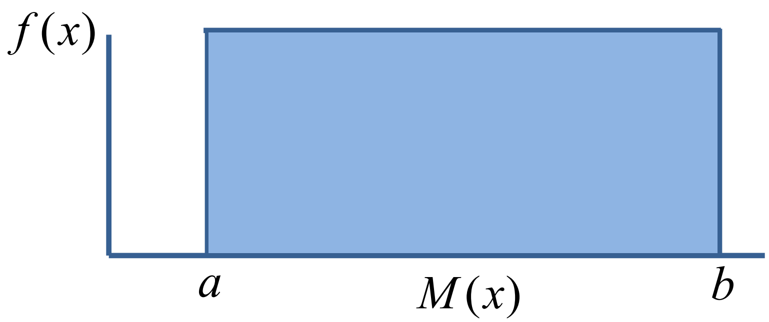
Skewed

- Skewed either left or right



Uniform Distribution

The probabilities in a uniform distribution are the same for all possible outcomes



$$\mu = \frac{a+b}{2}$$

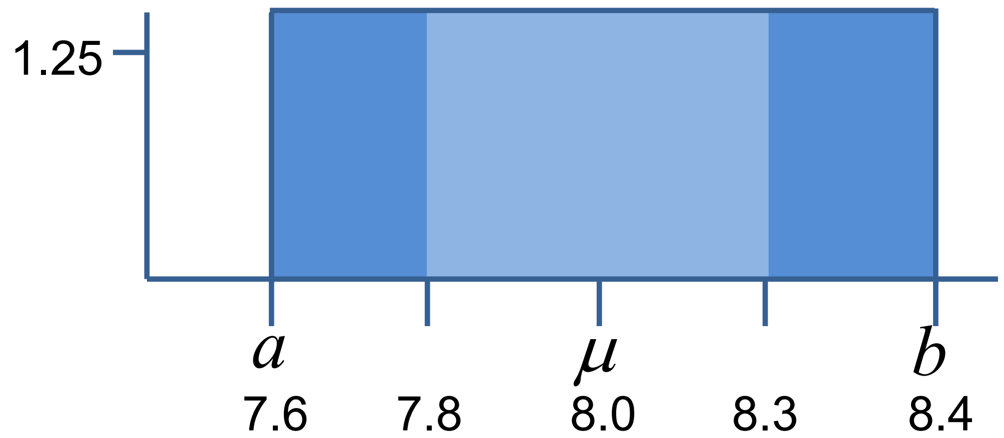
$$Height = \frac{area}{width} = \frac{1}{b-a}$$

Characteristic – symmetrical and area under distribution = 1

Example

A juice box food processing plants manufactures 8 oz boxes of apple juice that range anywhere from 7.6 oz to 8.4 oz and fit a uniform distribution

- Mean = (7.6 + 8.4)/2 = 8
- Height = 1/(8.4 - 7.6) = 1.25
- Assume the company wanted to find the probability a single box weighted between 7.8 and 8.3 oz.
- *Probability* = 1.25 (8.3 - 7.8)= 62.5%



Notes on the Uniform Distribution

The uniform distribution is symmetrical

- Symmetrical about its center μ_x
- μ_x is the median

The uniform distribution is rectangular

- For endpoints c and $d\,(a<b)$ the width of the distribution is $b-a$ and the height is $1/(b-a)$
- The area under the entire uniform distribution is 1
    1. Because $width \times height = (b-a)\times[1/(b-a)]=1$
    2. So $P(a \le x \le b)=1$

Normal Probability Distribution

The normal probability distribution is defined by the equation

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

for all values x on the real number line

- μ is the mean and σ is the standard deviation
- π = 3.14159… and e = 2.71828 is the base of natural logarithms

Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Normal Probability Distribution *continued*

