Business Analytics

UNIVERSITETI I EJL
ЈИЕ УНИВЕРЗИТЕТ
SEE UNIVERSITY

Cotinuous Random Variables

Faton Berisha

Chapter 5

Continuous Random Variables

Continuous Random Variables

Continuous Probability Distributions

Recall: 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

A curve $f(x)$ is a *continuous probability distribution* of the continuous 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*, or
  - *probability density function*

Properties of Continuous
Probability Distributions

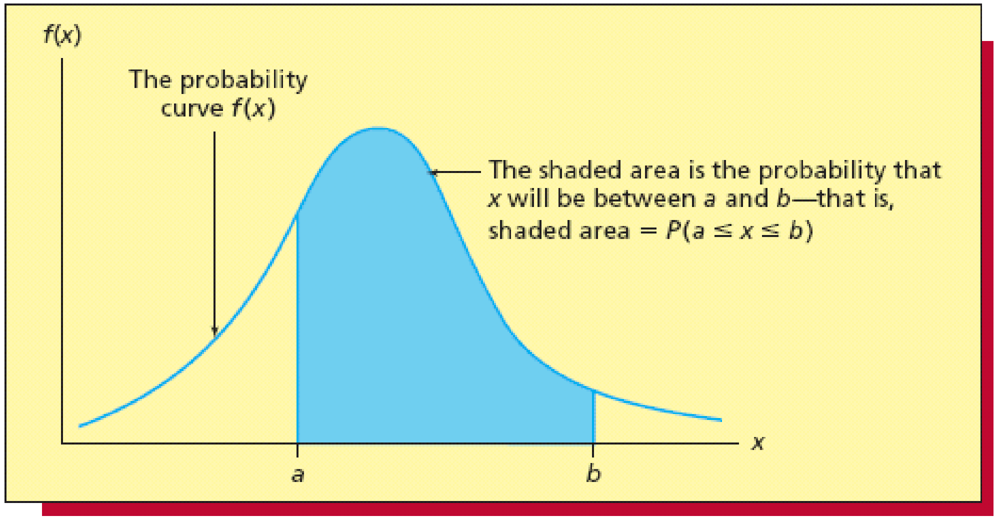
Properties of $f(x)$: $f(x)$ is a continuous function such that

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

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

Area and Probability

- The blue-colored area under the probability curve $f(x)$ from the value $x = a$ 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



Distribution Shapes

- Symmetrical and rectangular
  - The uniform distribution
    - Section 5.2
- Symmetrical and bell-shaped
  - The normal distribution
    - Section 5.3
- Skewed
  - Skewed either left or right
    - Section 5.5 for the right-skewed exponential distribution

Continuous probability

- ***Probability***: $$P(a \le x \le b) = \int_a^b f(x)\,dx$$
- ***Mean***: $$\mu_X = \int_a^b x f(x)\,dx$$
- ***Variance***: $$\sigma_X^2 = \int_a^b (x - \mu_X)^2 f(x)\,dx$$
- ***Standard deviation***: $$\sigma_X = \sqrt{\sigma_X^2}$$

The Uniform Distribution

If c and d are numbers on the real line ($c < d$), the probability curve describing the *uniform distribution* is

$$f(x) = \begin{cases} \dfrac{1}{d-c} & \text{for } c \le x \le d \\ 0 & \text{otherwise} \end{cases}$$

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

$$P(a \le x \le b) = \frac{b-a}{d-c}$$

Note: The number ordering is $c < a < b < d$

The Uniform Distribution (Cont.)

The mean μ_X and standard deviation σ_X of a uniform random variable x are

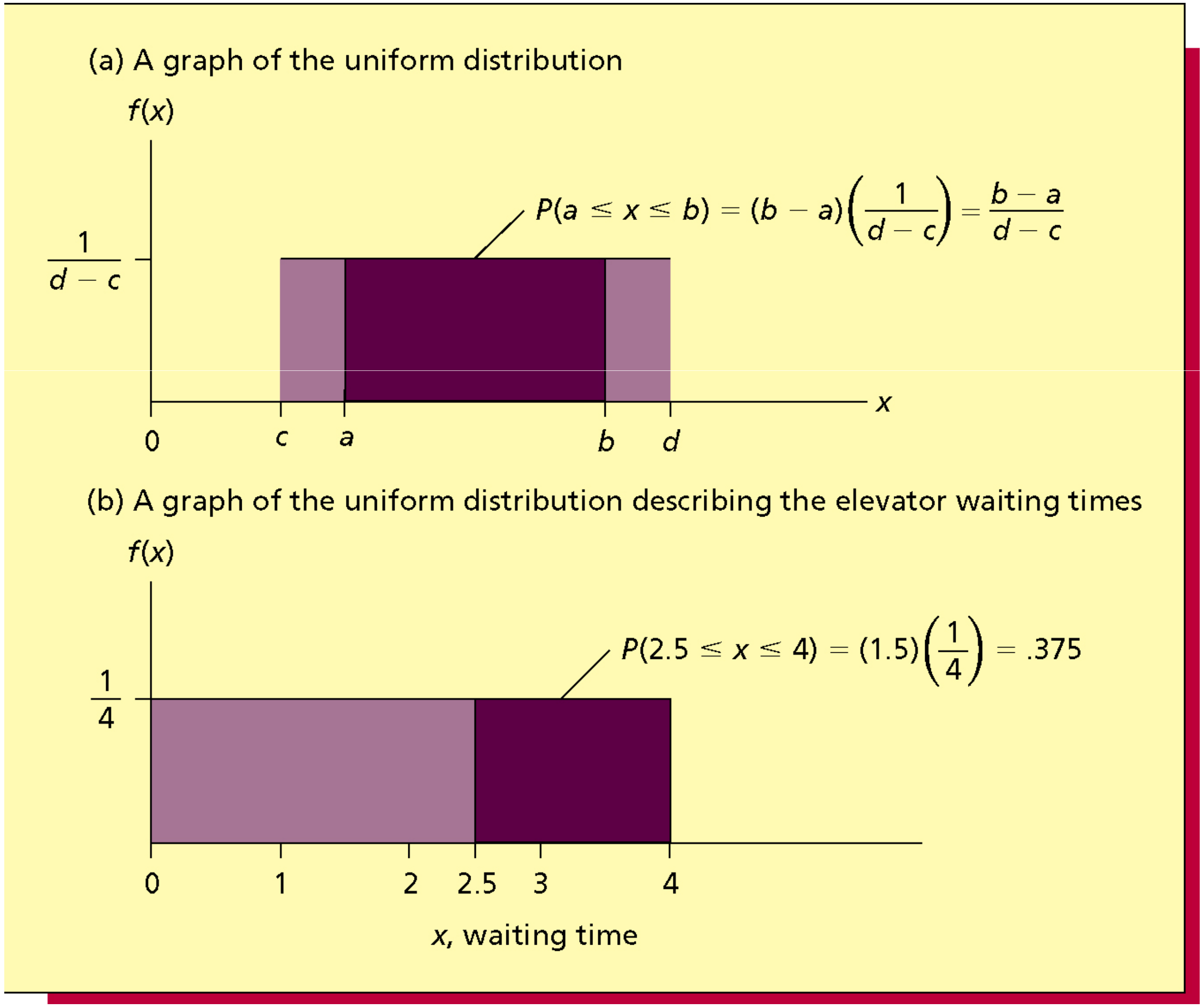
$$\mu_X = \frac{c+d}{2}$$

$$\sigma_X = \frac{d-c}{\sqrt{12}}$$

- These are the parameters of the uniform distribution with endpoints c and d ($c < d$)

The Uniform Probability Curve

Example 5.1. Elevator waiting time



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* ($c < d$) the width of the distribution is $d - c$ and the height is $1/(d - c)$
  - The area under the entire uniform distribution is 1
    - Because $(d - c) \times [1/(d - c)] = 1$
    - So $P(c \leq x \leq d) = 1$

The Normal Probability Distribution

The *normal probability distribution* is defined by the equation

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

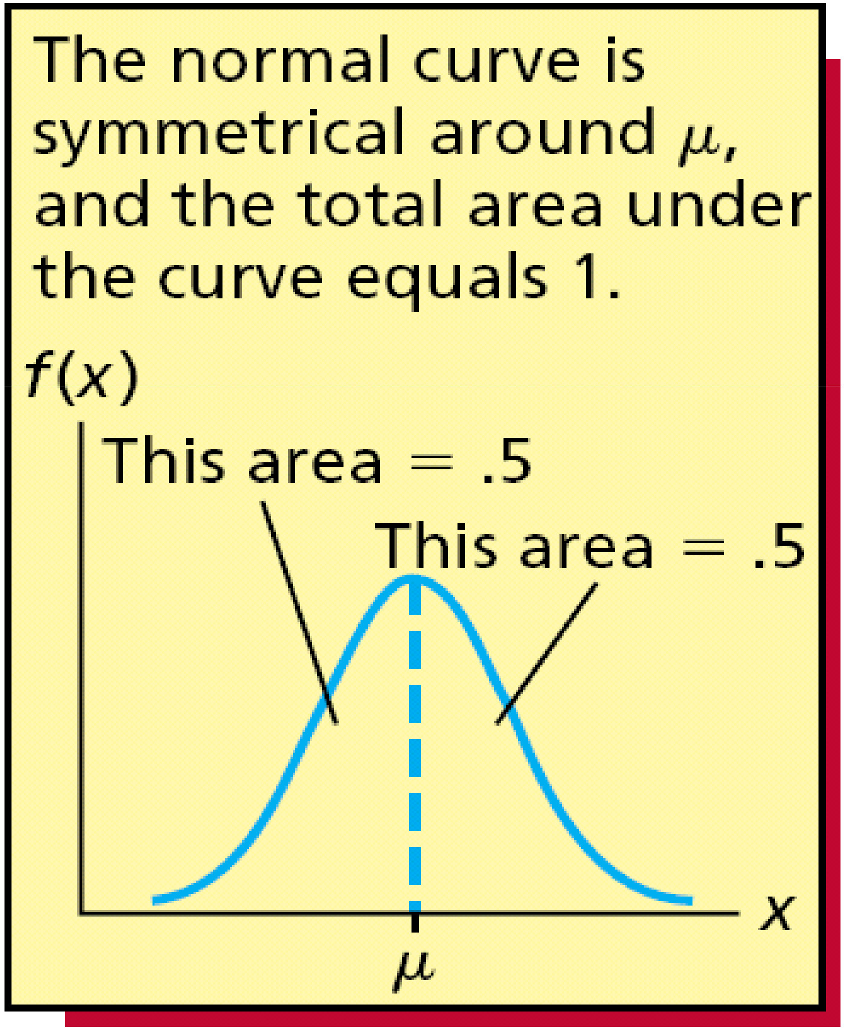
for all values x on the real number line, where

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

The Normal Probability Distribution (Cont.)

The *normal curve* is symmetrical and bell-shaped

- The normal is symmetrical about its mean μ
  - The mean is in the middle under the curve
  - So μ is also the median
- The normal is tallest over its mean μ
  - So μ is also the mode
- The area under the entire normal curve is 1
  - The area under either half of the curve is 0.5



Properties of the Normal Distribution

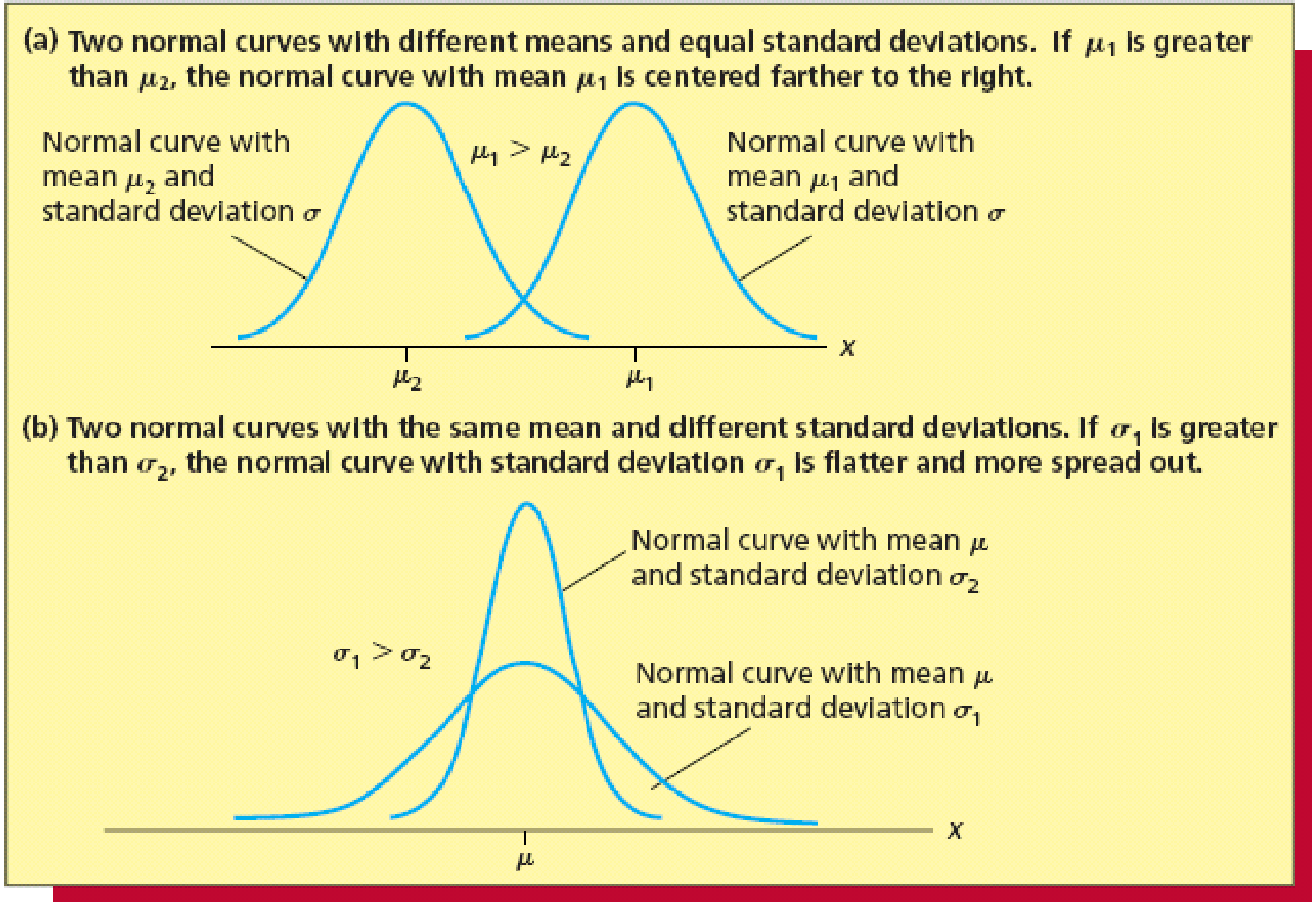
- There is an infinite number of possible normal curves
  - The particular shape of any individual normal depends on its specific mean μ and standard deviation σ
- The highest point of the curve is located over the mean
- Mean = Median = Mode
  - All the measures of central tendency equal each other
    - This is the only probability distribution for which this is true

Properties of the Normal Distribution (Cont.)