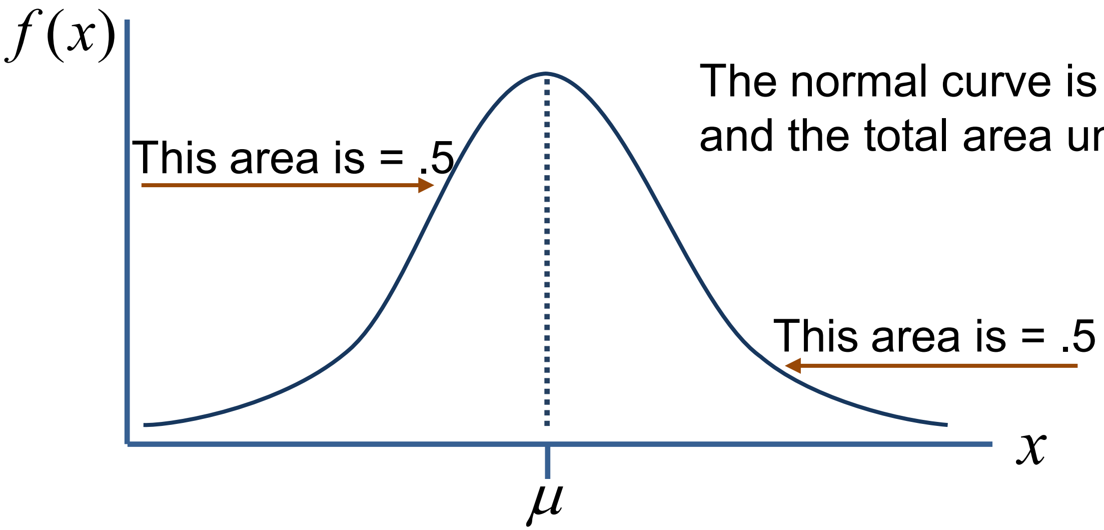
The normal curve is symmetrical about its mean μ

- The mean is in the middle under the curve
- So μ is also the median

It is tallest over its mean μ

The area under the entire normal curve is 1

- The area under either half of the curve is 0.5



The normal curve is symmetrical around μ, and the total area under the curve equals 1.

Properties of the Normal Distribution

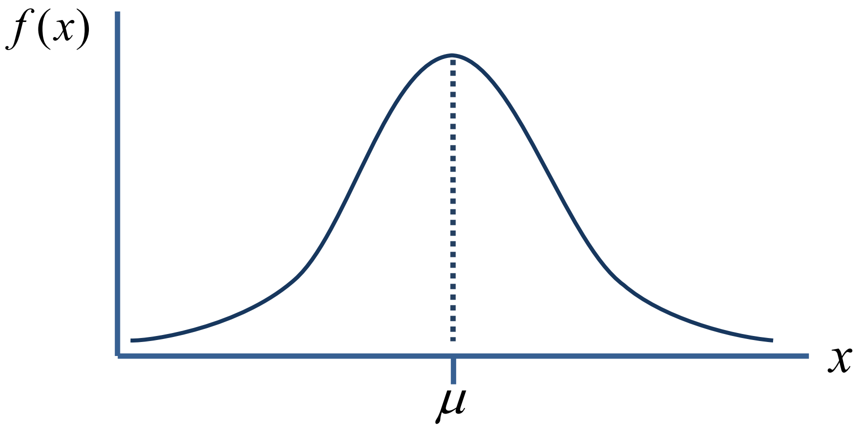
There are an infinite number of normal curves

- The shape of any individual normal curve depends on its specific mean μ and standard deviation σ

Mean = median = mode

- All measures of central tendency equal each other
- The only probability distribution for which this is true

The highest point is over the mean



Properties of the Normal Distribution *continued*

The curve is symmetrical about its mean

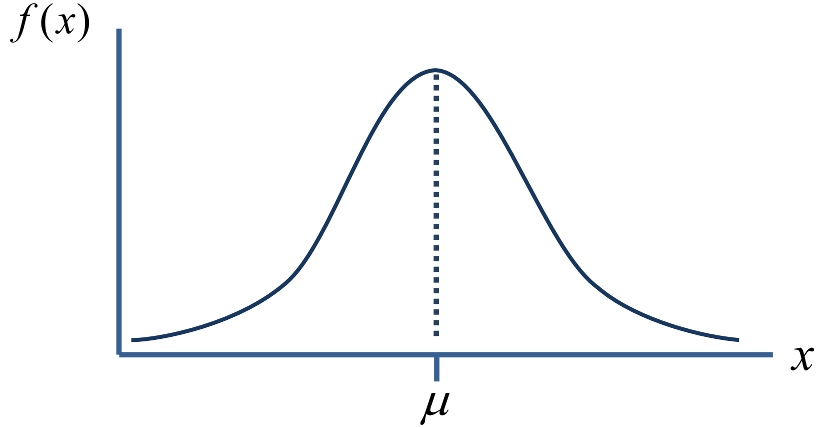
- The left and right halves of the curve are mirror images