- The curve is symmetrical about its mean
  - The left and right halves of the curve are mirror images of each other
- The tails of the normal curve 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

The Position and Shape

of the Normal Curve



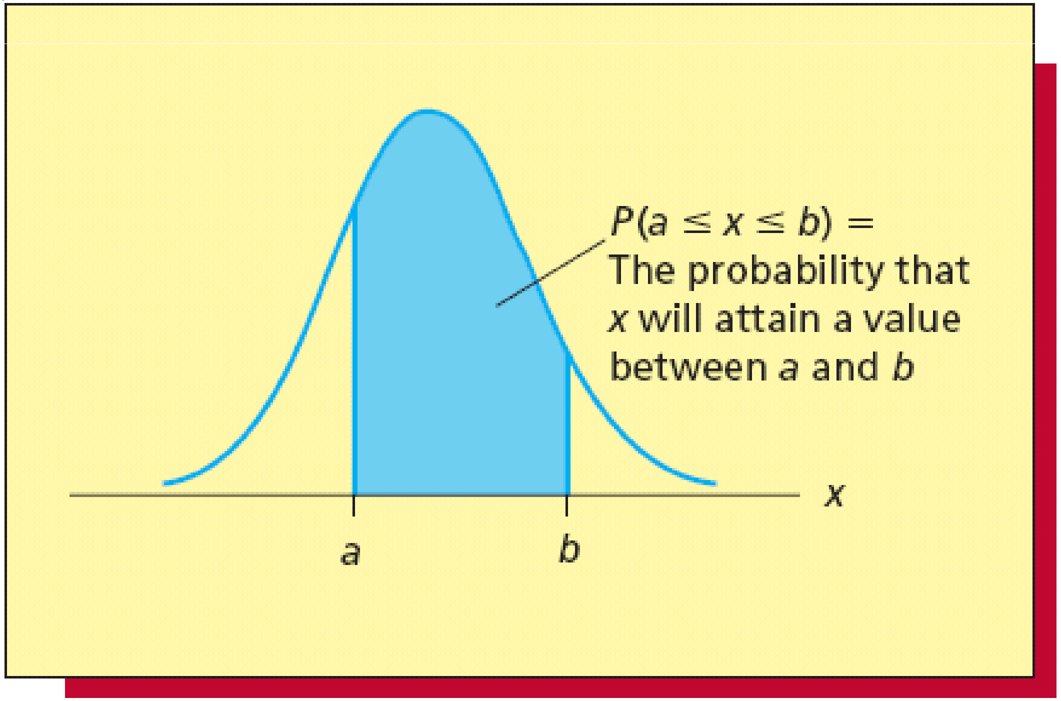
(a) The mean μ positions the peak of the normal curve over the real axis

(b) The variance σ^2 measures the width or spread of the normal curve

Normal Probabilities

Suppose x is a normally distributed random variable with mean μ and standard deviation σ

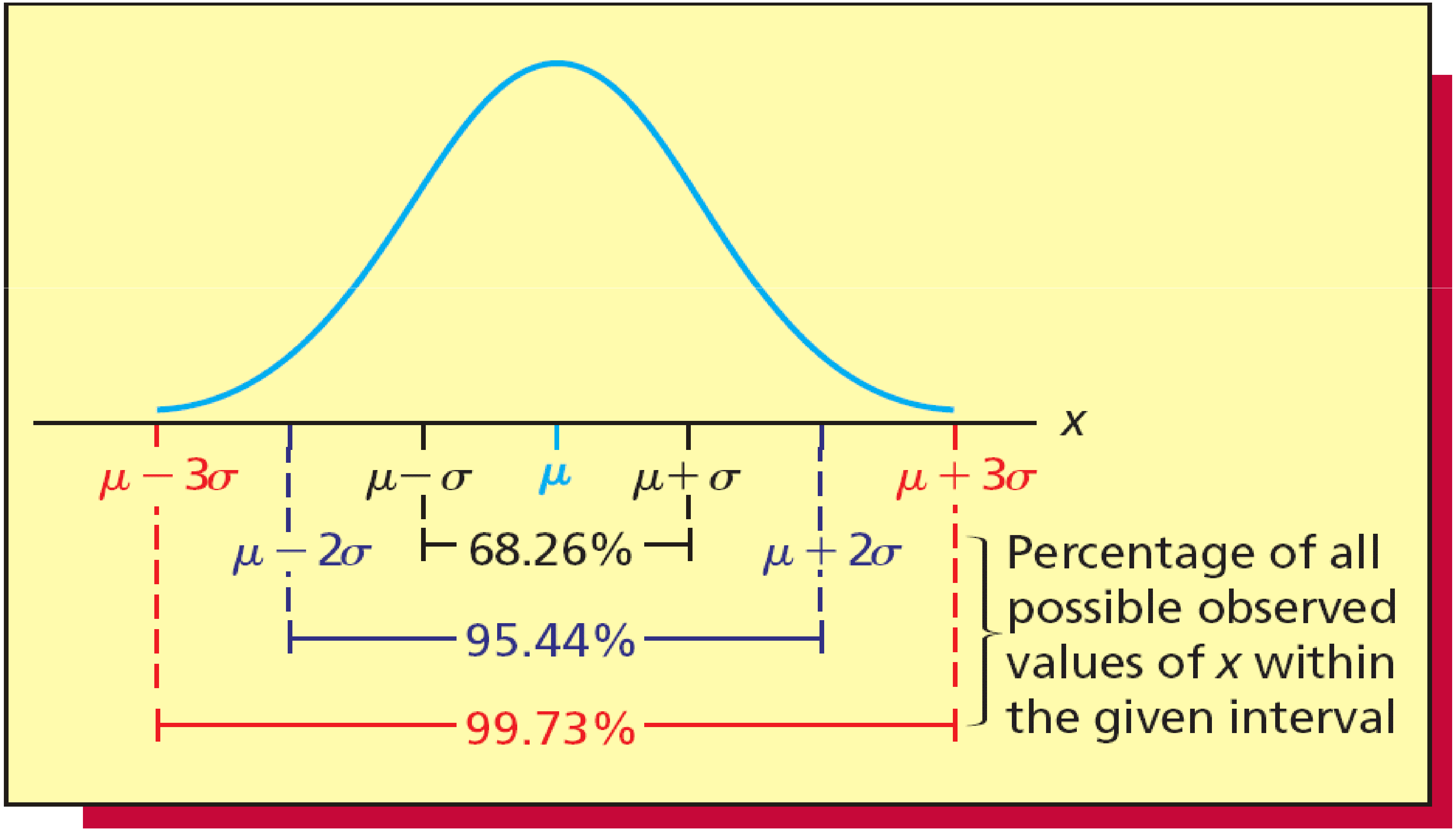
The probability that x could take any value in the range between two given values a and b ($a < b$) is $P(a \leq x \leq b)$



$P(a \leq x \leq b)$ is the area colored in **blue** under the normal curve and between the values $x = a$ and $x = b$