The tails of the normal extend to infinity in both directions

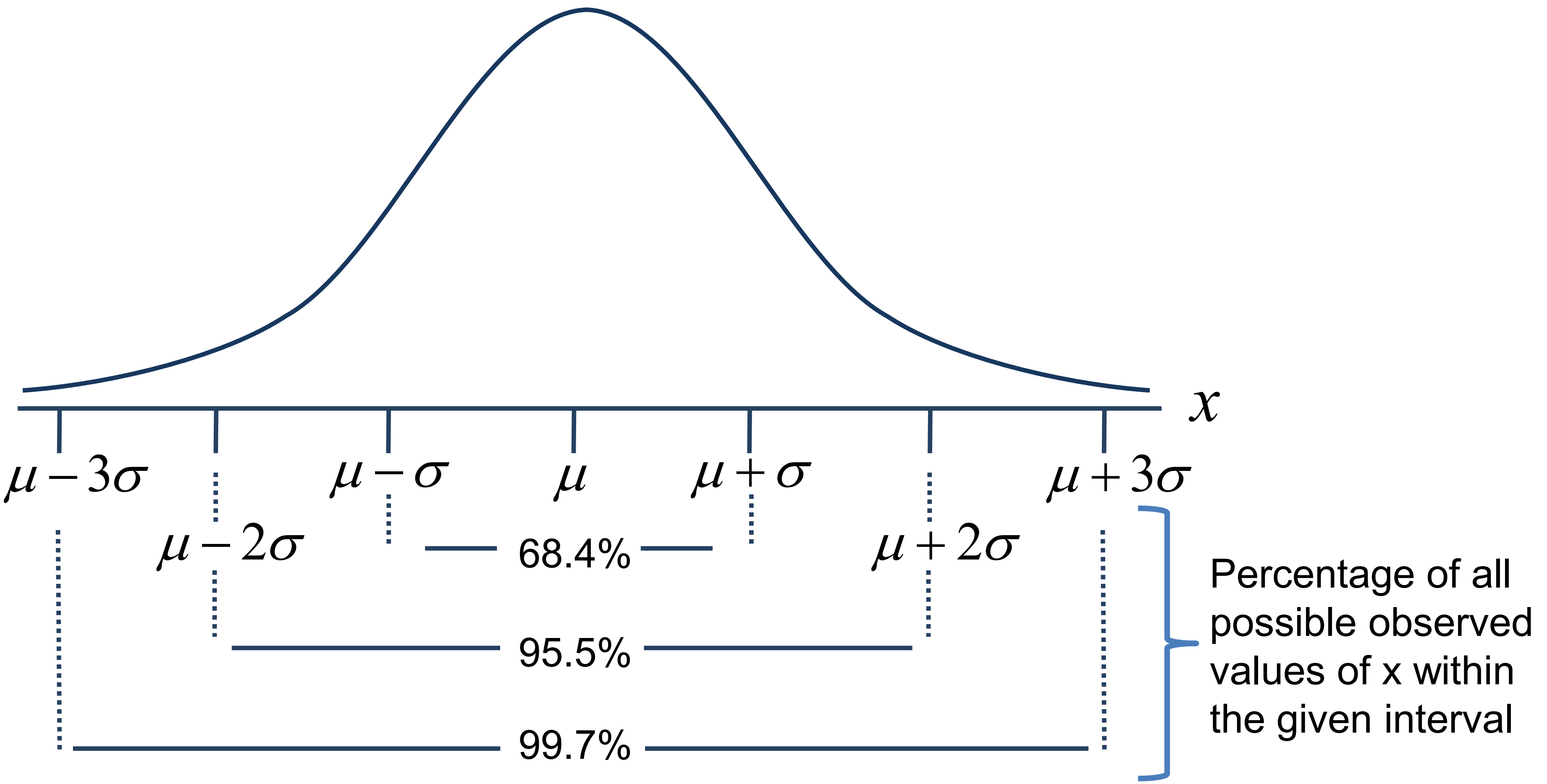
- The tails get closer to the horizontal axis but never touch it

The area under the normal curve to the right of the mean equals the area under the normal to the left of the mean

- The area under each half is 0.5

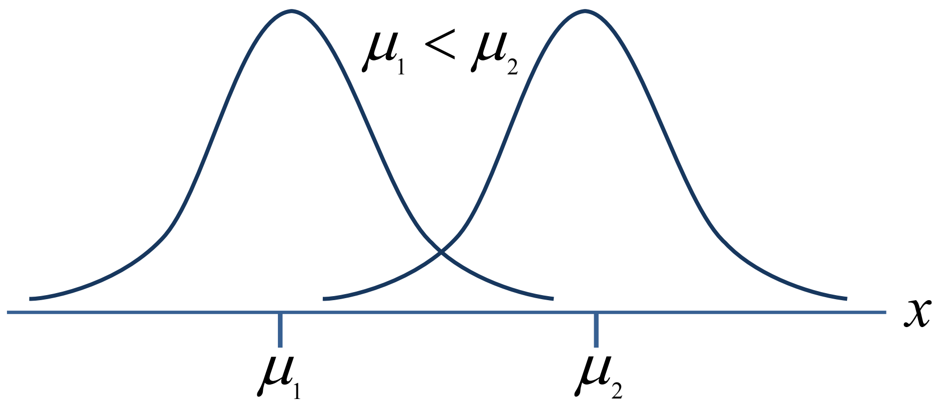


Three Important Percentages

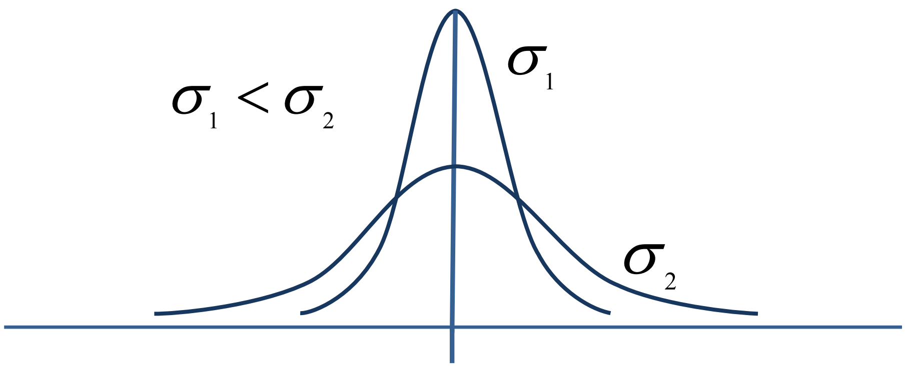


Position and Shape of the Normal Curve

Two data sets with same standard deviation σ, but different means (μ)



Same average (μ), but different standard deviation σ



All normal distributions have 50% of area each side of μ

Adapted from Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Standard Normal Distribution

If x is normally distributed with mean μ and standard deviation σ, then the random variable z

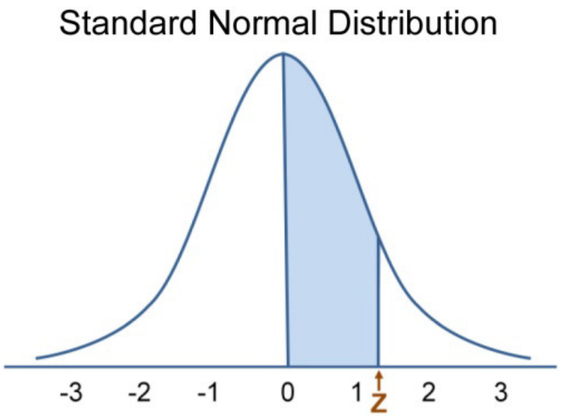
$$z = \frac{x - \mu}{\sigma}$$

is normally distributed with mean 0 and standard deviation 1

- This normal is called the **standard normal distribution**
- z is the number of standard deviation from the mean

Standard Normal Table

The standard normal table is a table that lists the area under the standard normal curve to the right of the mean ($z = 0$) up to the z value of interest



This is the "bell-shaped" curve of the Standard Normal Distribution.

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517

Adapted from Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

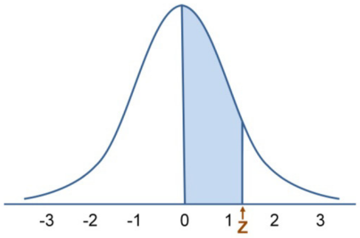
Standard Normal Table *continued*

The values of z (accurate to the nearest tenth) in the table range from 0.00 to 3.09 in increments of 0.01

- z accurate to tenths are listed in the far left column
- The hundredths digit of z is listed across the top of the table

The areas under the normal curve to the right of the mean up to any value of z are given in the body of the table

Standard Normal Distribution



This is the "bell-shaped" curve of the Standard Normal Distribution.