Three Important Percentages

Empirical Rule of a normally distributed population



The Standard Normal Distribution

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

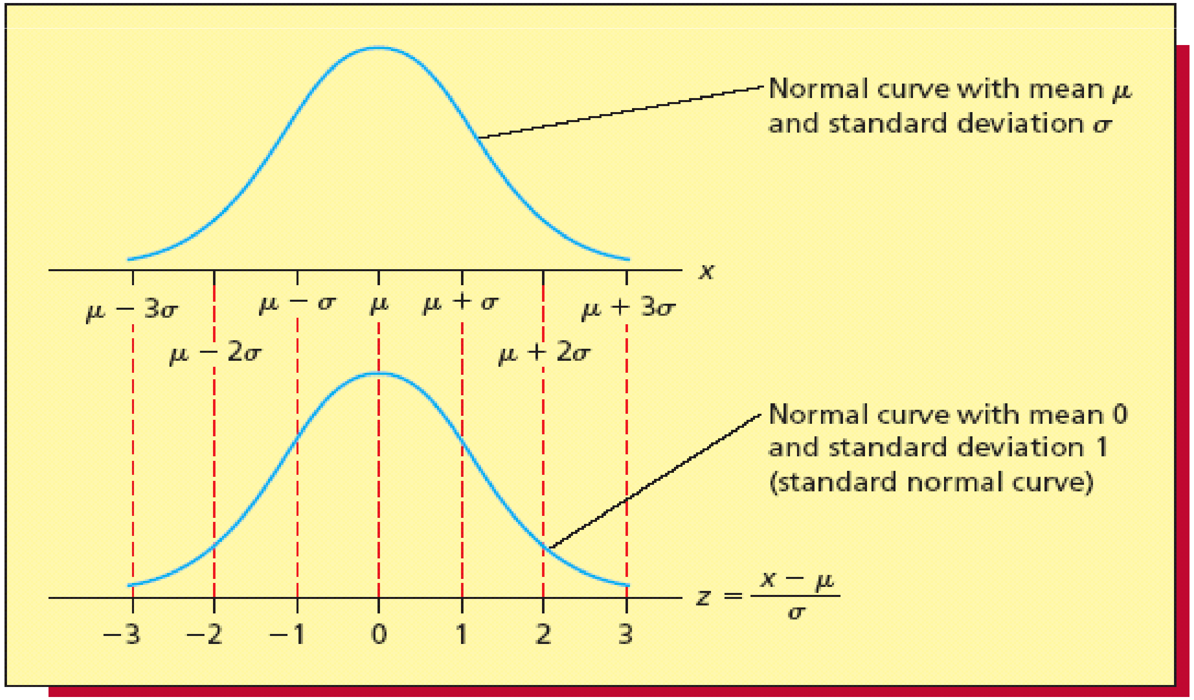
$$z = \frac{x - \mu}{\sigma}$$

is normally distributed with mean 0 and standard deviation 1; this normal is called the *standard normal distribution*

The Standard Normal Distribution
(Cont.)

z measures the number of standard deviations that x is from the mean μ

- The algebraic sign on z indicates on which side of μ is x
- z is positive if $x > \mu$ (x is to the right of μ on the number line)
- z is negative if $x < \mu$ (x is to the left of μ on the number line)



The 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
  - See Table 5.1
  - Also see Table A.3 in Appendix A and the table on the back of the front cover
    - This table is so important that it is repeated 3 times in the textbook!
    - Always look at the accompanying figure for guidance on how to use the table

The Standard Normal Table (Cont.)

- 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

The Standard Normal Table (Cont.)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

The 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

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

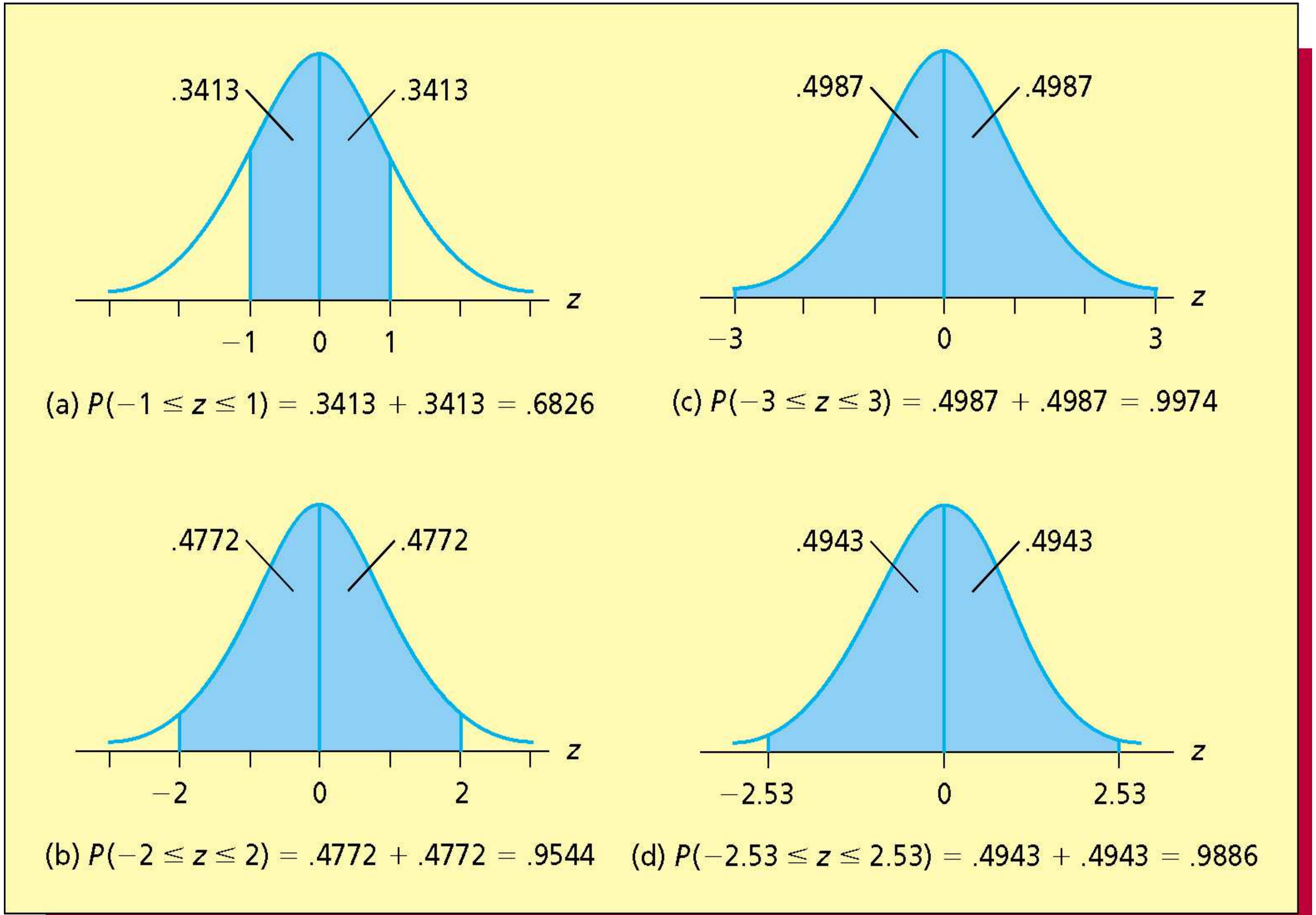
Calculating $P(-2.53 \leq z \leq 2.53)$

- First, find $P(0 \leq z \leq 2.53)$
  - Go to the table of areas under the standard normal curve
  - Go down left-most column for $z = 2.5$
  - Go across the row 2.5 to the column headed by .03
  - The area to the right of the mean up to a value of $z = 2.53$ is the value contained in the cell that is the intersection of the 2.5 row and the .03 column
  - The table value for the area is 0.4943

Calculating $P(-2.53 \le z \le 2.53)$ (Cont.)

- From last slide, $P(0 \le z \le 2.53)=0.4943$
- By symmetry of the normal curve, this is also the area to the **left** of the mean down to a value of $z = -2.53$
  - Then $P(-2.53 \le z \le 2.53) = 0.4943 + 0.4943 = 0.9886$