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517

Standard Normal Table *example*

Find $P(0 \leq z \leq 1)$

- Find the area listed in the table corresponding to a z value of 1.00
- Starting from the top of the far left column, go down to "1.0"
- Read across the row $z = 1.0$ until under the column headed by "$.00$"
- The area is in the cell that is the intersection of this row with this column
- As listed in the table, the area is 0.3413, so

$$P(0 \leq z \leq 1) = 0.3413$$

Finding Normal Probabilities

1. Formulate the problem in terms of x values
2. Calculate the corresponding z values, and restate the problem in terms of these z values
3. Find the required areas under the standard normal curve by using the table

Note: It is always useful to draw a picture showing the required areas before using the normal table

Bowerman, B. L., O'Connell, R. T., & Murphree, E. S., (2010). Business Statistics in Practice (6th Ed.), Copyright © McGraw-Hill Education

Example 1

Baltimore Inner Harbor in the month of July averages 4500 tourists per day with a population standard deviation $\sigma = 600$

What is the probability of the number of tourists on any day in July being:

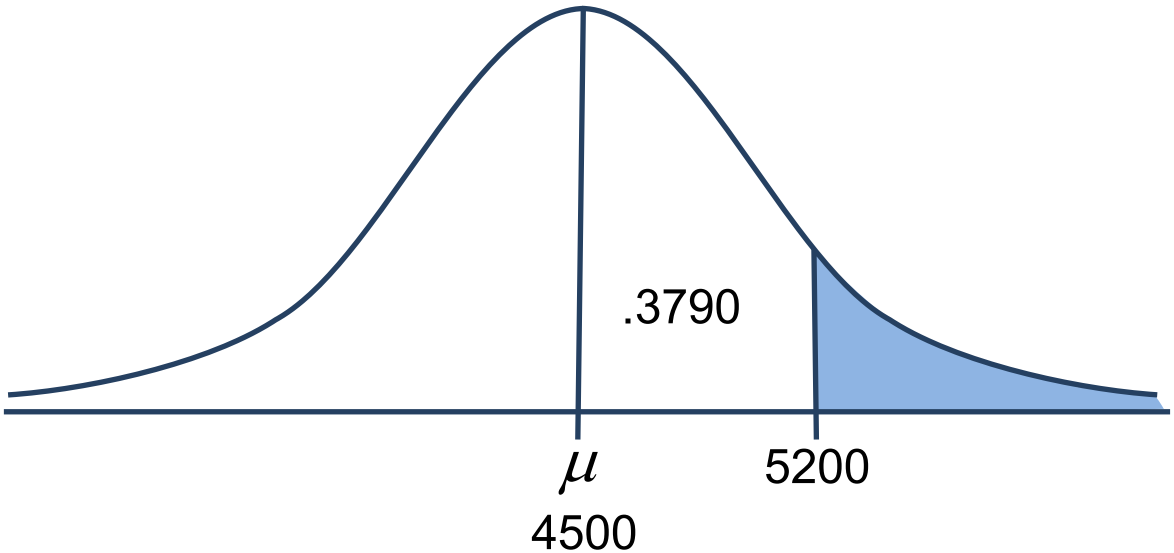
a) Greater than 5200
b) Less than 4000
c) Between 4200 and 5400
d) Between 5400 and 5600

Example 1 *continued*

a) Greater than 5200

$\mu = 4500$

$\sigma = 600$



$$z = \frac{x-\mu}{\sigma} = \frac{5200-4500}{600} = 1.17$$

From Table z = .3790

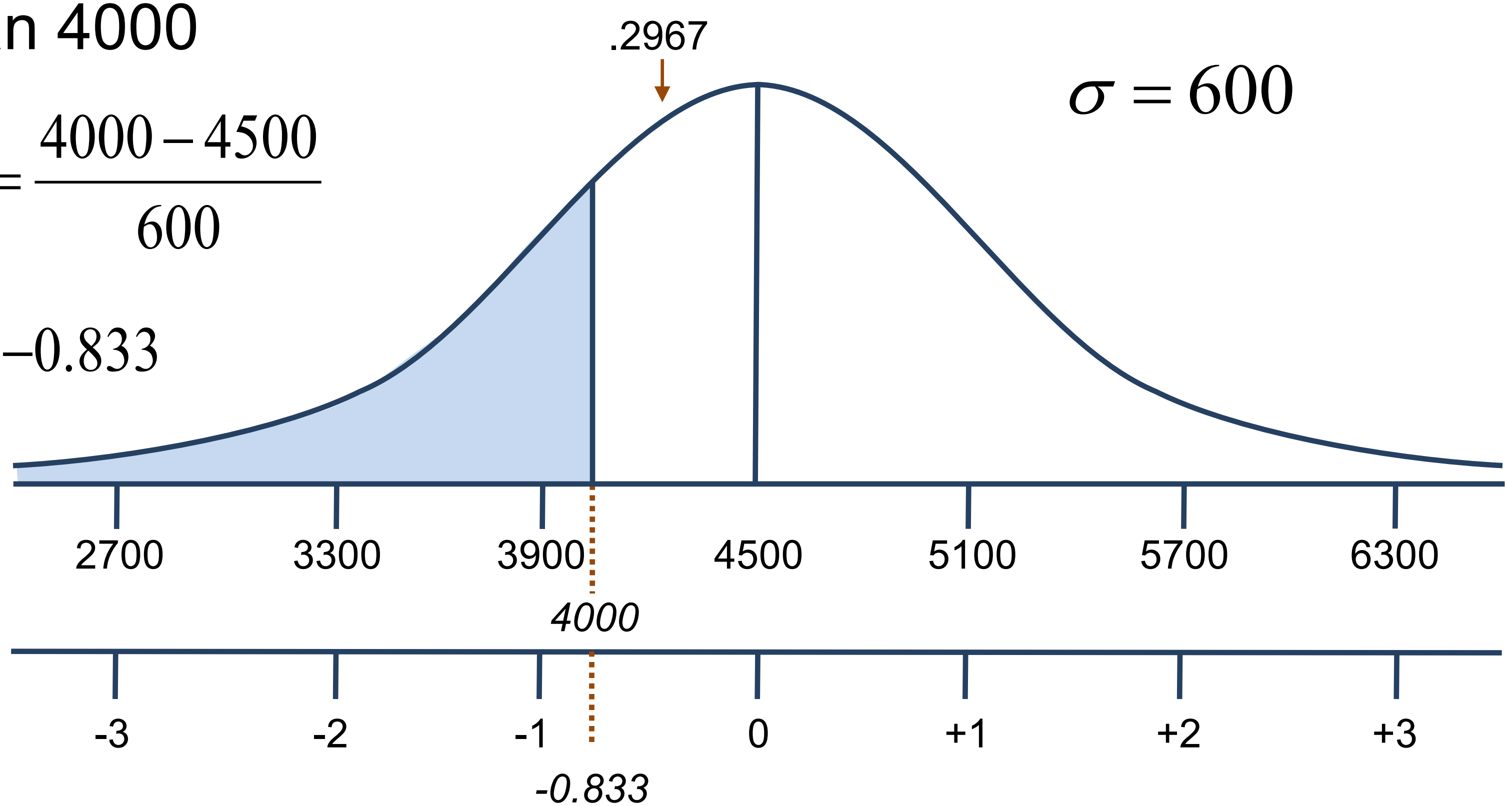
Therefore the probability of greater than 5200 tourists is .5000 - .3790 = .1210 or 12.10%.