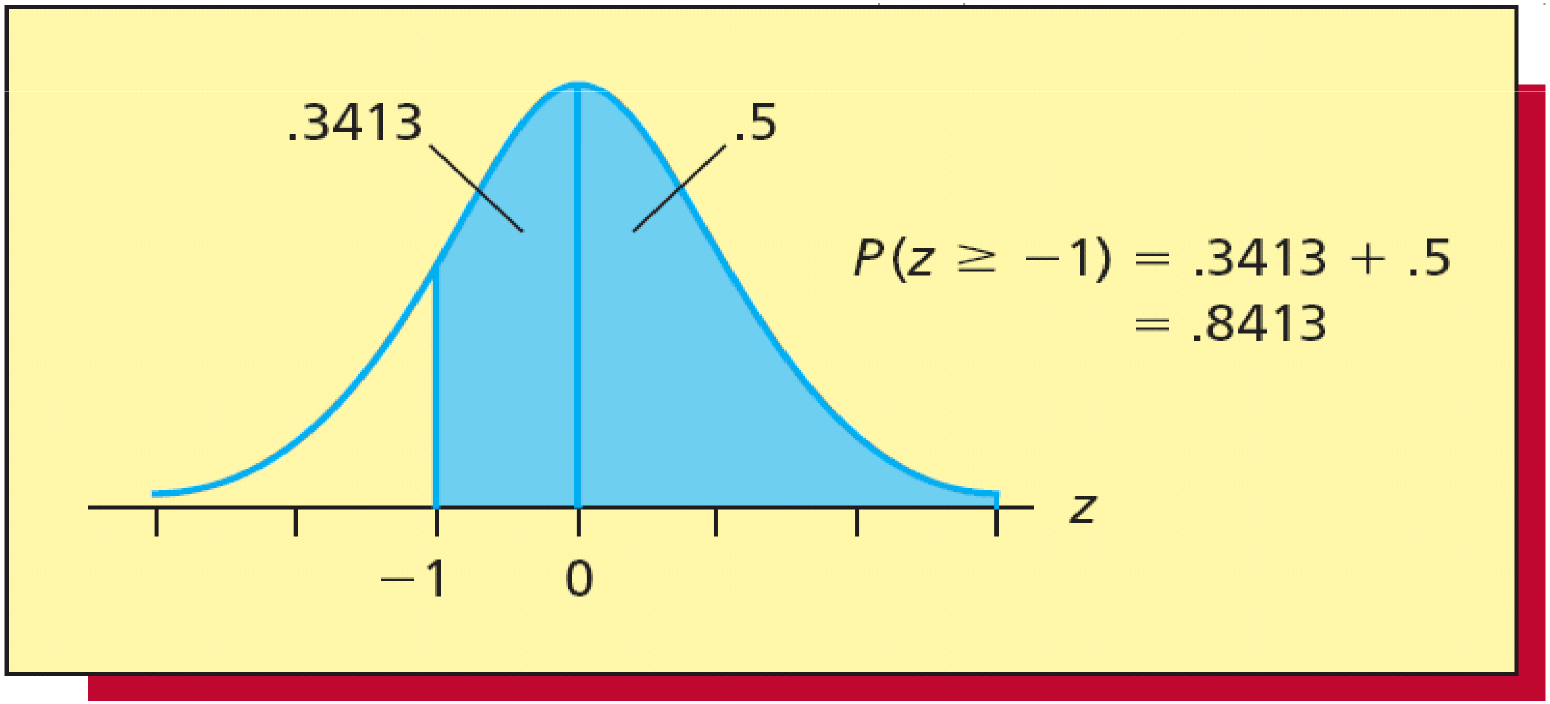
Some Areas Under the Standard Normal Curve



Calculating $P(z \geq -1)$

An example of finding the area under the standard normal curve to the right of a negative z value

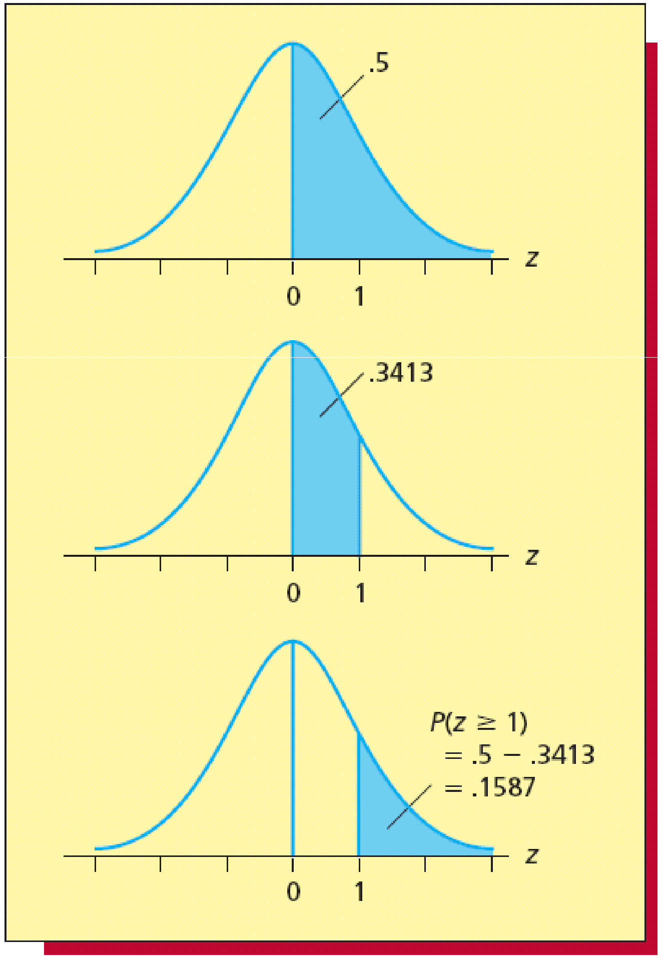
- Shown is finding the under the standard normal for $z \geq -1$



Calculating $P(z \geq 1)$

An example of finding tail areas

- Shown is finding the right-hand tail area for $z \geq 1.00$
  - Equivalent to the left-hand tail area for $z \leq -1.00$



Finding Normal Probabilities

General procedure:

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

Finding Normal Probabilities (Cont.)