Example 1 *continued*

b) Less than 4000

$$z = \frac{x-\mu}{\sigma} = \frac{4000-4500}{600}$$

$$= \frac{-500}{600} = -0.833$$



The negative indicates to left of mean

From table probability (area under curve)= .2967

Answer .5000 - .2967 = .2033 or 20.33%

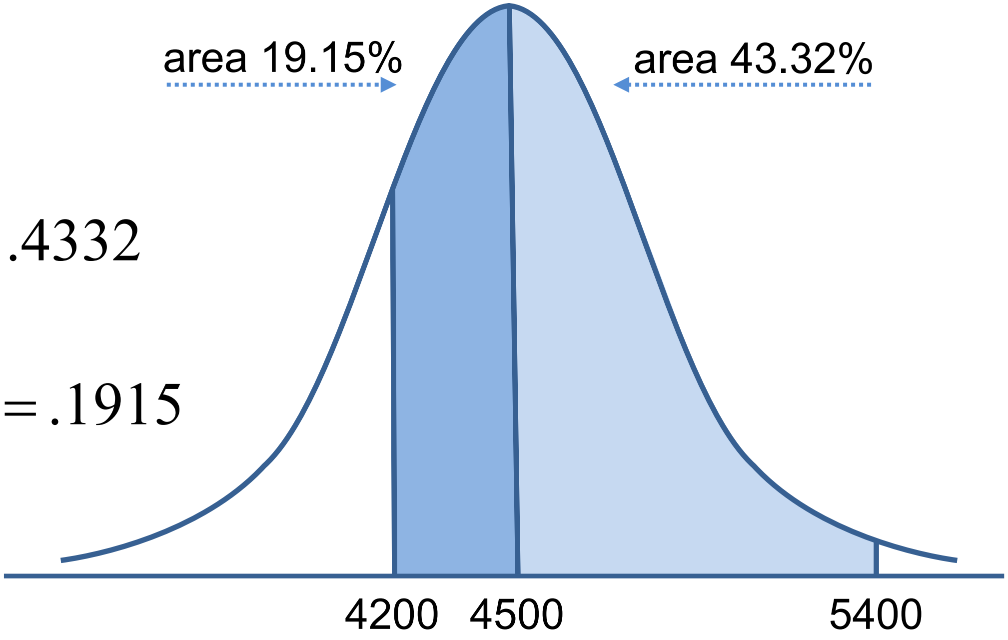
Example 1 *continued*

c) Between 4200 and 5400

Only work on one side of curve at a time

$$z = \frac{5400 - 4500}{600} = 1.5, \ \ area = .4332$$

$$z = \frac{4200 - 4500}{600} = -0.5, \ \ area = .1915$$



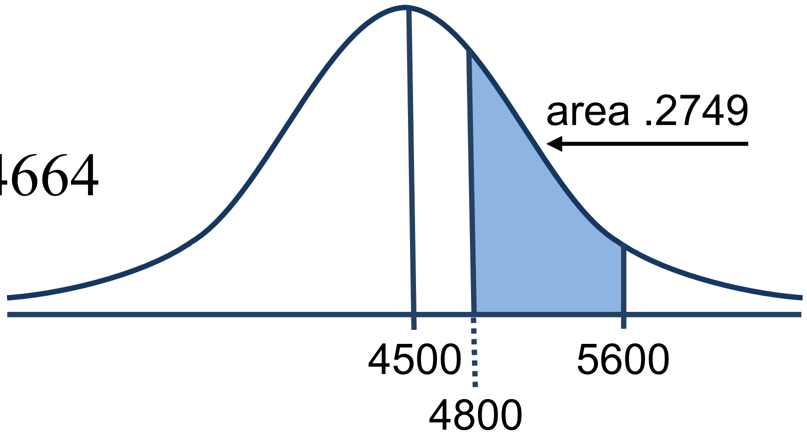
Add the two areas = .6247 or 62.47% probability the number of tourists will be between 4200 and 5400