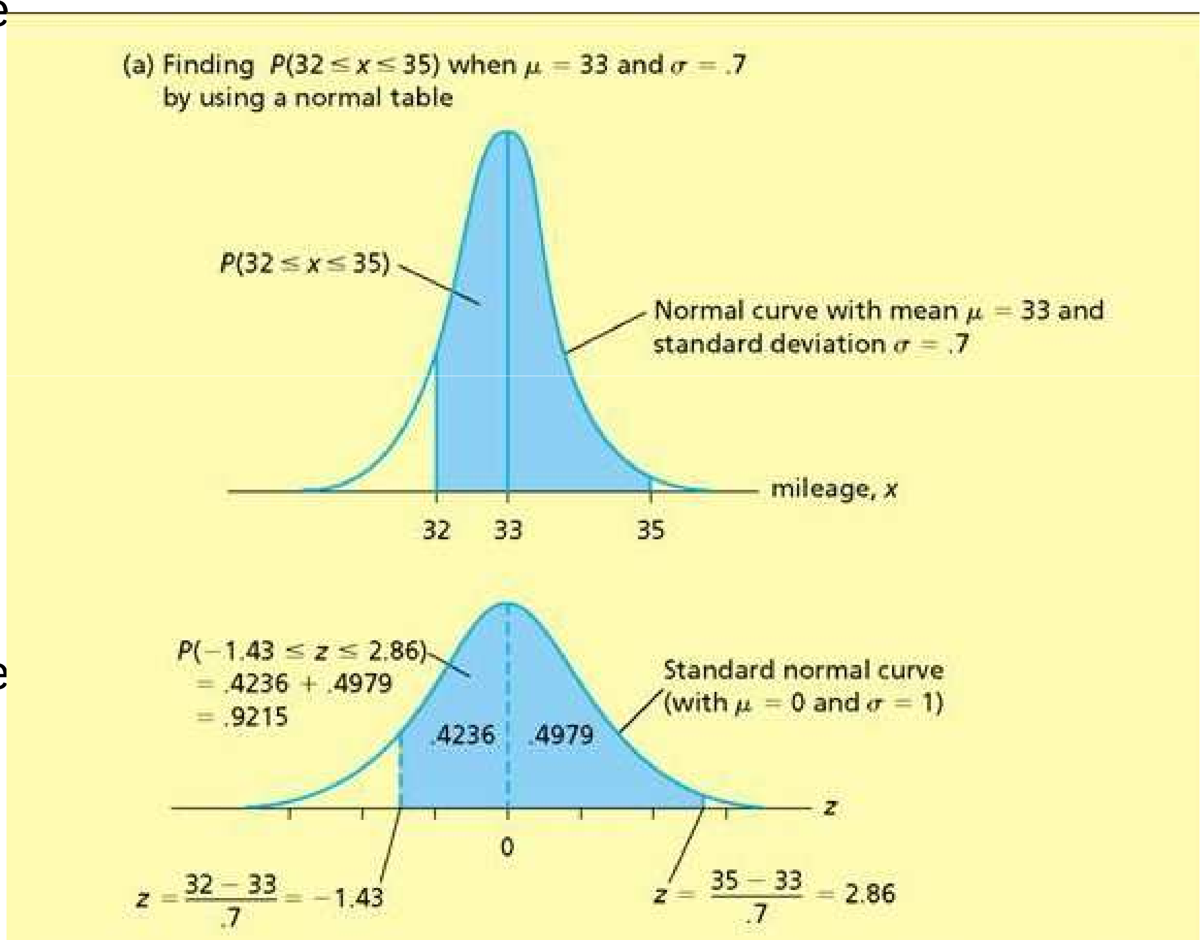
Example 5.2. The car mileage case

Procedure

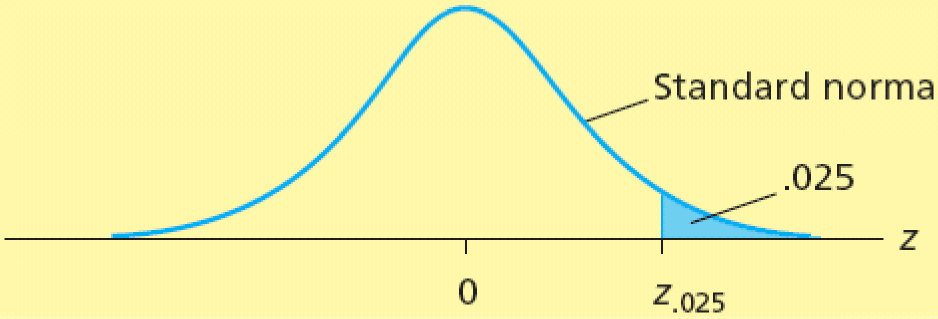
1. Formulate in terms of *x*
2. Reformulate in corresponding terms of *z*.
3. Find the obtained area under the standard normal curve.

Example 5.3. The car mileage case (Evaluating the strength of the evidence)

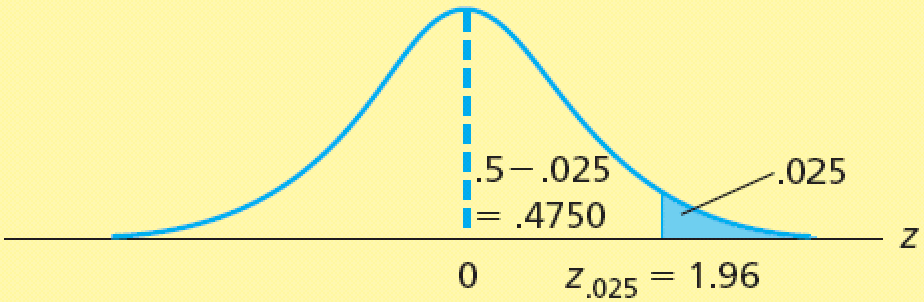
Example 5.4. The coffee temperature case



Finding z-Points on a Standard Normal Curve



(a) $z_{.025}$ is the point on the horizontal axis under the standard normal curve that gives a right-hand tail area equal to .025



(b) Finding $z_{.025}$

Finding *x*-Points on a normal curve

Example 5.5. Demand for video tape

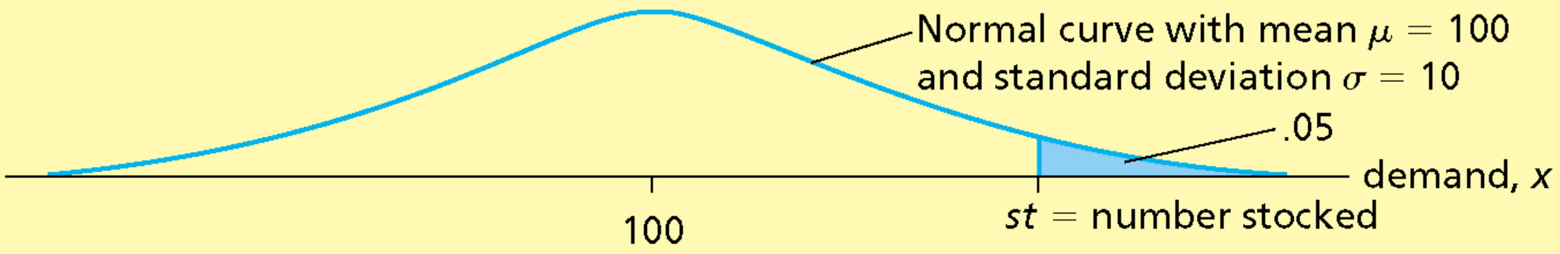
- Weekly demand distributed normally with μ = 100 tapes and σ = 10 tapes.
- Find the number *st* of stocked tapes so that there is onlt a 5% chance that the store will run short of tapes during the week.
- $P(x > st) = 0.05$

$$P(x > st) = P\left(\frac{x-\mu}{\sigma} > \frac{st-\mu}{\sigma}\right) = P\left(z > \frac{st-100}{10}\right) = 0.05$$

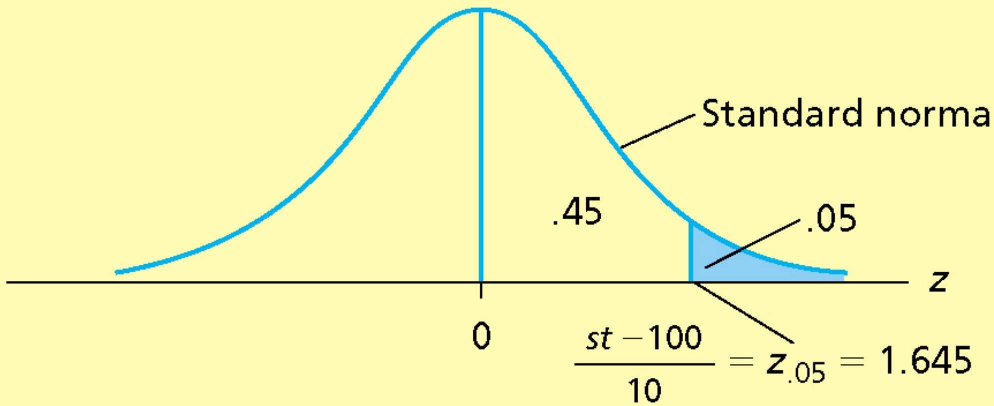
$$\frac{st-100}{10} = z_{0.05} = 1.645$$

$$st = 10 \cdot 1.645 + 100 = 116.45$$

Finding x-points (Cont.)



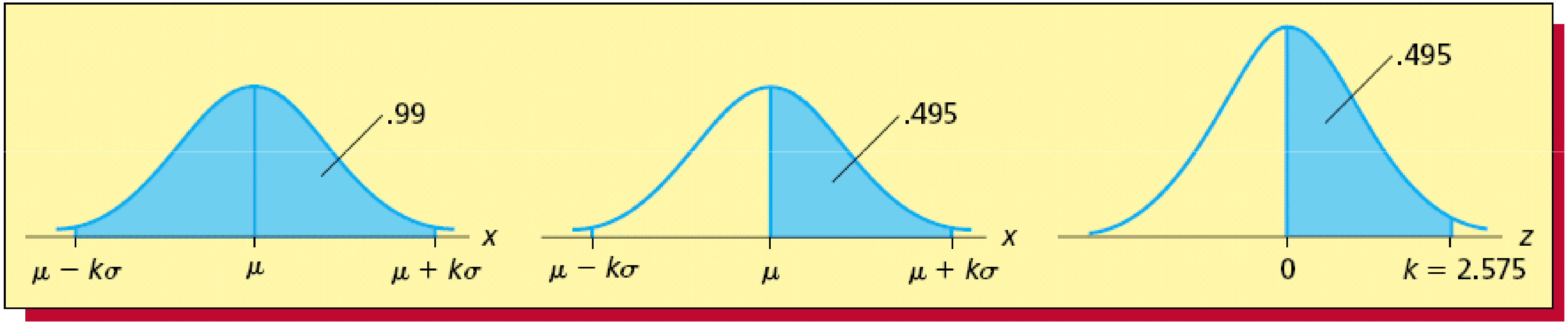
(a) The number of tapes stocked, *st*, must be chosen so that there is a .05 probability that the demand, *x*, will exceed *st*



(b) Finding $z_{.05}$, the z value corresponding to *st*

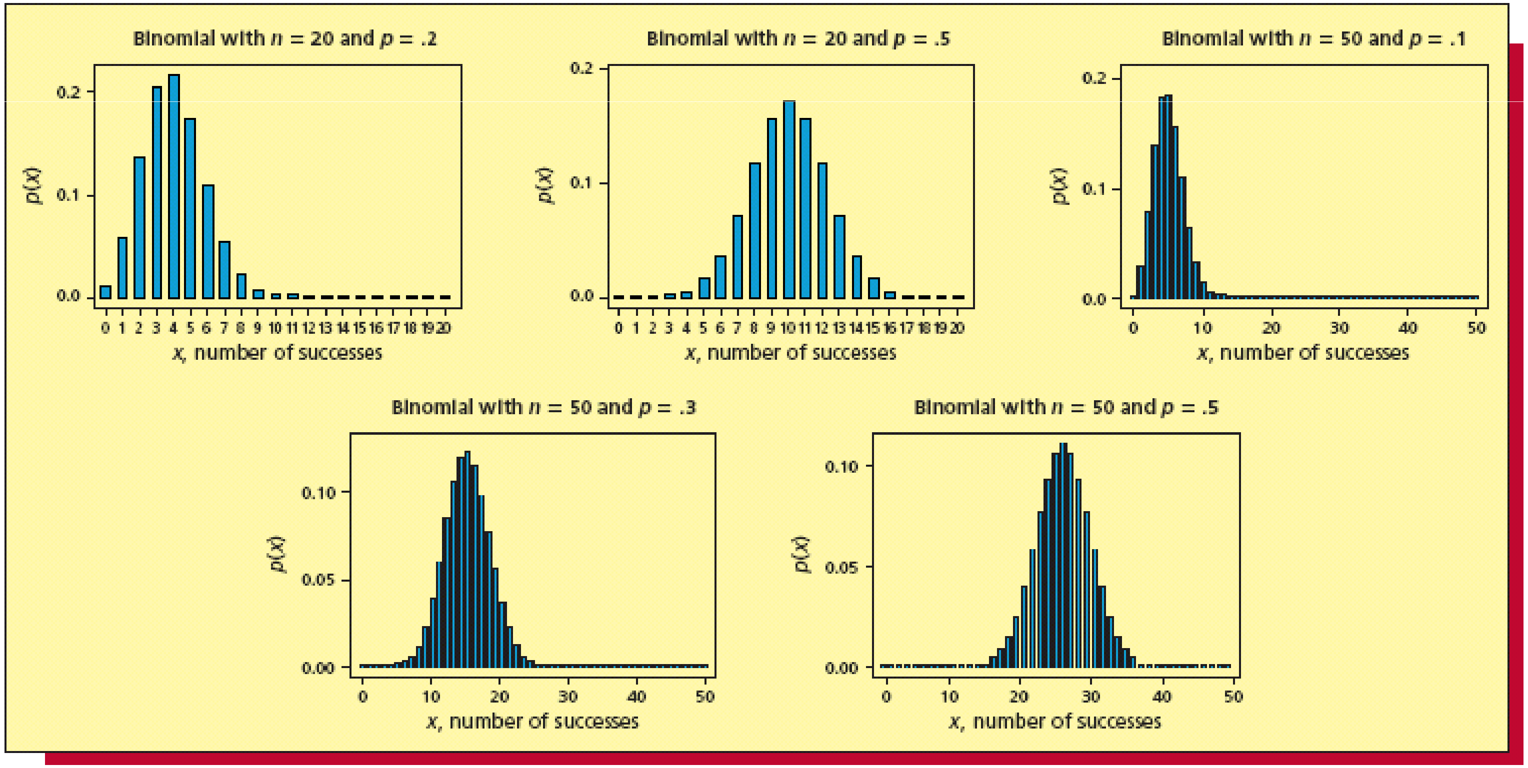
Finding a Tolerance Interval

Finding a tolerance interval $[\mu \pm k\sigma]$ that contains 99% of the measurements in a normal population



Normal Approximation
to the Binomial

- The figure below shows several binomial distributions
- Can see that as *n* gets larger and as *p* gets closer to 0.5, the graph of the binomial distribution tends to have the symmetrical, bell-shaped, form of the normal curve



Normal Approximation
to the Binomial (Cont.)

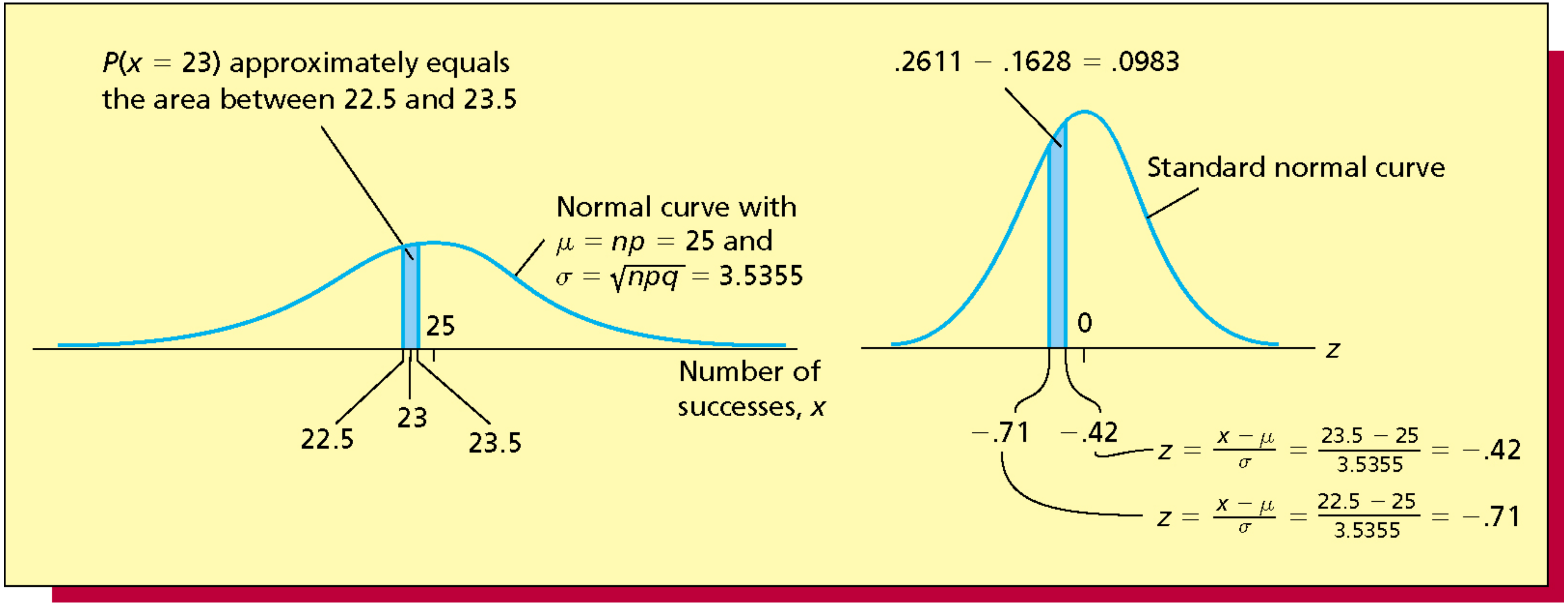
- Generalize observation from last slide for large p
- Suppose x is a binomial random variable, where n is the number of trials, each having a probability of success p
  - Then the probability of failure is $1 - p$
- If n and p are such that $np \geq 5$ and $n(1 - p) \geq 5$, then x is approximately normal with

$$\mu = np \quad \text{and} \quad \sigma = \sqrt{np(1-p)}$$

Example: Approximating Binomial
distribution by using normal distribution

Example 5.8: Approximating binomial probability $P(x = 23)$ by using normal curve when

$$\mu = np = 25 \text{ and } \sigma = \sqrt{np(1-p)} = 3.5355$$



Continuity correction: 22.5 ≤ z ≤ 23.5

The 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

The Exponential Distribution (Cont.)