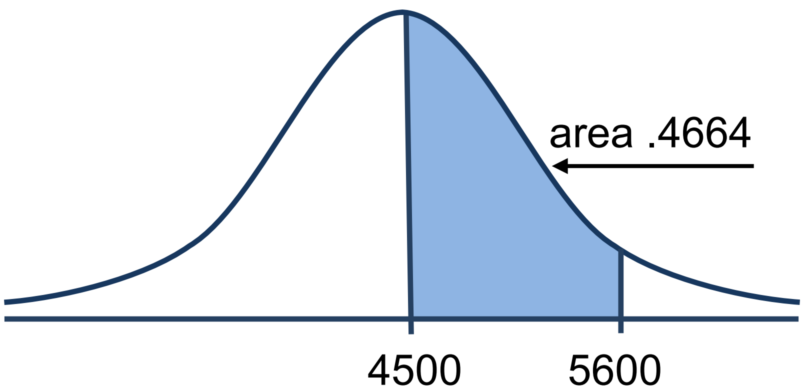
Example 1 *continued*

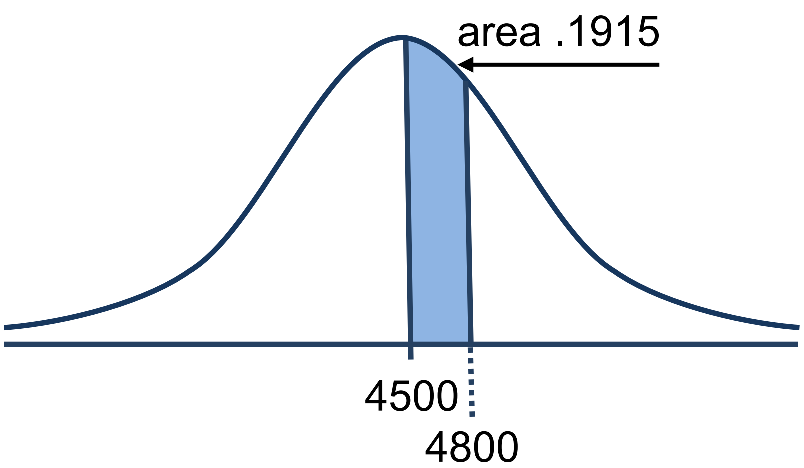
d) Between 4800 and 5600

$$z = \frac{x-\mu}{\sigma} = \frac{5600-4500}{600} = 1.83,\ \ area = .4664$$

$$z = \frac{4800-4500}{600} = 0.5,\ \ area = .1915$$







Subtract the difference to calculate shaded area

.4664 - .1915 = .2749 or 27.49% probability between 4800 and 5600

Finding x from a Known Probability

If we know a probability we can calculate an x-value by working the formula in reverse.

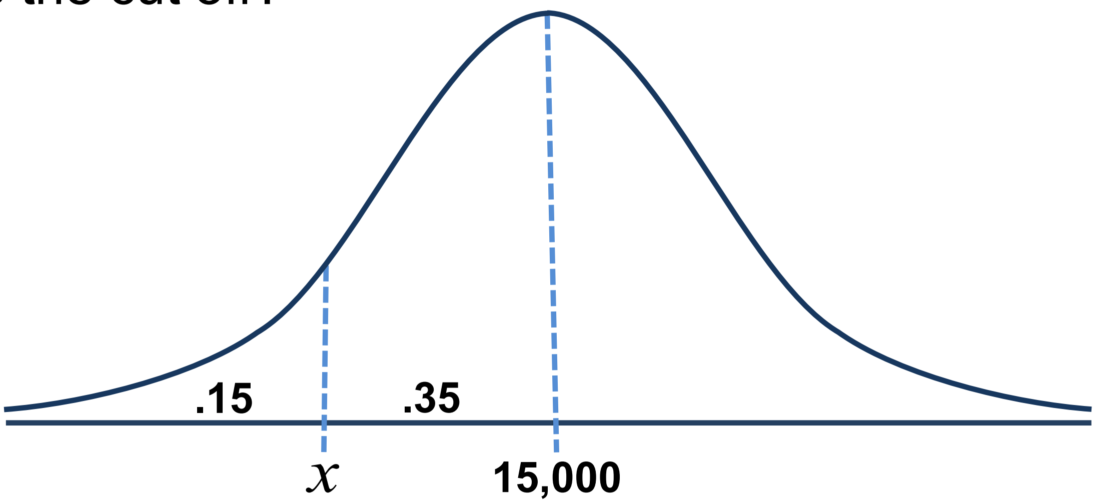
Example: The most recent census data has become available and a county must now recalculate the income of under privileged families to receive a government subsidy. The average income for the county is \$15,000 with a standard deviation of \$2,000. If the county wants the lowest 15% to receive the funds, what income level is the cut off?

$$z = \frac{x - \mu}{\sigma}$$

$$-1.04 = \frac{x - 15{,}000}{2{,}000}$$

$$x = (-1.04)(2{,}000) + 15{,}000$$

$$x = \$12{,}920$$



Interpretation: Therefore, for a family to qualify, the annual income must be below \$12,920.

Exponential Distribution

Suppose that some event occurs as a Poisson process

- That is, the number of times an event occurs is a Poisson random variable

Let x be the random variable of the interval between successive occurrences of the event

- The interval can be some unit of time or space

Then x is described by the **exponential distribution**

- With parameter λ, which is the mean number of events that can occur per given interval

Exponential Distribution *continued*

If λ is the mean number of events per given interval, then the equation of the exponential distribution is

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{otherwise}\end{cases}$$

The probability that x is any value between given values a and $b(a<b)$ is

$$P(a \leq x \leq b) = e^{-\lambda a} - e^{-\lambda b}$$

$$and$$

$$P(x \leq c) = 1 - e^{-\lambda c} \quad \text{and} \quad P(x \geq c) = e^{-\lambda c}$$

Exponential Distribution

The mean μ_x and standard deviation σ_x of an exponential random variable x are

$$\mu_x = 1/\lambda \text{ and } \sigma_x = 1/\lambda$$

$$f(x) = \lambda e^{-\lambda x}$$