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

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

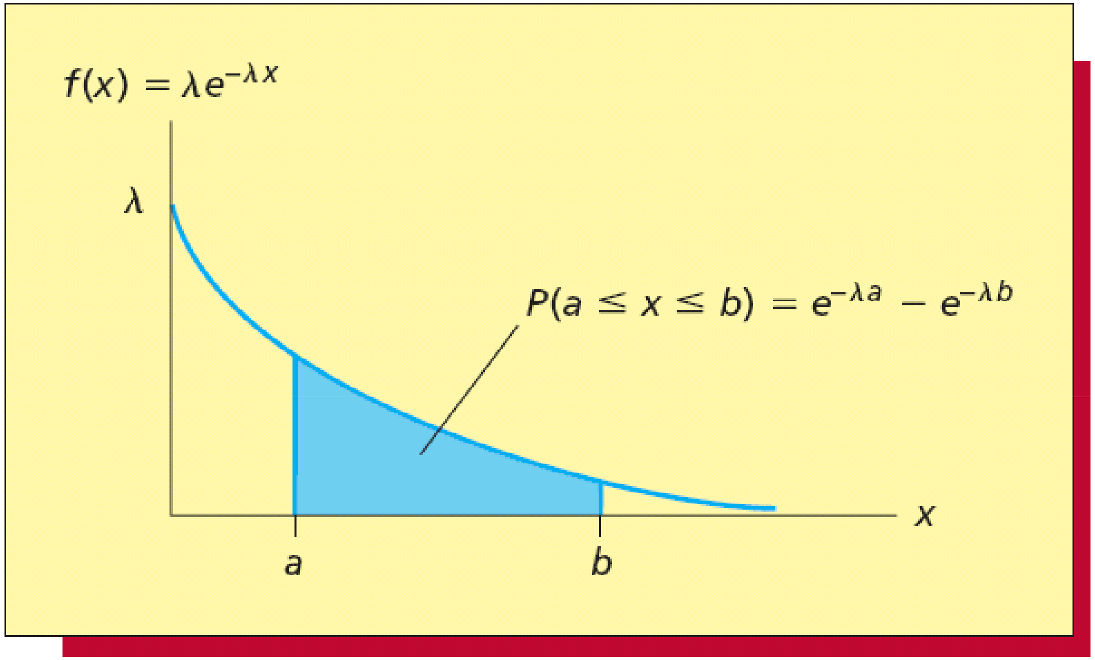
The probability that *x* is any value between the given values *a* and *b* (*a* < *b*) is

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

and

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

The Exponential Distribution (Cont.)



The mean μ_X and standard deviation σ_X of an exponential random variable x are

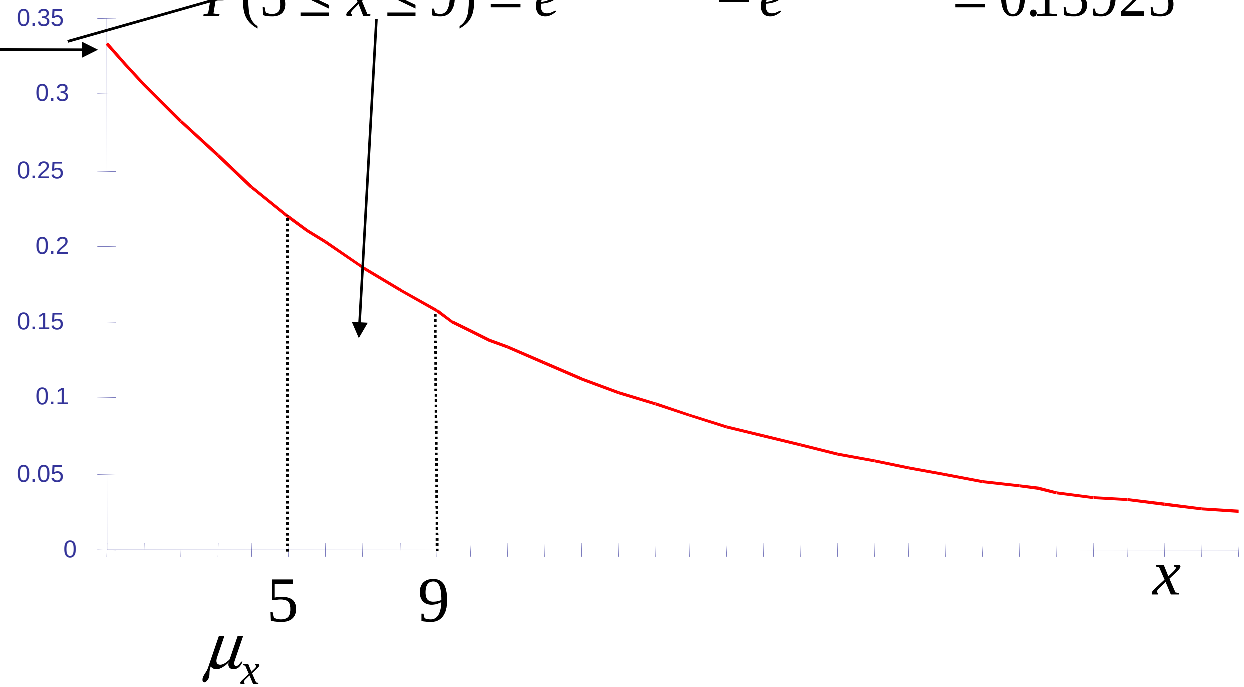
$$\mu_X = \frac{1}{\lambda} \text{ and } \sigma_X = \frac{1}{\lambda}$$

Example: Calculating exponential probabilities

Let μ_x=3.0 or 1/ μ_x=1/3=0.333,

$$P(5 \le x \le 9) = e^{-(0.333)5} - e^{-(0.333)9} = 0.13925$$

$\lambda = 0.333$

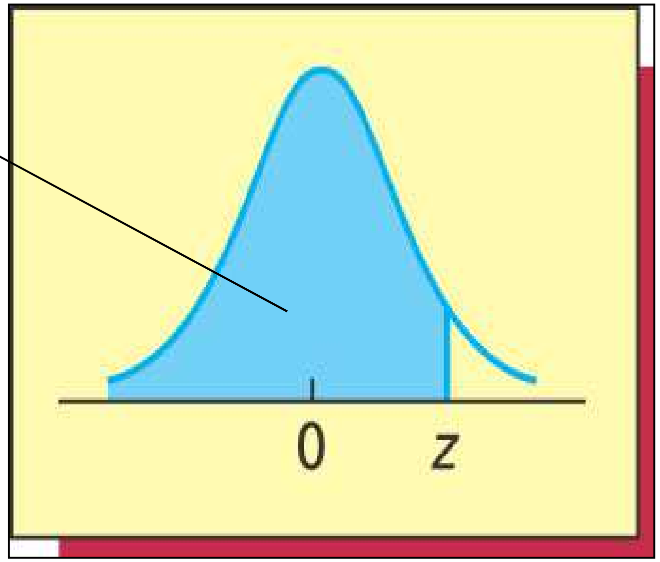


The Cumulative Normal Table

The cumulative normal table gives, for different values of z_0, the area under the standard normal curve to the left of z_0.

- Including negative *z values*
- The cumulative normal table gives the probability $P(z \leq z_0)$
- Table 5.3 and Table A.19 in Appendix A

The normal cumulative table gives the blue area



The Cumulative Normal Table

- Most useful for finding the probabilities of threshold values like $P(z \le a)$ or $P(z \ge b)$
  - Find $P(z \le 1)$
    - Find directly from cumulative normal table that

$$P(z \le 1) = 0.8413$$

  - Find $P(z \ge 1)$
    - Find directly from cumulative normal table that

$$P(z \le 1) = 0.8413$$

    - Since areas under the normal sum to 1

$$P(z \ge 1) = 1 - P(z \le 1)$$

get

$$P(z \ge 1) = 1 - 0.8413 = 0.1587$$